

**OPTIMIERUNG DER EFFEKTE VON PSYCHOTHERAPIE  
WIRKSAMKEIT, PRÄDIKTOREN UND NEBENWIRKUNGEN IN DER STATIONÄREN  
ROUTINEVERSORGUNG**

**Dissertation**

zur Erlangung des Doktorgrades der Naturwissenschaften

(Dr. rer. nat.)

dem Fachbereich Psychologie der Philipps-Universität Marburg

vorgelegt von

**Philipp Herzog**

aus Bad Langensalza

Marburg an der Lahn, Juni 2020

Vom Fachbereich Psychologie der Philipps-Universität Marburg

(Hochschulkennziffer 1180) am 21.08.2020 als Dissertation angenommen.

Erstgutachterin: Prof. Dr. Eva-Lotta Brakemeier

Zweitgutachter: Prof. Dr. Winfried Rief

Tag der mündlichen Prüfung: 21.08.2020

## Danksagungen

Die Danksagung ist nicht Teil der Veröffentlichung.



# Inhaltsverzeichnis

Abbildungsverzeichnis .....	VIII
Tabellenverzeichnis .....	IX
1 Zusammenfassung und Abstract .....	10
1.1 Zusammenfassung .....	10
1.2 Abstract .....	13
2 Theoretischer und empirischer Hintergrund .....	16
2.1 Psychotherapieforschung in der heutigen Zeit .....	24
2.1.1. Forschungsmodelle der Psychotherapie .....	24
2.1.2. Phasen der Psychotherapieforschung .....	26
2.1.3. Bedeutung von Phase IV-Studien im wissenschaftlichen Prozess: Benchmarking der Ergebnisse aus klinischen Studien .....	29
2.2 Praxisbasierte Evidenz als Ergänzung zur Evidenzbasierten Praxis: Rolle von Praxisorientierter Forschung .....	37
2.3 Probleme der Psychotherapieforschung .....	43
2.3.1. Definition von Psychotherapieerfolg: Wahl eines geeigneten Therapieoutcomes und Kriterien für klinisch-bedeutsame Veränderungen .....	43
2.3.2. Absolute vs. Relative Wirksamkeit: Dodo-Bird-Verdict oder die Frage nach spezifischen und allgemeinen Effekten .....	47
2.3.3. "End of average" oder warum nicht nur Effektstärken zählen .....	49
2.3.4. Verfahrensinflation .....	51
2.3.5. Stagnation der Effekte von Psychotherapie .....	52
2.3.6. Research-Practice-Gap oder warum die Elfenbeintürme weiter im Treibsand sinken .....	54
2.4 Personalisierte Medizin als Anhaltspunkt zur Optimierung von Effekten in der Psychotherapie: Ein möglicher Wechsel im Forschungsparadigma .....	56

2.4.1.	Klinische multivariate Prädiktionsmodelle: Einbezug von Machine Learning und Bayesscher Statistik.....	58
2.4.2.	Vorhersage von Effekten in der Psychotherapie: Prognostische vs. preskriptive Prädiktoren.....	60
2.4.3.	Untersuchung von Veränderungsmechanismen: Mediatoren von Effekten in der Psychotherapie.....	65
2.5	Kehrseite der Medaille: Von positiven Effekten zu negativen Effekten von Psychotherapie .....	68
3	Darstellung des Dissertationsvorhabens.....	72
3.1	Relevanz und Herleitung der Fragestellungen .....	72
3.2	Fragestellungen des Dissertationsvorhabens .....	75
4	Zusammenfassung der Studien .....	76
4.1	Studie 1: Wirksamkeit und Prädiktoren einer stationären Psychotherapie für Borderline-Persönlichkeitsstörungen .....	76
4.2	Studie 2: Wirksamkeit und Prädiktoren einer stationären Psychotherapie für Posttraumatische Belastungsstörungen.....	78
4.3	Studie 3: Wirksamkeit und Prädiktoren einer stationären Psychotherapie für Zwangsstörungen.....	80
4.4	Studie 4: Wirksamkeit und Prädiktoren einer stationären Psychotherapie für depressive Störungen .....	82
4.5	Studie 5: Erfassung von negativen Effekten in der Psychotherapie .....	84
4.6	Studie 6: Zusammenhang zwischen dem Auftreten von negativen Effekten und dem Outcome eines stationären CBASP-Programms für chronische Depression .....	86
5	Zusammenfassende Diskussion und Ausblick .....	88
5.1	Zusammenfassung der Ergebnisse im Rahmen der Leitfragen.....	89
5.2	Stärken und Einschränkungen .....	94
5.3	Implikationen für die klinische Praxis .....	105

5.4 Perspektiven für zukünftige Forschung.....	110
5.5 Fazit.....	121
Literaturverzeichnis .....	123
Anhang A: Studie 1.....	187
Anhang B: Studie 2.....	244
Anhang C: Studie 3.....	294
Anhang D: Studie 4.....	340
Anhang E: Studie 5.....	400
Anhang F: Studie 6.....	413
Anhang G: Publikationsverzeichnis.....	433
Anhang H: Eidesstattliche Erklärung.....	437

## Abbildungsverzeichnis

<i>Abbildung 1.</i> Evidenzbasierte Praxis bei Psychotherapie nach Roth & Fonagy (2005).....	21
<i>Abbildung 2.</i> Wechselbeziehung zwischen Evidenzbasierter Praxis und Praxisbasierter Evidenz nach Barkham & Mellor-Clark (2003).....	42
<i>Abbildung 3.</i> Schematische Darstellung der prozentualen Verteilung der Verbesserung in Psychotherapiepatient*innen als Funktion von therapeutischen Faktoren (Lambert, 2013).....	54
<i>Abbildung 4.</i> Obere Abbildung: Patient*innencharakteristikum X sagt ein besseres Outcome vorher, ungeachtet von der Behandlung. Untere Abbildung: Patient*innencharakteristikum X sagt einen größeren Benefit von jeder Behandlung vorher; adaptiert nach Simon & Perlis (2010).....	62
<i>Abbildung 5.</i> Patient*innencharakteristikum X sagt einen größeren Benefit für die Behandlung A als von Behandlung B vorher (d.h., moderiert den Vergleich in der Response auf Behandlung A und Behandlung B), adaptiert nach Simon & Perlis (2010).....	64



## Tabellenverzeichnis

Tabelle 1	<i>Merkmale von Forschungsparadigmen zur Erhebung des Psychotherapieoutcomes nach Donenberg, Lyons, &amp; Howard (1999)</i> .....	25
Tabelle 2	<i>Verschiedene Ansätze zur Generalisierung der Psychotherapieforschung nach Lueger (2002)</i> .....	25
Tabelle 3	<i>4 Phasen der Psychotherapieforschung in Analogie zur Pharmaprüfung modifiziert nach Buchkremer &amp; Klingberg (2001)</i> .....	26

# 1 Zusammenfassung und Abstract

## 1.1 Zusammenfassung

In den letzten Jahrzehnten hat sich durch klinische Studien mit methodisch fundierten Designs eine breite Evidenzbasis von Psychotherapie für verschiedene psychische Störungen gebildet. Durch verschiedene randomisiert-kontrollierte Studien (engl.: „randomized controlled trial“; RCT), die in Meta-Analysen zusammengefasst werden, konnte die Wirksamkeit (im Sinne der „efficacy“) mit meist großen Effekten nachgewiesen werden. Gemäß des Phasenmodells der Psychotherapieforschung werden in der letzten Phase – Phase IV – Psychotherapien im naturalistischen Kontext, d.h. unter alltäglichen Routinebedingungen, untersucht. Im Rahmen eines praxisorientierten Forschungsparadigmas können solche praxisbasierten Studien eine Ergänzung zu RCTs darstellen, um vorhandene Probleme in der Psychotherapieforschung zu adressieren (z. B. Research-Practice-Gap, Stagnation von Effekten). Insbesondere in Deutschland werden psychisch schwer erkrankte Patient\*innen<sup>1</sup> oft im stationären Kontext (z. B. in psychosomatischen oder psychiatrischen Kliniken) behandelt. In diesem Rahmen mangelt es an Studien, welche diese Art der klinischen Wirksamkeit (im Sinne der „effectiveness“) mit geringeren Ein- und Ausschlusskriterien untersucht. Die Befunde zu Variablen, welche den Behandlungserfolg stabil vorhersagen, sind dabei oft heterogen. Daher soll im ersten Schritt dieser Dissertation die Wirksamkeit von stationärer Psychotherapie bei verschiedenen psychischen Störungen (Borderline-Persönlichkeitsstörung, Posttraumatische Belastungsstörung, Zwangsstörung, Depression) in deutschen psychosomatischen Kliniken untersucht als auch Ansatzpunkte zur Optimierung der Effekte durch die Analyse von Prädiktoren identifiziert werden.

Die Studien 1-4 deuten darauf hin, dass spezifische evidenzbasierte Psychotherapieprogramme gemäß der jeweiligen nationalen Behandlungsleitlinien in der stationären psychosomatischen Versorgung wirksam implementiert wurden, jedoch in ihrer Wirksamkeit Unterschiede aufweisen. Während die Effektstärken für die Behandlung von Zwangsstörungen (N = 1.595 bzw. N = 514) und depressiven Störungen (N = 22.681 bzw. N = 6.377) als groß beurteilt werden

---

<sup>1</sup> Im Rahmen der vorliegenden Arbeit wurde versucht, eine geschlechtsneutrale Sprache zu verwenden, indem beispielsweise von „Patient\*innen“ gesprochen wird. An Stellen, an denen dies nicht möglich war, wurde aus Gründen der besseren Lesbarkeit in der Regel die weibliche Form verwendet. Es sei jedoch ausdrücklich darauf hingewiesen, dass immer alle Geschlechter eingeschlossen sind.

können, sind sie für die Behandlung von Posttraumatischen Belastungsstörungen (N = 612) mittel bis groß sowie für die Behandlung von Borderline-Persönlichkeitsstörungen (N = 878 bzw. N = 703) klein bis mittel. Aufgrund der großen Stichproben können die Effekte als Benchmark dienen und in zukünftigen Studien als Vergleich herangezogen werden. Darüber hinaus zielt die vorliegende Dissertation im nächsten Schritt darauf ab, allgemeine Prädiktoren für den Behandlungserfolg von Psychotherapie bei den o. g. Störungsbildern zu identifizieren. Im Sinne von prognostischen Variablen werden – getrennt für die verschiedenen Störungsbilder – solche Charakteristika von Patient\*innen im Rahmen von klinischen Prädiktionsmodellen bestimmt, die angeben, welche Patientin besser bzw. schlechter von der Behandlung profitiert. Bei Patient\*innen mit Borderline-Persönlichkeitsstörungen erweisen sich als wichtige Prädiktoren für den Therapieabschluss ein höherer Schulabschluss sowie eine komorbide rezidivierende depressive Störung und für symptomspezifische Veränderungen niedrige Affektregulationsfähigkeiten und keine vorherige ambulante Psychotherapie. Bei Patient\*innen mit Posttraumatischen Belastungsstörungen scheint eine stärkere Ausprägung im Bereich Psychotizismus, die Anzahl an Diagnosen und eine Asthma-Diagnose konstant in einem negativen prädiktiven Zusammenhang zum Behandlungserfolg zu stehen. Bei Patient\*innen mit Zwangsstörungen sind die Symptomschwere bei Aufnahme, die allgemeine psychopathologische Belastung, eine höhere soziale Unterstützung und mehr Waschwänge wichtige Prädiktoren für Symptomveränderungen. Bei Patient\*innen mit Depressionen beeinflussen zwar fast alle soziodemographischen und klinischen Faktoren vor der Behandlung die psychopathologische Gesamtbelastung auf unterschiedlichen Faktoren zu Beginn der Behandlung, dieser direkte Einfluss schwächt sich im Verlauf der Behandlung jedoch ab. Während dieser Behandlung sind diverse Faktoren veränderungsresistent und deuten auf stabile psychopathologische Faktoren hin, wie z. B.: Suizidalität, agoraphobische Ängste, Lebensunzufriedenheit, physische Beeinträchtigung und Schmerzen. Der stärkste Einfluss ergibt sich von der Suizidalität zu Behandlungsbeginn auf negative Kognitionen bei Behandlungsende, von agoraphobischen Symptomen bei Behandlungsbeginn auf die Ängstlichkeit bei Behandlungsende und von der körperlichen Beeinträchtigung bei Behandlungsbeginn auf die wahrgenommene Beeinträchtigung bei Behandlungsende.

Letztlich wird auch die andere Seite der Medaille betrachtet: neben positiven Effekten kann Psychotherapie auch negative Effekte hervorrufen. Im Rahmen eines systematischen Reviews werden in Studie 5 die Messinstrumente zur Erhebung von negativen Effekten unter

psychometrischen Gesichtspunkten evaluiert sowie eine konsensuale Definition und Konzeptualisierung von negativen Effekten vorgeschlagen und Empfehlungen zur Verbesserung der Erfassung von negativen Effekten abgeleitet. Dieses systematische Review ergibt, dass die vorhandenen Messinstrumente zwar viele Bereiche abdecken, ihre psychometrischen Gütekriterien hingegen zumeist als unbefriedigend zu beurteilen sind. Anschließend wird aus den Implikationen des Reviews der Zusammenhang zwischen dem Auftreten von negativen Effekten und dem Behandlungserfolg während eines stationären CBASP<sup>2</sup>-Programms für chronische Depression<sup>3</sup> in Studie 6 untersucht. Es zeigt sich, dass fast alle Patient\*innen mindestens einen negativen Effekt berichten, aber lediglich die wahrgenommene Abhängigkeit von der Therapeutin in einem negativen Zusammenhang zum Behandlungserfolg steht.

Insgesamt lässt sich feststellen, dass sich die Ergebnisse aus Phase III-Studien der Wirksamkeitsforschung auf ein angewandtes stationäres Setting übertragen lassen. Im Rahmen der Studien 1-4 konnten einige Prädiktoren (z. B. Symptomschwere zur Aufnahme) aus vorherigen Befunden von RCTs repliziert werden; zum Teil spielen jedoch bei stationären Psychotherapiepatient\*innen auch andere Variablen eine wichtige Rolle zur Vorhersage des Behandlungserfolgs, was auf unterschiedliche Gründe vor allem im Hinblick auf die Spezifika des Settings (z. B. höhere Therapiedosis in kürzerer Zeitspanne) und der Heterogenität der Stichproben zurückzuführen ist. Bei den Studien 5-6 konnten sowohl definitorische Aspekte von negativen Effekten adressiert werden, um deren Erfassung noch zu verbessern, als auch die Relevanz durch die Zusammenhänge zum Behandlungserfolg näher beleuchtet werden. Durch alle Studien können insgesamt Ansatzpunkte dafür gefunden werden, um die Effekte von stationärer Psychotherapie in Deutschland noch zu optimieren. Die Studien 1-4 liefern beispielsweise sowohl Benchmarks für die absolute Wirksamkeit von Psychotherapie in der stationären Routineversorgung als auch durch die Prädiktoranalysen praxisbasierte Evidenz, welche wiederum in zukünftigen RCTs z. B. bei Randomisierungsprozeduren berücksichtigt werden kann. Durch Rückmeldeschleifen der Ergebnisse aller Studien an klinisch-tätige Praktiker\*innen kann kurzfristig ein evidenzbasiertes Outcome-Monitoring stattfinden und langfristig das Praxis-Forschungsnetzwerk in Deutschland gestärkt werden.

---

<sup>2</sup> Cognitive Behavioral Analysis System of Psychotherapy (McCullough, 2003)

<sup>3</sup> Im DSM-5 (Association American Psychiatric, 2015) wird die Diagnose der persistierenden depressiven Störung eingeführt (Klein, Willenborg, & Schweiger, 2015). Da die Studie 6 noch unter dem Deckmantel des Begriffs der chronischen Depression durchgeführt wurde, wird im Folgenden dieser Begriff verwendet.

### 1.2 Abstract

Over the last decades, clinical studies with methodologically sound designs have created a broad evidence base of psychotherapy for various mental disorders. Numerous randomized controlled trials (RCT), which are subsequently summarized in meta-analyses, have demonstrated the efficacy of psychotherapy with mostly large effects. According to the phase model of psychotherapy research, the last phase - phase IV - investigates psychotherapy in a naturalistic context, i.e. under everyday routine conditions. Within the framework of a practice-oriented research paradigm, such practice-based studies can complement RCTs to address existing problems in psychotherapy research (e.g. research-practice gap, stagnation of effects). In Germany in particular, patients with severe mental disorders are often treated in an inpatient setting (e.g. in psychosomatic or psychiatric clinics). In this naturalistic context, there is a lack of studies that investigate the clinical effectiveness with lower inclusion and exclusion criteria. Furthermore, the findings on variables that predict treatment outcome are often heterogeneous. Therefore, the objectives of this dissertation are (1) to examine the effectiveness of inpatient psychotherapy for various mental disorders (borderline personality disorder, post-traumatic stress disorder, obsessive-compulsive disorder, depression) in German psychosomatic clinics, and (2) to identify starting points for optimizing these effects by analyzing predictors of outcome.

Overall, studies 1-4 indicate that specific evidence-based psychotherapy programs have been effectively implemented in inpatient psychosomatic care according to the respective national treatment guidelines, but indicate differences in their effectiveness expressed by different effect sizes. While the effect sizes for the treatment of obsessive-compulsive disorders (OCD; N = 1,595 and N = 514, respectively) and depression (N = 22,681 and N = 6,377, respectively) can be classified as large, there are medium to large effects for the treatment of post-traumatic stress disorders (PTSD; N = 612) and small to medium effects for the treatment of borderline personality disorders (BPD; N = 878 and N = 703, respectively). Due to large sample sizes, the effects might serve as a benchmark and therefore can be used as a comparison in future studies. Moreover, the present dissertation aims to identify general predictors for treatment outcome of psychotherapy considering the above-mentioned mental disorders. In the sense of prognostic variables, patient characteristics were determined - separately for the various mental disorders - within the framework of clinical prediction models that indicate which

patient benefit more vs. less from the treatment. In patients suffering from BPD, the most important predictors for treatment completion are a higher education and a comorbid recurrent depressive disorder, and low affect regulation skills and no previous outpatient psychotherapy for symptom-specific changes. In patients with PTSD, a higher degree of psychoticism, total number of diagnoses, and a diagnosis of bronchial asthma consistently show a stable negative predictive relationship to treatment outcome. In patients with OCD, symptom severity at admission, general psychopathological distress, higher social support and more washing compulsions are important predictors of symptom changes. In patients suffering from depression, almost all sociodemographic and clinical pretreatment factors significantly influence the overall psychopathological burden on various baseline factors, but this direct influence weakens over the course of treatment. Several factors are resistant to change and indicate stable psychopathological factors during treatment, such as suicidality, agoraphobia, life dissatisfaction, physical disability and pain. The strongest impact is exerted by suicidality at the beginning of treatment on negative cognitions at the end of treatment, by agoraphobia at the beginning of treatment on anxiety at the end of treatment and by physical disability at the beginning of treatment on perceived disability at the end of treatment. Ultimately, the other side of the coin is also considered: in addition to positive effects, psychotherapy can also produce negative effects. Within the framework of a systematic review, study 5 evaluates instruments for assessing negative effects in the light of psychometric properties, proposes a consual definition and conceptualization of negative effects and derives recommendations for improving the assessment. This systematic review shows that although the existing instruments cover many relevant domains, their psychometric properties are usually unsatisfactory. Subsequently, the implications of this review are used to investigate the relationship between the occurrence of negative effects and treatment outcome during an inpatient CBASP program for chronic depression as described in study 6. It is shown that although almost all patients report having experienced at least one negative effect, only a perceived dependence on the therapist is negatively related to treatment outcome.

Overall, it can be stated that the results from phase III studies of efficacy research can be transferred to an applied inpatient setting. With regard to studies 1-4, some predictors (e.g. symptom severity at admission) could be replicated to previous findings of RCTs; however, other variables also play an important role in predicting treatment outcome in these samples of inpatients, which can be attributed to different reasons, such as the specifics of the setting

(e.g. higher treatment dose in a shorter period of time) and the heterogeneity of the samples. In studies 5-6, both definitional aspects of negative effects could be addressed in order to further improve their detection, and the relevance of specific negative effects could be determined in more detail by establishing the relationship to treatment outcome. In summary, all studies can provide starting points for optimizing the effects of inpatient psychotherapy in Germany. For example, studies 1-4 in particular provide benchmarks for the absolute effectiveness of psychotherapy in routine inpatient care as well as practice-based evidence through the predictor analyses, which in turn can be taken into account in future RCTs, e.g. in randomisation procedures. By feedback loops of the results of studies 1-6 to clinical practitioners, an evidence-based outcome monitoring can take place in the short term and, in the long term, the practice-research network in Germany can be strengthened.

### 2 Theoretischer und empirischer Hintergrund

Psychotherapie<sup>4</sup> wird in Deutschland in verschiedenen Versorgungseinrichtungen im ambulanten, tagesklinischen bzw. teilstationären und stationären Setting angeboten. Die Durchführung von Psychotherapieprogrammen im teil- als auch vollstationären Kontext unterscheiden dabei deutschsprachige Länder insbesondere von den englischsprachigen forschungstarken Ländern wie USA und UK insofern, als dass dort fast ausschließlich Psychotherapie im ambulanten Setting angeboten und evaluiert wird. Dieser Umstand ist vor allem durch geschichtliche Prozesse in Deutschland (Reimer, 2007) und dem daraus resultierenden Gesundheitssystem sowie Unterschieden in der Erstattung von Gesundheitsdienstleistungen bedingt. Viele der in den englischsprachigen Ländern entwickelten Psychotherapien sind daher auf das ambulante Setting manualisiert zugeschnitten, wohingegen in Deutschland die stationäre Psychotherapie in psychiatrischen und psychosomatischen Einrichtungen ein sehr großes Versorgungsangebot widerspiegelt. Die stationäre Psychotherapie kann im Vergleich zur ambulanten Psychotherapie verschiedene Vorteile bieten: Unterschiede in der Therapiedosis, multilaterale Übertragungsmöglichkeiten durch ein multiprofessionelles Team, multimodales Therapieangebot, Gemeinschaft von Mitpatient\*innen und die begrenzte Behandlungsdauer (Reimer, 2007). In den deutschen S-3 Leitlinien lassen sich die Indikationsbereiche für stationäre Psychotherapien nachlesen (z. B. für Depression: DGPPN, BÄK, KBV, & AWMF, 2017). Darin wird empfohlen, dass Patient\*innen insbesondere bei langjährig bestehender und ambulant nicht ausreichend behandelbarer psychischer Störung, ausgeprägter Komorbidität, Gefährdung durch Suizidalität und sehr hoher Krankheitsschwere stationär behandelt werden sollten. Aber auch bei therapieverhindernden Faktoren in der häuslichen Lebensumwelt der Patientin sowie bei ausbleibendem Erfolg ambulanter Behandlungsversuche kann sich die Notwendigkeit der Behandlung durch ein multiprofessionelles Team mit Krankenhaus typischen Heilmethoden ergeben. Obwohl viele Patient\*innen mit psychischen Erkrankungen in Deutschland im stationären Rahmen behandelt werden, ist die Kluft zwischen der derzeitigen Versorgungssituation und der Anzahl an

---

<sup>4</sup> Barlow (2004) schlägt vor, dass Behandlungen für bestimmte Störungen, die sich als wirksam erwiesen haben, im Englischen als psychologische Behandlungen (engl.: „psychological treatments“) bezeichnet werden sollten, um sie von der allgemeinen Psychotherapie zu unterscheiden. Im englischsprachigen Ausland scheinen erstere etablierte Behandlungen innerhalb der Gesundheitsversorgungssysteme (d.h. von Dritten erstattungsfähig) zu sein und letztere häufig außerhalb des Geltungsbereichs der Gesundheitsversorgungssysteme eingesetzt werden. Im deutschsprachigen Raum ist jedoch der Begriff Psychotherapie üblicher, sodass im Folgenden dieser verwendet wird.



wissenschaftlichen Studien und Publikationen insbesondere in internationalen Fachzeitschriften relativ groß. Daher soll in der vorliegenden Dissertation dieser Diskrepanz Rechnung getragen werden, indem Studien zur Wirksamkeit von Psychotherapie in der stationären Routineversorgung in Deutschland durchgeführt und Prädiktoren für den Behandlungserfolg für verschiedene Störungsbilder identifiziert werden. Darüber hinaus sollen neben positiven Effekten auch negative Effekte von Psychotherapie näher beleuchtet werden.

Bei Betrachtung der Geschichte der wissenschaftlich fundierten Psychotherapie, wird deutlich, dass es ein weiter Weg war, um zum Schluss zu gelangen, dass Psychotherapie eine wirksame Behandlung von psychischen Störungen darstellt. Die Anzahl an klinischen Studien bzgl. Psychotherapie über die Zeit ist von einzelnen Studien im Jahr 1970 bis hin zu über 10.000 Studien im Jahr 2020 rasant angestiegen, ebenso die Meta-Analysen bzgl. der Wirksamkeit von Psychotherapie von einer Meta-Analyse im Jahr 1980 bis weit über 700 Meta-Analysen im Jahr 2020. Die Psychotherapie ist daher lange über Eysenck (1952) hinausgewachsen, der provokativ behauptete, Psychotherapie könne im besten Fall eine spontane Remission nicht verhindern, auf jeden Fall keinen substanziellen über die spontane Remission hinausgehenden Beitrag zur Verbesserung der psychischen Gesundheit leisten. Durch unzählige RCTs, Meta-Analysen und systematische Reviews von unterschiedlichen internationalen Arbeitsgruppen ist in der Forschung hinlänglich bekannt, dass Psychotherapien wirken und positive Effekte haben.

Je nach Störungsbild werden für die für kognitive Verhaltenstherapie (KVT, auch: kognitiv-behaviorale Therapie) sehr hohe Effektstärken (ES) im Prä-Post-Vergleich berichtet, so zum Beispiel für die Behandlung von Angststörungen mit  $ES = 1.30$  (Bandelow et al., 2015), für Posttraumatische Belastungsstörungen (PTBS) mit  $ES = 1.63$  (Watts et al., 2013) und Zwangsstörungen mit  $ES = 1.31$  (Öst, Havnen, Hansen, & Kvale, 2015). Für die KVT der Depression liegen im Vergleich zu anderen Psychotherapien die meisten Publikationen vor (Barth et al., 2013). Zwar wurden gerade im letzten Jahr bei einer Re-Analyse erstaunlich geringe Effektstärken für die Psychotherapie der Depression berichtet (Cuijpers, Karyotaki, Reijnders, & Ebert, 2019), insgesamt kann aber davon ausgegangen werden, dass Psychotherapie zur Behandlung von Depressionen wirkt (Munder et al., 2019a). In einer groß angelegten RCT zeigte sich, dass insbesondere die KVT und Interpersonelle Psychotherapie (IPT) sowohl kurz- als auch langfristig wirksam sind bei episodischer Depression (Lemmens et al., 2015; Lemmens et al., 2019), wobei die Befunde zur chronischen Depression geringer ausfallen

(Cuijpers, van Straten, et al., 2010; Schramm et al., 2019, 2017). Solche Studien tragen immer wieder zur Erneuerung von diversen nationalen und internationalen Richtlinien bei (z. B. S-3 Leitlinien in Deutschland, NICE Guidelines in UK).

Grawe (1997) bezeichnete diese Etablierung der Psychotherapie in der Versorgung als *Legitimationsphase*, d.h. eine Etablierung der Psychotherapieschulen, der auch durch einen dogmatischen Schulenstreit begleitet wurde. Die Diskussion über die Sinnhaftigkeit von Verfahrenskonzepten und damit einhergehend die Richtlinien-Psychotherapien wird bis heute in Deutschland immer wieder aufgegriffen und vor allem von Seiten psychodynamisch orientierter Forscher\*innen und Praktiker\*innen befürwortet (Leichsenring et al., 2019). In der anschließenden *Wettbewerbsphase*, der durch das Paradigma der evidenzbasierten Medizin zu einer noch stärkeren wissenschaftlichen Orientierung führte, vollzog sich der sog. „Horse run“ zwischen verschiedenen Psychotherapieschulen. Diese Phase war geprägt von der Etablierung von empirisch unterstützten Behandlungen (engl.: „empirically supported treatments (EST)“<sup>5</sup>) durch RCTs von manualisierten – meist ambulanten – Psychotherapien, dessen Einbezug dann in systematischen Reviews und Meta-Analysen letztlich zu internationalen Leitlinien geführt hat. Diese Epoche hat ebenso zur Entwicklung von neuen psychologischen Therapien für spezifische diagnostische Gruppen geführt und integrative, schulenübergreifende Ansätze mit wissenschaftlichen Evaluationen im Sinne einer absoluten Wirksamkeit hervorgebracht (z. B. Dialektische Behaviorale Therapie (DBT), Schematherapie, Acceptance- and Commitment-Therapie (ACT), CBASP, IPT). Dadurch wurde das Portfolio zur Auswahl von diversen Psychotherapien bei einer Vielzahl an psychischen Störungen erheblich erweitert, was Kritiker\*innen jedoch als Verfahrensinflation bezeichnen (Rubel & Lutz, 2016), da viele der neuen Verfahren durch zu wenige Studien zur relativen Wirksamkeit keine empirischen Handlungsrichtlinien für Praktiker\*innen mit sich bringen.

---

<sup>5</sup> Empirisch unterstützte Behandlungen (auch: Evidenzbasierte Behandlungen bzw. „Evidence-based treatments (EBT)“) können definiert werden als Therapien (bzw. spezifische Interventionen und Techniken), die sich bei der Evaluation im Rahmen von RCTs mit spezifischen Patient\*innenpopulationen als wirksam erwiesen haben (Wampold & Imel, 2015).

Diese Phase hat zuweilen in Forschungskreisen zu einem „Dodoismus-Klima“<sup>6</sup> geführt, welches sich auf die fehlende Evidenz von relativer Wirksamkeit durch komparative Non-inferiority-RCTs bezieht (Gibbons et al., 2016; Steinert et al., 2017). Dieser Diskurs dauert bis heute an, was in Deutschland durch die Veröffentlichung der Ergebnisse der LAC-Studie einhergehend mit einer kritischen Diskussion insbesondere von methodischen Problemen untermauert wird (Hautzinger et al., 2019; Kaiser, Kolar, Kok, & Schuster, 2019; Leuzinger-Bohleber, Hautzinger, Keller, et al., 2019). Diese grundsätzliche Debatte bzgl. der Verwendung von Non-Inferiority-Designs findet ebenso in internationalen Fachzeitschriften statt (Leichsenring, Abbass, Driessen, et al., 2018; Rief & Hofmann, 2018, 2019). Eine solche Non-Inferiority-Studie zeigt beispielsweise, dass die psychodynamische Therapie der KVT bei der Behandlung von depressiven Störungen nicht unterlegen sein muss (Driessen et al., 2013). Unabhängig von dieser Debatte über einzelne Effekte kann jedoch festgehalten werden, dass die KVT für viele psychische Störungen die breiteste und somit robusteste Evidenzbasis vorweist (Barth et al., 2013). Ohne eine adäquate Implementierung und spätere Disseminierung von ESTs in den klinischen Alltag, ist der Wert solchen Fortschritts in der Forschung hingegen limitiert (Steinfeld, Coffman, & Keyes, 2009).

Die „Division 12: Society of Clinical Psychology“ der *American Psychological Association* (APA) hatte im letzten Jahr 84 Verfahren gelistet für insgesamt 32 verschiedene psychische Störungen (Stand: August 2019). Im englischsprachigen Ausland hat die APA die beste Forschungsevidenz als Hauptkomponente der evidenzbasierten Praxis (EBP) durch die „APA Presidential Task Force on Evidence-Based Practice“ formuliert (APA, 2006). Die EBP ist ein erweiterter Begriff für die klinische Praxis, welche durch Erkenntnisse über Interventionen, klinisches Fachwissen bzw. Expertise und die Bedürfnisse, Werte und Präferenzen der Patientin und deren Integration informiert ist, um Entscheidungen über die individuelle Versorgung für optimale Behandlungsansätze zu treffen (APA, 2006).<sup>7</sup>

---

<sup>6</sup> Das sog. Dodo-Bird-Verdict (auch: Äquivalenzparadoxon) in der Psychotherapie bezeichnet eine kontroverse Debatte in der Forschung, wonach einige Forscher\*innen davon ausgehen, dass alle psychologischen Therapie wirksam sind, d.h. alle empirisch validierten Psychotherapien ohne Berücksichtigung von spezifischen Komponenten ungefähr ähnliche Outcomes produzieren. Sie basiert auf dem gleichnamigen Dodo-Bird aus Alice im Wunderland („Everybody has won, and all must have prizes.“).

<sup>7</sup> Die evidenzbasierte Praxis ist angelehnt an die evidenzbasierte Medizin. Diese bezeichnet den Versuch zur gewissenhaften, expliziten und vernünftigen Verwendung der aktuell besten Beweise bei der Entscheidungsfindung über die Behandlung einzelner Patient\*innen, wobei die Praxis der evidenzbasierten Medizin die Integration individueller klinischer Expertise mit den besten verfügbaren externen klinischen Evidenzen aus der systematischen Forschung bedeutet (Sackett & Rosenberg, 1995b, 1995a).

In Abbildung 1 ist ein Modell zur Erreichung von EBP für die Psychotherapie nach Roth & Fonagy (2005) dargestellt. Die Vorzüge und Spannungsfelder von EBP werden in Barkham et al. (2010) gut beschrieben. Dazu zählen wissenschaftliche Spannungsfelder wie die Debatten zu:

- (1) 'No evidence of effectiveness' vs. 'Evidence of no effectiveness'
- (2) Interne vs. Externe Validität
- (3) Experiment vs. Beobachtung
- (4) Klinische Wirksamkeit vs. Kosteneffektivität
- (5) Quantitative vs. Qualitative Forschungsansätze
- (6) Spezifische vs. allgemeine Faktoren
- (7) Empirisch-unterstützte Behandlungen vs. Veränderungsprozesse (siehe „Hourglass model“ von Salkovskis (1995))

Darüberhinaus gibt es professionelle Probleme, z. B. statistische vs. klinische Beurteilung als auch die Unterstützung der Entscheidungsfindung vs. Bedrohung der Legitimität in der Ausübung der Profession, die zu bedenken sind.

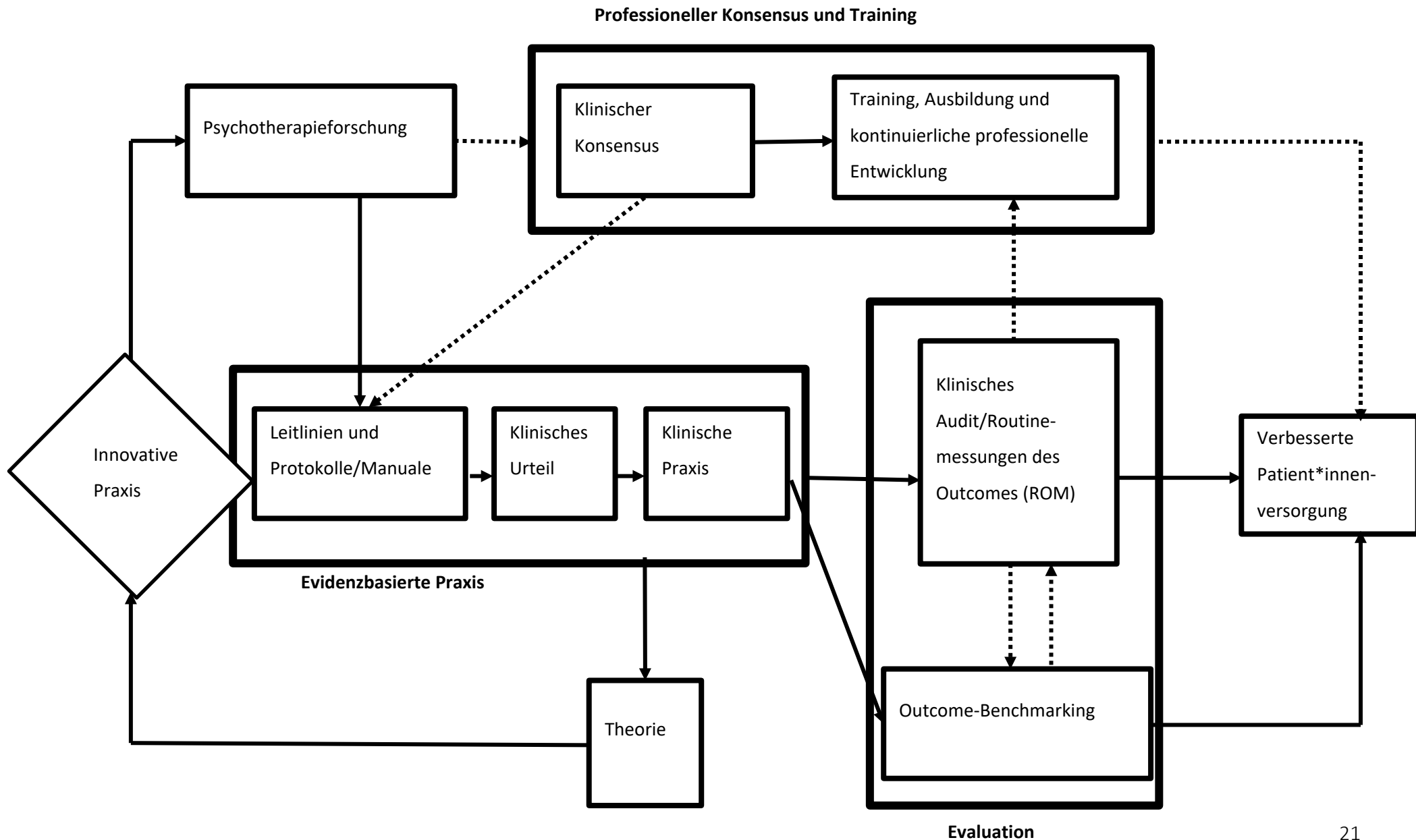


Abbildung 1. Evidenzbasierte Praxis bei Psychotherapie nach Roth & Fonagy (2005).

In der *Verschreibungsphase* und seit der Etablierung von ESTs wurde die von Kazdin bereits 1956 formulierte und von anderen Psychotherapieforscher\*innen immer wieder aufgegriffene Frage „What works for whom?“ (Fonagy, 2010; Roth & Fonagy, 2005) adressiert, um die Vielzahl an bereits vorhandenen etablierten Psychotherapien noch zu optimieren. Diese Frage ist heute aktueller denn je: In dieser Phase hat sich durch Ansätze aus der Prozessforschung viel den Wirkmechanismen gewidmet, um algorithmenbasierte Entscheidungen treffen zu können. Beispielsweise konnte übergreifend gezeigt werden, dass die Realitätsbetrachtung und Problemlösung und nicht die Bearbeitung des Vermeidungsverhalten die wirksamen Komponenten der Psychotherapien sind (Lambert, 2013). Durch derartige Wirkmechanismusstudien kann die Versorgung von Menschen mit psychischen Erkrankungen insofern verbessert werden, als dass auf wissenschaftlicher Grundlage Entscheidungsbäume entwickelt werden können, welche Patient\*innen bei welchen Problemen in welchem Setting welche Behandlung im Sinne der notwendigen Therapieintensität eines „Stepped-care“-Ansatzes erhalten sollen (Delgadillo, Gellatly, & Stephenson-Bellwood, 2015). Dadurch kann die klinische Entscheidungsfindung verbessert werden, welche wiederum auf praxisbasierter Evidenz erfolgen und zu empirisch-unterstützten Entscheidungen führen sollte. So könnten langfristig im Rahmen von computationalen Ansätzen optimale klinische Entscheidungen z. B. durch die Nutzung von Markov-Entscheidungsprozessen erleichtert werden (z. B. Bennett & Hauser, 2013).

Wie den Publikationen zu entnehmen ist, handelt es sich zwar um Phasen, die in unterschiedlichen Jahrzehnten der wissenschaftlich fundierten Psychotherapie starteten, aber sich bis heute in der Forschung in unterschiedlichem Ausmaß weiterhin finden lassen. Insgesamt kann über verschiedene Meta-Analysen über die Jahre gesagt werden, dass die aggregierte Effektstärke in Zusammenhang mit der absoluten Wirksamkeit erstaunlich konsistent ist über die Jahrzehnte und sich zwischen  $.75 < ES < .85$  für Psychotherapie bei verschiedenen psychischen Störungen bewegt. Eine angemessene und vertretbare Punktschätzung der Wirksamkeit von Psychotherapie kann daher generell als  $ES = 0.80$  im Sinne einer Daumenregel angenommen werden (Wampold & Imel, 2015). Ob sich diese ES auch in der psychotherapeutischen Versorgung unter Routinebedingungen finden lassen, bleibt weiterhin eine offene Forschungsfrage. Um Klaus Grawes Vision „Von der Konfession zur Profession“ (Grawe, 1994) weiterhin gerecht zu werden, ist es sinnvoll, die Effekte von Psychotherapie stetig intensiv zu untersuchen und weiter zu optimieren insofern, als dass jede

Patientin die Behandlung für seine individuelle psychischen Problemen erhält, von der sie mit großer Wahrscheinlichkeit am meisten profitieren wird.

Im Kontext dieser Optimierung von Effekten in der Psychotherapie bewegt sich die vorliegende Dissertation. Im Folgenden sollen Grundlagen der Psychotherapieforschung in der heutigen Zeit skizziert werden, die für die vorliegende Dissertation relevant sind. Im Anschluss sollen aktuelle Probleme benannt, alternative Forschungsparadigmen identifiziert und die Kehrseite der Psychotherapie beleuchtet werden.

### 2.1 Psychotherapieforschung in der heutigen Zeit

#### 2.1.1. Forschungsmodelle der Psychotherapie

Empirische Studien zu Interventionen der psychischen Gesundheit können in zwei Modelle eingeteilt werden: Forschung durch klinische Studien (engl.: „clinical trial research“) und Forschung im Rahmen der psychischen Gesundheitsversorgung (engl.: „mental health service research“). Diese können sich auf vier verschiedenen Kernkomponenten unterscheiden (Donenberg et al., 1999; Howard, Moras, Brill, Martinovich, & Lutz, 1996; Lambert, Hansen, & Finch, 2001; Lueger, 2002):

1. Ebene der Behandlungen vs. Ebene der Patient\*innen
2. Ebene der Methodik (konfirmatorische Studien vs. exploratorische Studien)
3. Vorgeschriebene Prozesse des „Behandlungspakets“ (theoretisch-informierte, kontrollierte Behandlung vs. Behandlung, wie sie tatsächlich durchgeführt werden)
4. Vorhersagen von Gruppen vs. Vorhersage von individuellen Patient\*innen

Donenberg et al. (1999) clustert diese Ansätze im Sinne einer Dichotomisierung in zwei Bereiche, welche die Merkmale auf einem Kontinuum von „Mental Health Services Research“ bis hin zu „Clinical Trial Research“ anordnet. Diese Organisation ist der Tabelle 1 zu entnehmen.



Tabelle 1

*Merkmale von Forschungsparadigmen zur Erhebung des Psychotherapieoutcomes nach Donenberg et al. (1999)*

<b>„Mental health service“-Forschungsmodell</b>	<b>„Clinical trial“-Forschungsmodell</b>
Wirksamkeit im Sinne von „Effectiveness“	Wirksamkeit im Sinne von „Efficacy“
Exploratorisches Vorgehen	Konfirmatorisches Vorgehen
Induktive Schlussfolgerungen	Deduktive Schlussfolgerungen
Patient*innenfokussierte Studien	Behandlungsfokussierte Studien
Externe Validität	Interne Validität
Große, heterogene Stichproben	Homogene, diagnostisch klare Stichproben
Natürlicher Verlauf der Behandlung	Protokoll-bestimmte Behandlung
Statische Kontrolle für klinische Risikovariablen	Methodologische Kontrolle von Risikofaktoren

Zudem hat sich eine Erweiterung dieses dichotomen Ansatzes zur Generalisierung von Psychotherapieforschung in den Folgejahren etabliert (Lueger, 2002). Diese Erweiterung ist in Tabelle 2 schematisch dargestellt.

Tabelle 2

*Verschiedene Ansätze zur Generalisierung der Psychotherapieforschung nach Lueger (2002)*

<b>Teilnehmer*innen</b>	<b>Forschungsfokus</b>	
	<b>Behandlung</b>	<b>Patient*innen</b>
<b>Population</b>	Technologie-Transfer-Studien (z. B. klinische Studien wie RCTs)	Naturalistische Studien (z. B. Versorgungsforschungsstudien)
<b>Fall</b>	Einzelfallstudien	Studien zur Unterstützung der klinischen Entscheidungsfindung

Die Studien in der vorliegenden Dissertation können eher dem „Mental Health Service“-Forschungsparadigma (Donenberg et al., 1999) sowie den naturalistischen Studien (Lueger, 2002) zugeordnet werden.

### 2.1.2. Phasen der Psychotherapieforschung

Die wissenschaftliche Fundierung von Psychotherapieverfahren lässt sich in verschiedene Phasen unterteilen (Metzler & Krause, 1997). Die vorgeschlagene Einteilung dieser Phasen ist in Tabelle 3 dargestellt und lehnt sich an Buchkremer & Klingberg (2001) an. Das Modell wurde im weiteren Verlauf noch erweitert (Krampen, Schui, & Hütter, 2008).

Tabelle 3

*4 Phasen der Psychotherapieforschung in Analogie zur Pharmaprüfung modifiziert nach Buchkremer & Klingberg (2001)*

Phase	Inhaltlicher Schwerpunkt	Art der Studien
I	Konzeptentwicklung	Explizierung theoretischer Annahmen Analogstudien Kasuistiken Entwicklung von Manualen
II	Exploration mit (nicht kontrollierten) Pilotstudien	Verlaufsbeschreibende Einzelfallstudien Prä-Post-Analysen in verschiedenen Populationen Offene Studien Durchführbarkeitsstudien Prozess-Outcome-Studien
III	Wirksamkeitsprüfung	Efficacy-Studien / RCTs Experimentelle Einzelfallstudien Randomisierte Effectiveness-Studien
IV	Anwendung unter Routinebedingungen, klinische Nützlichkeit	Nicht randomisierte Effectiveness-Studien Erweiterte Einschlusskriterien Zunehmend naturalistische Bedingungen Anwendungsbeobachtungen

Der inhaltliche Schwerpunkt der *Phase I* liegt auf der Konzeptentwicklung mit Hilfe von Analogstudien, Kasuistiken, Manualentwicklung und Explizierung theoretischer Annahmen. In der Phase II wird der Schwerpunkt auf der Exploration mit nicht-kontrollierten Pilotstudien gelegt und kann durch verlaufsbeschreibende Einzelfallstudien, Prä-Post-Analysen in verschiedenen Patient\*innenpopulationen, offenen Studien, Durchführbarkeitsstudien und Prozess-Outcome-Studien durchgeführt werden. Die ersten zwei Phasen ebnen den Weg für die Phase, welche für die wissenschaftliche Anerkennung am wichtigsten ist: Die *Phase III* dient der Wirksamkeitsprüfung, welche anhand von Wirksamkeitsstudien, experimentellen Einzelfallstudien und randomisierten Effektivitätsstudien durchgeführt wird (im Sinne von „efficacy“-Studien). Studien zur Untersuchung der Wirksamkeit (häufig auch: Outcome) von Psychotherapie können in sog. „efficacy“- und „effectiveness“-Studien unterteilt werden (Mintz, Drake, & Crits-Christoph, 1996; Seligman, 1995). Die erste Form der Wirksamkeitsforschung (engl.: „efficacy research“) zielt darauf ab, zu untersuchen, ob eine bestimmte Intervention einen spezifischen, messbaren Effekt hat, und kann sich auch mit Fragen der Sicherheit, Durchführbarkeit, Nebenwirkungen und dem Level einer angemessenen Dosis befassen, wohingegen das Ziel der zweiten Form der Wirksamkeitsforschung (engl.: „effectiveness research“) es ist, zu ermitteln, ob wirksame Behandlungen einen messbaren, positiven Effekt haben können, wenn sie in breiten Bevölkerungsschichten und in anderen Gesundheitsversorgungsbereichen eingesetzt werden (Barkham et al., 2010; Freeman & Power, 2007)<sup>8</sup>. Der Goldstandard sind RCTs, welche langfristig im Rahmen von Meta-Analysen und systematischen Reviews zusammengefasst werden können, um mittlere Effektstärken zu berechnen. Die Definition von ESTs basiert auf den kumulierten Daten und Befunden von verschiedenen Studien zur Wirksamkeit („efficacy“) von Psychotherapie (Crits-Christoph, Hollon, & Wilson, 2005; Weisz, Weersing, & Henggeler, 2005; Westen, Novotny, & Thompson-Brenner, 2004).

Bezugnehmend auf die in der vorliegenden Dissertationen untersuchten psychischen Störungen deuten Meta-Analysen und systematische Reviews auf die Wirksamkeit insbesondere der von trauma-fokussierter KVT für die Posttraumatische Belastungsstörung hin (Bisson et al., 2007; Cusack et al., 2016; Ehlers et al., 2010) sowie von KVT bzw. Expositionstherapie mit Reaktionsmanagement (engl.: „Exposure and Response Prevention

---

<sup>8</sup> Die Autor\*innen unterscheiden auch noch die *Praxisforschung* und die *Gesundheitsversorgungssystemforschung* (Barkham et al., 2010; Freeman & Power, 2007).

(ERP“) für Zwangsstörungen (Öst et al., 2015), wobei bei sehr schweren Zwangsstörungen durchaus eine Kombinationsbehandlung aus Psychotherapie und pharmakologischer Therapie sich als sinnvoll erweist (Skapinakis et al., 2016). Bei der Behandlung von Borderline-Persönlichkeitsstörungen zeigen Studien, dass insbesondere die DBT effektiv ist (Cristea et al., 2017; Kliem, Kröger, & Kosfelder, 2010; Linehan, Armstrong, Suarez, Allmon, & Heard, 1991; Linehan et al., 2015; McCauley et al., 2018; Mehlum et al., 2016; Ougrin, Tranah, Stahl, Moran, & Asarnow, 2015; Panos, Jackson, Hasan, & Panos, 2014; Priebe et al., 2012). Eine aktuelle Meta-Analyse zeigt, dass DBT nicht-suizidales selbstverletzendes Verhalten sowie suizidale Handlungen mit mittleren Effektstärken reduziert (DeCou, Comtois, & Landes, 2019) und auch Behandlungsabbrüche reduzieren kann (Barnicot, Katsakou, Marougka, & Priebe, 2011; Linehan, 1993; Swenson, Sanderson, Dulit, & Linehan, 2001). Zur Wirksamkeit von Psychotherapie bei Depressionen liegen vielzählige Befunde vor (Cuijpers et al., 2020), insbesondere die KVT und IPT scheinen besonders kurz- sowie langfristig wirkungsvolle Therapien bei episodischer Depression zu sein (Lemmens et al., 2015; Lemmens et al., 2019), wohingegen die kurz- und langfristigen Effekte bei chronischer Depression geringer sind (Cuijpers, van Straten, et al., 2010; Schramm et al., 2019, 2017), ebenso wie die Effekte bei Depression in stationären Behandlungssettings (Schefft, Guhn, Brakemeier, Sterzer, & Köhler, 2019). Gerade im Vergleich zu einigen medizinischen Erfolgsraten sind diese Befunde relativ gute Ergebnisse: z. B. hat Chemotherapie bei Brustkrebs eine ES von  $d = 0.08 - 0.11$  (Lutz, 2010).

Die letzte Phase – *Phase IV* – richtet ihren inhaltlichen Schwerpunkt auf die Anwendung unter Routinebedingungen und klinische Nützlichkeit. Dabei spielen zumeist Aspekte von Generalisierbarkeit, Machbarkeit und Kosteneffektivität eine Rolle. Unter naturalistischen Versorgungsbedingungen und Anwendungsbeobachtungen werden nicht-randomisierte Wirksamkeitsstudien mit erweiterten Einschlusskriterien und geringeren Ausschlusskriterien durchgeführt.

Der Fokus der traditionellen Psychotherapieforschung, wie in der Einleitung skizziert, liegt klassischerweise auf der Outcomeforschung im Sinne von RCTs zur Untersuchung von Wirksamkeit bzw. Effektivität, d.h. die Durchführung von Phase III-Studien. Dieser essentielle Teil hat maßgeblich zur Entwicklung von Leitlinien geführt, worin insbesondere die KVT durch die robuste und überzeugende Datenlage oft als die Methode der 1. Wahl genannt wird. Hingegen stellt diese Phase nur einen Teil dar und in den letzten Jahren wurden vermehrt

naturalistische Studien im Rahmen der Phase IV in der Psychotherapieforschung durchgeführt. Relevant für die vorliegende Dissertation sind beispielsweise Ergebnisse einer Studie, die darauf hinweisen, dass die kognitive Therapie für PTBS effektiv in der ambulanten Routineversorgung implementiert werden kann (Ehlers et al., 2013). Eine andere Studie zeigt, dass KVT in einem stationären Routinesetting große Effekte erzielen kann für multimorbide, vorbehandelte Patient\*innen mit Zwangsstörungen (Gönner, Limbacher, & Ecker, 2012), was sich auch mit Ergebnissen aus einem Review deckt (Veale et al., 2016). Einige weitere Studien untersuchen die naturalistische Wirksamkeit von Psychotherapie bei depressiven Störungen, deren vielversprechende Ergebnisse auf eine gute Umsetzung in der Routineversorgung hindeuten (Gyani, Shafran, Layard, & Clark, 2013; Karlin et al., 2012; Kilbourne, Williams, Bauer, & Areal, 2012; Lutz, Schiefele, Wucherpfennig, Rubel, & Stulz, 2016). Nur eine geringe Anzahl an Studien untersucht hingegen das Ausmaß, in dem DBT in der Routineversorgung unter alltäglichen Bedingungen untersucht wurde (z. B. Higgins & Green, 2011; Kröger et al., 2006). Insgesamt mangelt es noch an solchen naturalistischen Studien in der internationalen Forschungsliteratur, insbesondere in unterschiedlichen Behandlungssettings wie z. B. stationäre Psychotherapie, welche einen wichtigen Stellenwert in der Versorgung von Menschen mit psychischen Erkrankungen in Deutschland einnimmt. Diese Forschungslücke versucht die vorliegende Dissertation zu adressieren.

### 2.1.3. Bedeutung von Phase IV-Studien im wissenschaftlichen Prozess: Benchmarking der Ergebnisse aus klinischen Studien

Die Durchführung von Behandlungen in klinischen Studien und in der Praxis kann durchaus variieren (Westen et al., 2004), wobei die Unterscheidung zwischen klinischen Studien und naturalistischen Settings nicht immer trennscharf ist (Shadish, Navarro, Matt, & Phillips, 2000). In qualitativ hochwertigen Psychotherapiestudien (v.a. RCTs) wird sich daher oft an den sog. CONSORT<sup>9</sup>-Kriterien orientiert (Pittler, Blümle, Meerpohl, & Antes, 2011). Für weitere detailliertere Ausführungen zu qualitativ hochwertigen Psychotherapiestudien wird auf Areal & Kraemer (2015) verwiesen.

Kurz gesagt: Eine Behandlung wird in Phase III-Studien (z. B. RCTs) durch ausgewählte Therapeut\*innen durchgeführt, die ein Training und Supervision erhalten, um die Behandlung

---

<sup>9</sup> CONSORT = 'Consolidated Standards of Reporting Trials' (Pittler et al., 2011)

manualgeleitet durchzuführen. Meistens sind die Patient\*innenpopulationen homogen, werden ausgewählt durch gewisse Ein- und Ausschlusskriterien, welche oft Patient\*innen mit Komorbiditäten (z. B. bei Depression: psychotische Symptome, Substanzmissbrauch, Persönlichkeitsstörungen) sowie bei Gefährlichkeit durch z. B. Suizidalität und Einnahme von psychotroper Medikation systematisch ausschließen. Patient\*innen werden rekrutiert, um an einer klinischen Studie teilzunehmen, die oft an einer universitären Klinik oder medizinischen Einrichtung angegliedert ist. Gegner\*innen von solchen klinischen Studien behaupten daher, dass validere Antworten zum Therapieoutcome und zu Veränderungsprozessen eher aus Studien in „real-world“-Settings als in kontrollierten Laborstudien hervorgebracht werden (Weisz, Doss, & Hawley, 2005; Weisz, Jensen-Doss, & Hawley, 2006). Entsprechend kann folgendes Argument, welches die Durchführung von Phase IV-Studien unterstützt, angebracht werden: Die engen Kriterien in Psychotherapiestudien aus der Phase-III erhöhen immens die interne Validität und den Ausschluss von Alternativerklärungen, wodurch die Robustheit und Aussagekraft der Ergebnisse sichergestellt wird. Durch diese hohen Standards haben solche Studien zur Etablierung der Psychotherapie als anerkanntes Heilverfahren im Gesundheitsversorgungssystem geführt. Gleichzeitig ist die externe Validität solcher Studien zu einem gewissen Grad eingeschränkt, welche insbesondere ergänzend in Phase IV-Studien Rechnung getragen werden soll.

Wie bereits erwähnt, wird in der Phase IV die Wirksamkeit unter Versorgungsbedingungen („Effectiveness“-Studien) festgestellt. Diese Phase ist insbesondere gekennzeichnet durch nicht randomisierte Studien, große Prä-Post-Studien, Studien mit erweiterten Einschlusskriterien und unter insgesamt zunehmend naturalistischen Bedingungen. Die Unterschiede zwischen klinischen Studien bzw. randomisiert-kontrollierten Studien und praxisbasierten Studien ist in einer Tabelle in Barkham, Hardy, & Mellor-Clark (2010) für verschiedene Merkmale (u.a. Design, Hypothesen, Allegianz der Forscherin, Stichprobe, Behandlungen, Messungen) aufgelistet sowie die Kriterien für die Patient\*innenselektion in solchen Studien in Shadish et al. (1997) nachzulesen. Der große Vorteil dieser Studien ist neben den oft großen Stichproben die hohe externe sowie ökologische Validität, die durch weniger enge Kriterien allerdings zulasten der internen Validität geht.

Es existieren verschiedene Wege, „Effectiveness“-Studien durchzuführen (Wampold & Imel, 2015):

- (1) Untersuchung der Effekte von Studien über Psychotherapie-"Behandlungspakete", die in klinisch relevanteren Umgebungen untersucht werden (Studien zur klinischen Repräsentativität)<sup>10</sup>
- (2) Benchmarking-Studien, die die aus naturalistischen Umgebungen gewonnenen Erkenntnisse über die Wirksamkeit der Behandlung (im Sinne der „effectiveness“) mit den aus klinischen Studien (i.e. RCTs) abgeleiteten Effekten (im Sinne der „efficacy“) vergleichen (Barkham et al., 2008; Merrill, Tolbert, & Wade, 2003; Minami, Wampold, Serlin, Kircher, & Brown, 2007), und
- (3) klinische Studien, die eine evidenzbasierte Behandlung mit der Standardbehandlung (engl.: „treatment as usual (TAU)“) vergleichen. Die Auswahl von TAU muss jedoch sorgfältig getroffen werden, da dies wichtige Implikationen für die Aussagekraft der Ergebnisse hat.

Eine wichtige meta-analytische Studie zur klinischen Repräsentativität von KVT zur Behandlung von diversen Angststörungen ergibt, dass diese im Allgemeinen in der Routinepraxis wirksam sind mit großen Prä-Post-Effektstärken innerhalb eines Range von  $ES = 0.90 - 1.45$  und dadurch auch innerhalb der Benchmarks für die meisten Störungen, wobei im Speziellen die Behandlung von PTBS sogar noch effektiver und die Behandlung von Panikstörungen etwas weniger effektiv ist (Stewart & Chambless, 2009). Die Ergebnisse bedeuten, dass die durchschnittliche Patientin zu Behandlungsende bessere Werte als 83% und 93% der Patient\*innen vor der Behandlung erzielt. Eine wichtige Limitation ist hingegen, dass nur 2 von 56 Studien das maximale Rating zur klinischen Repräsentativität auf allen Dimensionen erfüllt, sodass Praktiker\*innen die Ergebnisse anzweifeln können.

---

<sup>10</sup> Die Strategie zur Feststellung klinischer Repräsentativität ist die älteste Methode und überprüft meta-analytisch bereits publizierte Studien, um zu bestimmen, inwieweit Faktoren, die zwischen der realen Welt und dem Labor unterscheiden (z. B. Einstellung, Methode der Patient\*innenrekrutierung, zufällige Zuweisung von Patient\*innen zur Behandlung, Verwendung des Behandlungsmanuals), die Behandlungseffekte moderieren (Wampold & Imel, 2015).

Im Folgenden wird vertiefend auf Benchmarking-Studien eingegangen, da dieser Strategie in der vorliegenden Arbeit gefolgt wurde:

Der Kontext des Benchmarking kann im Rahmen der Qualitätsverbesserung verstanden werden und unterscheidet das absolute Qualitätsmanagement (engl.: „Total Quality Management (TQM)“) und später die kontinuierliche Qualitätsverbesserung (engl.: „Continuous Quality Improvement (CQI)“) (Barkham et al., 2010). Das CQI versucht, die zugrundeliegenden Arbeitsprozesse innerhalb eines Systems vor der tatsächlichen Produktion des Outcomes zu verstehen und zu verbessern. Diese haben zu verschiedenen Forschungsprogrammen geführt (z. B. National Outcomes Management Project (NOMP) durch Dewan et al., 2000; klinische Informatik durch Howard, Moras, Brill, Martinovich, & Lutz, 1996; Feedbacksysteme: Lambert, Hansen, & Finch, 2001), die bis heute ergänzt und erweitert wurden (z. B. Harmon, Hawkins, Lambert, Slade, & Whipple, 2005; Lambert, Whipple, & Kleinstäuber, 2018; Lutz, Rubel, Schwartz, Schilling, & Deisenhofer, 2019). Im Rahmen solcher Forschungsprogramme werden oft relative kurze Skalen verwendet, die eine gute Reliabilität und konkurrente Validität aufzeigen, atheoretisch sind und normative Daten sowohl aus Patient\*innenstichproben und Kontrollstichproben haben. Eine Augenscheinvalidität sowohl für Patient\*innen und Therapeut\*innen zeigen sie auf, da sie oft sowohl symptom-spezifische als auch funktionsbezogene Maße beinhalten (Barkham et al., 2010). Der große Vorteil ist, dass diese Messinstrumente als Outcomemaße entwickelt wurden im Gegensatz zu allgemein verwendeten Messinstrumenten, deren Konstrukte auf Theorien (z. B. BDI-II, Beck, Steer, & Brown, 1996) oder Diagnosen (z. B. PHQ-9, Kroenke, Spitzer, & Williams, 2001) basieren. Ein Nachteil ist, dass sie nicht sonderlich spezifisch sind und somit wichtige Aspekte nicht berücksichtigen.

Ein Schätzer für die wahre Wirksamkeit und Indikator für Implementierung evidenzbasierter Verfahren im Feld können solche Benchmarking-Studien unter Routinebedingungen darstellen (vgl. Durham et al., 2004; Ehlers et al., 2013b; Franklin, Abramowitz, Kozak, Levitt, & Foa, 2000; Gillespie, Duffy, Hackmann, & Clark, 2002; Hahlweg, Fiegenbaum, Frank, Schroeder, & von Witzleben, 2001; Lincoln et al., 2003; Merrill et al., 2003). Solche Studien teilen das gemeinsame Ziel, eine Evidenzbasis bzgl. der Wirksamkeit von Psychotherapie, wie sie in der Routineversorgung erbracht werden, zu liefern und lassen sich als Phase IV-Studien einordnen. Diese Studien können daher als Repräsentant von klinischen



Studien im Sinne einer Benchmark verstanden werden und bieten normative Standards zur Orientierung. In den meisten Benchmarkstudien wird dabei die klinische Wirksamkeit berücksichtigt, wohingegen die Kosteneffektivität oft vernachlässigt wird und vergleichsweise wenige Studien dazu existieren (Priebe et al., 2012; Slade et al., 2017; Vos, Corry, Haby, Carter, & Andrews, 2005; Zala, Brabban, Stirzaker, Kartha, & McCrone, 2019).

Im weiteren Sinne existieren vier Formen von Benchmarking nach Camp (1989): *Internes Benchmarking* bezieht sich auf Benchmarking innerhalb einer Klinik bzw. innerhalb eines Klinikverbunds, *Kompetitives Benchmarking* auf das Benchmarking zwischen Kliniken und Klinikverbänden, *Funktionelles Benchmarking* auf das Benchmarking bei ähnlichen Funktionen über verschiedene Einrichtungen hinweg (z. B. ähnliche Zielsetzung) und *generisches Prozess-Benchmarking* auf das Benchmarking bei allgemeinen Prozessen in verschiedenen Kontexten (z. B. positive Therapieerwartungen). Im engeren Sinne beinhaltet Benchmarking daher den Vergleich zwischen der Schätzung des Ausmaßes, indem Psychotherapie in klinischen Studien wirksam ist, mit den Effekten, die in naturalistischen Settings gefunden werden. Dabei werden oft Effektstärken aus klinischen Studien zur Herstellung einer Benchmark verwendet, gegen die die Effekte aus dem naturalistischen Setting verglichen werden. Die naturalistische Behandlung kann entweder die TAU oder eine evidenzbasierte Psychotherapie sein, welche in einem naturalistischen Setting angewendet wird (Wampold & Imel, 2015).

Barkham et al. (2010) unterscheidet im Rahmen des „Effectiveness“-Paradigmas behandlungsfokussierte Benchmarkingstudien und im Rahmen des Benchmarkings für praxisbasierte Studien die sog. patient\*innenfokussierten Benchmarkingstudien.

Für behandlungsfokussierte Benchmarkingstudien existieren zahlreiche Beispiele mit unterschiedlichen Ergebnissen, u. a. Merrill et al. (2003) zur kognitiven Therapie der Depression, Wade et al. (1998) zur evidenzbasierten Behandlung der Panikstörung, Weersing & Weisz (2002) zur Routinebehandlung von Jugendlichen mit depressiven Störungen, Scheeres, Wensing, Knoop, & Bleijenberg (2008) zur KVT bei chronischer Erschöpfung und Westbrook & Kirk (2005) zur KVT im Allgemeinen. Die Ergebnisse eines Reviews zu solchen Studien von Hunsley & Lee (2007) deuten vielversprechenderweise daraufhin, dass die Implementierung von etablierten Behandlungen in das klinische Setting mit ähnlichen Effekten erfolgreich gelingen kann.

Im Rahmen von patient\*innenfokussierten Benchmarkingstudien existieren diverse Forschungsprogramme zur Aggregation von Outcomedaten, welche gekennzeichnet sind durch das Sammeln von Daten zu psychotherapeutischen Behandlungen, die mindestens ein kurzes Messinstrument zur allgemeinen Beurteilung des psychischen Gesundheitsstatus umfasst, geleitet ist durch die Bemühungen eines Forscher\*innenteams im Rahmen der Prinzipien der patient\*innenorientierten Forschung (siehe Abschnitt 2.2) und letztlich Ausweitungen in große Versorgungssysteme beinhaltet. Diese aggregierten Datensets bestehen meistens aus naturalistischen Studien zur Wirksamkeit (im Sinne der „Effectiveness“) von Psychotherapie und beinhalten mindestens Maße zu Prä-Post-Behandlungsveränderungen zusätzlich zur Erhebung von Patient\*innencharakteristika bei Aufnahme (Barkham et al., 2010). Die wichtigsten Beispiele solcher Forschungsansätze sind bei Barkham et al. (2010) nachzulesen. Es wurden verschiedene Methoden zur Bestimmung von Benchmarks entwickelt, welche Vergleiche zu klinischen Studien ermöglichen sollen (Minami, Serlin, Wampold, Kircher, & Brown, 2008; Minami, Wampold, et al., 2008; Minami et al., 2007):

- (1) Bilden von Prä-Post-Benchmarks aus klinischen Studien
- (2) Schätzen der Wirksamkeit („effectiveness“) in der Routineversorgung durch die Berechnung von Prä-Post-Effektstärken (ES)
- (3) Benchmarking der Effektstärke in der Routineversorgung gegen die Benchmark aus den Daten der klinischen Studien (oft wird Cohen’s Regel von  $d = 0.2$  als Grenze verwendet)

Die Frage, nach welchen Kriterien eine passende klinische Studie ausgewählt wird, gegen die als Benchmark getestet wird, bleibt jedoch offen. Eine logische Strategie wäre es, die zuletzt veröffentlichte Studie als Benchmark zu verwenden, sodass sukzessive immer wieder neue und unterschiedliche Vergleichsbedingungen über die Zeit hinweg generiert werden (Lambert, 2013). Um eine arbiträre Entscheidungsfindung für eine gewählte Benchmark zu reduzieren, könnte eine andere Strategie sein, jede neue Studie gegen den kumulativen Bestand an früheren Studien zu benchmarken (z. B. Houghton, Saxon, Bradburn, Ricketts, & Hardy, 2010). Sobald neue Studien publiziert werden und zur Beweislage hinzukommen, sollte auch die Benchmark zunehmend robuster werden, da sie weniger anfällig für den Einfluss von einer einzelnen Studie ist und zunehmend engere Konfidenzintervalle hat. Hier spielen jedoch – wie bei Meta-Analysen – auch Probleme wie z. B. das „Äpfel und Birnen-Problem“ eine Rolle (vgl. Eysenck, 1978; Purgato & Adams, 2012; Sharpe, 1997).

Als zusätzliche Vergleichsbedingung haben Minami et al. (2007) eine Benchmark für den natürlichen Verlauf der Verbesserung für depressive Störungen auf der Basis von meta-analytischen Befunden (vgl. Posternak & Miller, 2001) entwickelt. Dieser Wert kann als Baseline für keine Behandlungsbedingung verwendet werden, sodass jede Art von Intervention diese Schwelle überschreiten sollte, wodurch Behandlungen unter Routinebedingungen als (a) nicht effektiv, (b) effektiver als keine Behandlungsbedingungen aber weniger effektiv als eine angemessene Vergleichsbehandlung unter optimalen Bedingungen, oder (c) äquivalent zu einer angemessenen Vergleichsbehandlung unter optimalen Bedingungen beurteilt werden kann. Dieser Ansatz könnte auf andere psychische Störungen ausgeweitet werden. Daher benötigt man Benchmarks für spezifische Messinstrumente und spezifische klinische Bedingungen (Minami et al., 2007).

Die Ergebnisse von einigen solcher Studien zeichnen ein heterogenes Bild und deuten tendenziell daraufhin, dass Psychotherapie in der Routineversorgung wirksam ist, aber in den meisten Fällen hinter den Effekten aus klinischen Studien zurückbleibt. Die psychologische Behandlung, die in routinemäßigen Kliniken durchgeführt wird, ist bei einer Reihe von Patient\*innenproblemen wirksam, insbesondere bei häufigen psychischen Gesundheitsproblemen mit einer unkontrollierten Effekstärke von  $ES = 1.29$  (95% CI=1.26–1.33,  $N = 10.842$ ) (Cahill, Barkham, & Stiles, 2010). Eine frühe Studie kann zwar zeigen, dass die Ergebnisse von RCTs der Kognitiven Therapie der Depression sich auf ambulante Praxen generalisieren lassen (Persons, Bostrom, & Bertagnolli, 1999), wobei diese Studie an einer kleinen Stichprobe ( $N=45$ ) durchgeführt wurde. Eine neuere Studie zur klinischen Wirksamkeit (im Sinne der „effectiveness“) der KVT der Depression zeigt vergleichbare Effekte in der Routineversorgung, wobei die Effizienz im Zusammenhang mit der Anzahl an Sitzungen noch Optimierungspotenzial bietet (Lutz et al., 2016). Ein Vergleich zwischen praxisbasierten Daten und einer Auswahl an klinischen Studien zur Behandlung von Depressionen kommt jedoch zu dem Ergebnis, dass das Outcome in klinischen Studien 12% höher in Bezug zur Prä-Postbehandlungsveränderung ist als in praxisbasierten Studien (Barkham et al., 2008). Eine solche durchaus reliable Differenz zwischen den Effekten von psychotherapeutischer Routineversorgung (sog. „true-to-life methods“) und Psychotherapie in klinischen Studien (sog. „laboratory methods“) spricht gegen die Generalisierbarkeit von den Ergebnissen der Psychotherapien in klinischen Studien und für naturalistische Evaluationen zur Unterstützung der Forschung unter kontrollierten Bedingungen (Wampold & Imel, 2015). Einer

internationalen Vergleichbarkeit zwischen Benchmarks mangelt es jedoch aus diversen Gründen (Barkham et al., 2010), sodass wenig gute Vergleichsbenchmarks vorhanden sind (Barkham & Margison, 2008). Es gibt diverse Probleme, die das Benchmarking erschweren:

Ein Problem bei solchen Vergleichen ist die Tatsache, dass die ES-Berechnung in RCTs mit eingeschränkter Varianz bzw. Standardabweichungen erfolgt, da viele Patient\*innen in solchen Studien erst mit einem Mindestwert auf dem primären Outcomemaß eingeschlossen und sehr schwer erkrankte Patient\*innen durch z. B. Komorbiditäten ausgeschlossen werden, wodurch größere ES erzielt werden (Barkham et al., 2008). Ebenso wichtig für die Behauptung der Vergleichbarkeit könnten Äquivalenzanalysen (Rogers, Howard, & Vessey, 1993) bzw. Tests zur Signifikanz von Unterschieden sein, welche oft fehlen (z.B. Wade et al., 1998), wobei dabei ebenso diverse Probleme existieren (z. B. unterschiedliche Länge der Therapie, anderes Outcome in internationalen klinischen Studien). Signifikanztestungen von Benchmarks sind daher schwierig, da bei sehr großen Stichproben ein Unterschied zwischen Benchmarks immer signifikant werden kann (Minami, Wampold, et al., 2008). In den Studien existiert zudem oft ein methodischer Pluralismus: Die Messreliabilität ist jedoch methodenabhängig, da die Messung des Outcomes variieren kann in Bezug zur Reaktivität und Spezifität. Für Benchmarks wäre daher ein einheitliches Maß geeignet, wodurch diese Benchmark zwar instrumentenspezifisch aber auch vergleichbar wäre. Eine einheitliche Messung würde dadurch den Vergleich von Gesundheitsdienstleistungen über verschiedene Gesundheitsversorgungssysteme im internationalen Vergleich ermöglichen (Barkham et al., 2010). In solchen naturalistischen Settings eignet sich die sog. Risikoanpassung bzw. Fall-Mix-Anpassung zur Generierung von robusteren Aussagen – ein Prozess, der Variablen verwendet, die mit dem Therapieoutcome korrelieren, um das Outcome statistisch zu korrigieren (Kraus et al., 2016). Ein Problem dabei ist, dass solche Variablen nur unzureichend stabil bisher identifiziert wurden. Die Identifikation von solchen Variablen unter Routinebedingungen anhand von großen Datensätzen können replizierbare Variablen hervorbringen, welche wiederum in zukünftigen klinischen Studien für diese Prozeduren berücksichtigt werden können. Im Rahmen von Stratifikationsprozessen und multivariaten statistischen Kontrollen könnten Ergebnisse zur Wirksamkeit von klinischen Studien so robuster werden (Steyerberg, 2019).

Die Studien 1-4 der vorliegenden Dissertation sollen aus diesen genannten Gründen als interne Benchmark für die stationäre Psychotherapie in psychosomatischen Kliniken in Deutschland dienen und eher deskriptive Vergleiche zu ähnlichen nationalen Studien ziehen anstatt gegen Benchmarks zu testen, da aufgrund des unterschiedlichen und international fast einzigartigen Settings, i.e. stationäre Psychotherapieprogramme, nicht genügend Vergleichsbedingungen existieren, bei denen die Generierung einer Vergleichsbenchmark sinnvoll wäre.

### **2.2 Praxisbasierte Evidenz als Ergänzung zur Evidenzbasierten Praxis: Rolle von Praxisorientierter Forschung**

Es gibt mehrere Trends in der Psychotherapieforschung und -praxis, die über die Bereitstellung evidenzbasierter Behandlungen und Behandlungspraktiken hinaus für Patient\*innen als auch Behandler\*innen sehr vielversprechend und nützlich zu sein scheinen. Zwei unterschiedliche, aber potenziell überlappende Forschungsparadigmen bergen dieses Potenzial: Patient\*innenorientierte Forschung (Sunderji, Nicholas Angl, Polaha, & Gao, 2019) und Praxisorientierte Forschung (Lambert, 2013), wobei insbesondere die letztere für die vorliegende Dissertation relevant ist.

Studien mit größeren und überregional repräsentativen Stichproben können die Basis für einen wertvollen Austausch zwischen Forschung und Praxis legen, welche den sog. „Research-Practice-Gap“ (siehe Abschnitt 2.3.6.) verringern können. Ziel für Therapeut\*innen kann es sein, ihre Entscheidungen in der praktischen Implementierung und Umsetzung ihrer Therapie auf wissenschaftlich abgeleiteten Kriterien zu gründen, die wiederum aus der Praxis generiert wurden. Diese Forschung wird Praxisorientierte Forschung genannt (Castonguay, Barkham, Lutz, & McAleavey, 2013) und ist ein Forschungsparadigma (Castonguay & Muran, 2016), welches sich in drei Hauptansätze unterteilt: Patient\*innenfokussierte Forschung, Praxisbasierte Evidenz und Praxis-Forschungsnetzwerke (Castonguay et al., 2013).

Patient\*innenfokussierte Forschung nutzt Methoden zur Modellierung des erwarteten Ansprechens der Patient\*innen auf eine Therapie im Verhältnis zum tatsächlichen Ansprechen auf die Therapie und meldet diese Informationen an Therapeut\*innen, Supervisor\*innen und den Patient\*innen zurück (Lambert, 2013). Diese Forschung verwendet klinisch-bedeutsame Veränderungskriterien, wie z. B. das Konzept der klinischen Signifikanz (Jacobson, Follette, &

Revenstorf, 1984; Jacobson & Truax, 1991) für die individuelle Patientin, um der Therapeutin zurückzumelden, wenn die Patientin eine reliable Verbesserung zeigt. Dabei werden Outcome-Management-Forschungsstrategien genutzt, um der Therapeutin formell zu helfen, das Therapieoutcome der Patientin zu monitoren, sodass Anpassungen der Behandlungsplanung in Echtzeit möglich ist. So wird es ermöglicht, dass die Empirie ein Teil der Routinepraxis auf einer Fall-zu-Fall-Basis darstellt, was aus verschiedenen Gründen wichtig zu sein scheint. Im Sinne des Qualitätsmanagements kann es sich langfristig aus Kosten-Effektivitätsgründen lohnen, solche Strategien in der Routineversorgung zu implementieren (Lambert, 2013). Es hat sich gezeigt, dass Therapeut\*innen in der Regel nicht zuverlässig einschätzen können, wenn eine Patientin nicht oder sogar negativ anspricht auf die Therapie vor Therapieende und man somit von einem schlechten klinischen Therapeut\*innenurteil ausgehen kann (Hannan et al., 2005; Hatfield, McCullough, Frantz, & Krieger, 2010; Hill, Nutt-Williams, Heaton, Thompson, & Rhodes, 1996; Hill, Thompson, Cogar, & Denman, 1993; Regan & Hill, 1992; Rhodes, Hill, Thompson, & Elliott, 1994). In einer Studie hat sich eine Verzerrung in der Selbsteinschätzung gezeigt: 90% aller Fachkräfte im psychischen Gesundheitssektor sahen ihre Fähigkeiten im Vergleich zu ihren Kolleg\*innen unter den besten 25%, dabei schätzte sich selbst keiner der Teilnehmenden als unter dem Durchschnitt liegend ein (Walfish, McAlister, O'Donnell, & Lambert, 2012). Darüber hinaus neigen die Therapeut\*innen im Vergleich zur Empirie dazu, ihre Verbesserungsraten bei den Patient\*innen zu überschätzen und ihre Verschlechterungsraten bei den Patient\*innen zu unterschätzen. Gemäß aktueller Daten beträgt laut Empirie die Nonresponse-Rate über verschiedene Störungen hinweg ca. 50% (Westen & Morrison, 2001), ca. 10% der Patient\*innen zeigen negative Effekte (Lilienfeld, 2007), wovon nur ca. 20-30% der negativen Effekte und Verschlechterungen erkannt werden (Hatfield et al., 2010). Solche Zahlen deuten auf massive Folgen hin, wie z. B. geringere Erwartungen an eine Psychotherapie als Wirkfaktor, Therapieabbrüche, Chronifizierung und in der Folge Behandlungsresistenz. Studien zeigen, dass das Feedback von Fortschritts- bzw. Verlaufsmessungen in Rahmen von entwickelten Instrumenten hilfreich sein kann und solche Methoden – auch: klinische Unterstützungstools – das Therapieoutcome signifikant erhöhen, indem bspw. Verschlechterungsraten bei prädizierten Therapiemisserfolgen durch die Verwendung von Warnsignalen an Therapeut\*innen und die Unterstützung durch Problemlösestrategien reduziert werden (Delgadillo et al., 2018, 2017; Harmon, Hawkins, Lambert, Slade, & Whipple, 2005; Lambert et al., 2001; Lambert & Harmon, 2018; Lambert &

Shimokawa, 2011; Lambert et al., 2018; Shimokawa, Lambert, & Smart, 2010; Simon, Lambert, Harris, Busath, & Vazquez, 2012). Solche klinischen Unterstützungstools unterstützen die klinische Arbeit durch Technologie – wie empfohlen (Holmes et al., 2018) – und können zum Monitoring (Bar-Kalifa et al., 2016), Feedback (Lambert, 2017), Personalisierung (z. B. automatisierte Bereitstellung von Therapiematerialien), Prognostik und letztlich zur Evaluation im Sinne des Qualitätsmanagements eingesetzt werden. Allein das regelmäßige Monitoring und dadurch die Entdeckung von Schwankungen in der Symptombelastung der Patientin kann zu einem besseren Therapieergebnis führen (Bar-Kalifa et al., 2016). Die Implementierung solcher Tools stößt jedoch noch zuweilen auf Widerstand unter Therapeut\*innen (Jensen-Doss & Hawley, 2010). In Deutschland wurde dieser Forschungsansatz in der Trierer Arbeitsgruppe implementiert, welche den Trierer Therapienavigator (TTN) entwickelt haben (Lutz, De Jong, & Rubel, 2015; Lutz, Rubel, et al., 2019). Dieser ermöglicht es, sowohl personalisierte Behandlungsempfehlungen (z. B. Vorhersage des Therapieabbruchrisikos, Vorhersage und Auswahl der optimalen Behandlungsstrategie) als auch personalisierte Behandlungsanpassungen (z. B. Erkennen von negativen Therapieverläufen, klinisches Unterstützungstool) vorzunehmen. Die aktuelle Forschung, die auf dem Monitoring des Ansprechens der Behandlung der Patient\*in und Alarmsignalen nach jeder Behandlungssitzung beruht, kann eine leistungsfähige Methode zur Reduktion von negativen Therapieverläufen bzw. Verschlechterungsraten und Veränderungen sein und das Outcome durch ein zeitiges Feedback an die Behandler\*innen erhöhen (z. B. effektiv für „not-on-track“-Patient\*innen, Reduktion der Therapielänge bei „on-track“-Patient\*innen).

Unter praxisbasierter Evidenz versteht man die gewissenhafte, explizite und vernünftige Nutzung aktueller Erkenntnisse, die in praktischen Settings gewonnen wurden, bei Entscheidungen über die Versorgung individueller Patient\*innen (Barkham et al., 2010; Barkham & Margison, 2008; Holmqvist, Philips, & Barkham, 2015). Solche Evidenz rührt dabei aus der Integration sowohl individueller klinischer Expertise als auch von Parametern auf Versorgungsebene mit den besten verfügbaren Erkenntnissen aus rigorosen Forschungsaktivitäten, die in klinischen Routineumgebungen gewonnen werden. Praxisbasierte Forschung untersucht, wie und welche Behandlungen der individuellen Patientin im Versorgungssystem angeboten werden und evaluiert, wie man auf dem Kliniklevel die Behandlung und Versorgung noch verbessern kann (Castonguay et al., 2013). Das Ziel ist vor allem die Variation in der Versorgung zu untersuchen und Wege aufzutun, wie man

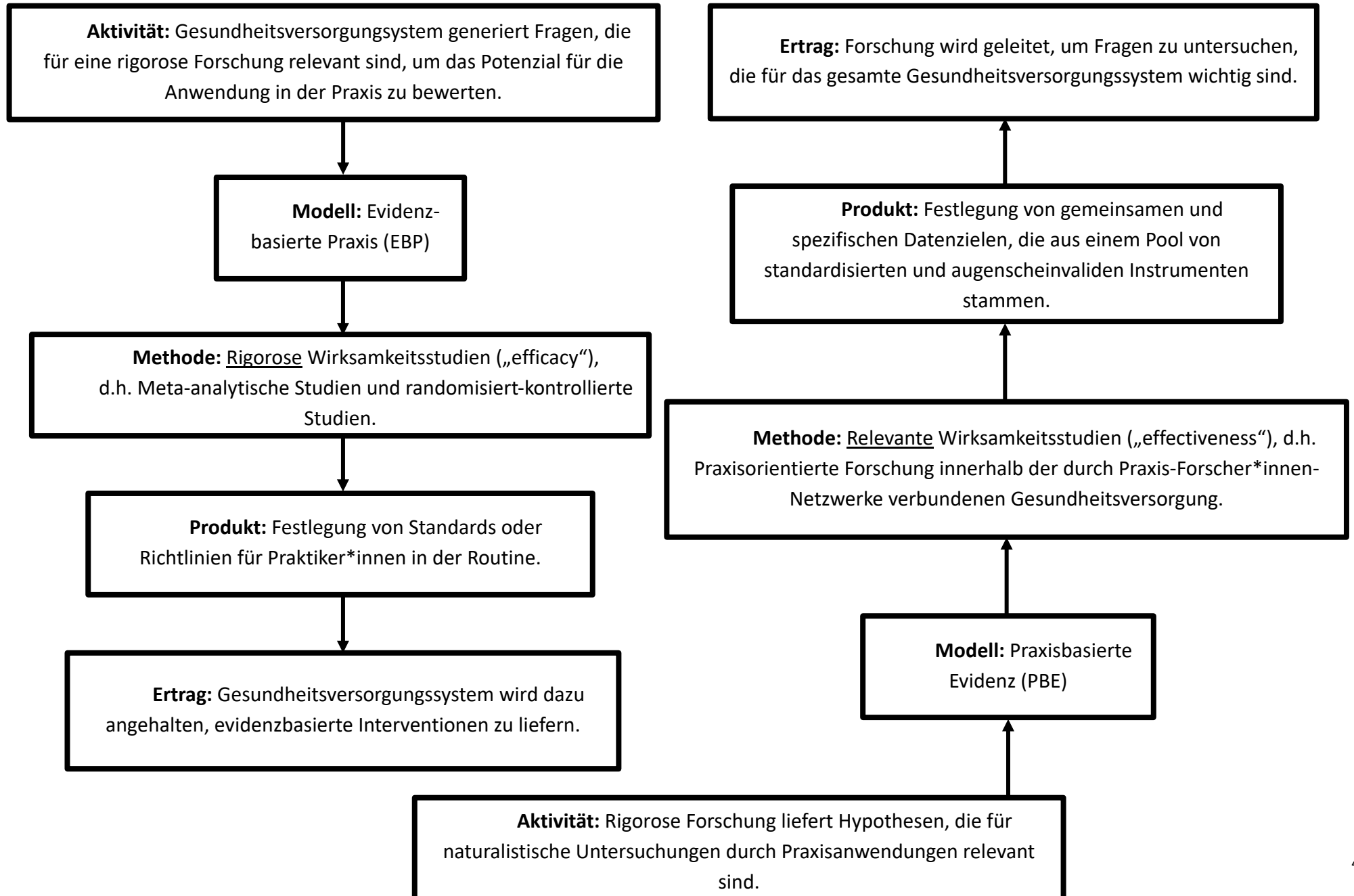
forschungsbasierte Behandlungen implementieren kann. Durch systematische Datenerhebungen in Versorgungseinrichtungen können große Datensätze generiert werden, die auf unterschiedliche Weise ausgewertet werden: z. B. die Ermittlung von Benchmarks über Versorgungseinrichtungen hinweg und innerhalb einer Einrichtung über die Zeit hinweg bis hin zur Identifikation von Behandlungen, die wirken bzw. nicht wirken (z. B. Stiles, Barkham, Mellor-Clark, & Connell, 2008). Eine in diesem Sinne wichtige Studie weist darauf hin, dass das Problem von Ausschlusskriterien bei klinischen Studien überschätzt sein könnte, indem gezeigt werden konnte, dass Patient\*innen mit Zwangsstörungen, die in RCTs aufgrund diverser Gründe ausgeschlossen wurden, mit ähnlichen Effekten durch KVT behandelt werden konnten (Franklin et al., 2000). Eine weitere aktuelle Studie zeigt, dass KVT in einem großen Langzeit-Follow-up von 5-20-Jahren nach einer ambulanten Routinebehandlung nachhaltig wirkt, insbesondere im Vergleich zur medikamentösen Therapie (von Brachel et al., 2019). Diese enge Kooperation und Vernetzung zwischen Praktiker\*innen und Forscher\*innen kann im Sinne eines Praxis-Forschungsnetzwerkes ein großes Potential für die Implementierung von evidenzbasierten Verfahren bilden (Lutz, Neu, & Rubel, 2019). Im Rahmen der Praxisorientierten Forschung können daher naturalistische Studien sowohl die Praxisbasierte Evidenz als auch Praxis-Forschungsnetzwerke fördern (Castonguay et al., 2013). Dieses Forschungsparadigma kann somit als wichtige Ergänzung zur evidenzbasierten Forschung betrachtet werden.

Das prominenteste Beispiel zur Dissemination und Implementierung von evidenzbasierten psychologischen Therapien in der Versorgung ist das Programm „Improving Access for Psychological Therapy (IAPT)“ aus dem Vereinigten Königreich (Clark, 2018; Clark et al., 2018), welches beispielhaft für Praxisorientierte Forschung genutzt werden und somit Praxisbasierte Evidenz generieren kann. Das Vereinigte Königreich hat das IAPT-Programm ins Leben gerufen, welche die Verbreitung von evidenzbasierter Psychotherapie (EBP: v. a. KVT, aber auch IPT, Paartherapie, und kurze psychodynamische Therapie) innerhalb des nationalen Gesundheitssystems ausdehnt, um so in großem Maße die Therapie für Depression und Angststörungen verfügbar zu machen (Clark, 2011). Das nationale Rollout beinhaltet die Finanzierung von Trainings und ein standardisiertes nationales Trainingscurriculum, ein Versorgungsmodell mit allgemeinen Standards aber einem gewissen Grad an lokaler Autonomie und das Monitoring von klinischen Outcomes. Im Rahmen eines „Stepped-Care“-Modells werden die Patient\*innen aufgrund des Schweregrads und Behandlungsresponse in hoch-intensive (z. B. Face-to-Face-Behandlung) und niedrig-intensive Behandlungen (z. B.



Telefon- und internetbasiert-geleitete Selbsthilfe) eingeteilt. Die Outcomedaten von jedem IAPT-Service im ganzen Land werden vierteljährlich veröffentlicht und zur Verfügung gestellt. Durch die Nutzung von solchen großen Routinedatensätzen kann es gelingen, die Entscheidungsfindung bzgl. einer Behandlung zu verbessern. Im Kontext von IAPT wurden bereits viele weitere vielversprechende Ergebnisse publiziert, im Zusammenhang mit dem Therapieerfolg (Gyani et al., 2013), den Kosten (Radhakrishnan et al., 2013) und der Kosten-Effektivität (Zala et al., 2019).

In Abbildung 5 ist die wechselseitige Beziehung zwischen Evidenzbasierter Praxis und Praxisbasierter Evidenz gemäß Barkham & Mellor-Clark (2003) dargestellt. Daher soll im Rahmen eines praxisorientierten Forschungsparadigmas in den Studien der vorliegenden Dissertation kurzfristig praxisbasierte Evidenz generiert und langfristig das Praxis-Forschungsnetzwerk gestärkt werden.



### 2.3 Probleme der Psychotherapieforschung

Neben der Verwendung eines geeigneten Forschungsparadigmas existieren durch den starken Fokus auf Phase III-Studien Probleme in der Psychotherapieforschung. Einige der Probleme, die im Zusammenhang mit der vorliegenden Dissertation stehen, werden im Folgenden vertiefend aufgegriffen.

#### 2.3.1. Definition von Psychotherapieerfolg: Wahl eines geeigneten Therapieoutcomes und Kriterien für klinisch-bedeutsame Veränderungen

Die Frage, ab wann eine Therapie erfolgreich ist, hängt von klinisch-inhaltlichen und statistischen Kriterien ab, die in der Forschung unterschiedlich verwendet werden. Dadurch unterscheiden sich auch die Ergebnisse, was eine Vergleichbarkeit zwischen Studien und die Entwicklung von universellen Standards zur Evaluation von Psychotherapie erschwert.

Den Behandlungserfolg kann man auf verschiedenen klinischen-inhaltlichen Ebenen abbilden (Caspar & Jacobi, 2007):

- (1) Symptomebene
- (2) Allgemeine psychopathologische Gesamtbelastung
- (3) Ebene der Störungsfolgen (Funktionalität, Lebensqualität)

Diese Definition ist angelehnt an die Arbeit von Schulte (1993), der die drei Ebenen Symptom- und Beschwerderückgang, Störungsursache und Störungsfolgen etablierte. Die Frage nach den Zielen und Outcomes in der Psychotherapie bleibt aber offen, so wird aktuell vorgeschlagen, dass insbesondere die Perspektive der Patient\*innen bei der Wahl eines Outcomes intensiver mit einbezogen werden sollte (Barber & Solomonov, 2019; Cuijpers, 2019). Die Wahl eines prognostischen Outcomes zur Bestimmung des Psychotherapieerfolgs sollte dabei durch den Kontext geleitet sein, wobei das Outcome so reliabel wie möglich gemessen werden sollte (Steyerberg, 2019). Bei der Wahl sollten Kriterien wie Validität, Veränderungssensitivität und praktische Kriterien unter dem Schirm der klinischen Nützlichkeit (d.h. kurz, kostengünstig, unkomplizierte Durchführung, Auswertung und Interpretation) berücksichtigt werden. Hingegen sollten jedoch auch die Outcomes über klinische Studien hinweg standardisiert werden, um die Vergleichbarkeit zu erhöhen.

Ebenso spielen auch statistische Überlegungen eine Rolle, ab wann eine Psychotherapie klinisch-signifikante bzw. klinisch bedeutsame Veränderungen erzeugt. Die absolute Wirksamkeit einer Psychotherapie kann über klinisch-bedeutsame Veränderungen untersucht werden. Dafür existieren verschiedene Methoden (Lambert & Bailey, 2012): In der Forschung haben sich sowohl Prä-Post-Vergleiche als absolutes Maß der Veränderung (z. B. durch Angabe der Mittelwertsveränderungen und in Form von Effektstärken), ‚Response‘ als Maß für das Ansprechen auf die Behandlung bzw. ‚Early Response‘ als frühzeitiges Ansprechen während der ersten Therapiesitzungen sowie Remission als Symptombefreiheit, i.e. bei Werten außerhalb des als pathologisch definierten Wertebereichs einer Skala, bewährt (Hiller, Bleichhardt, & Schindler, 2009; Hiller & Schindler, 2011). Ein weiteres Kriterium ist Therapieabbruch, welches z. B. standardmäßig bei Studien zur Evaluation der DBT bei Borderline-Persönlichkeitsstörungen gewählt wird (Panos et al., 2014). Auch diese Kriterien können unterschiedlich operationalisiert sein und daher zu unterschiedlichen Ergebnissen in der gleichen Studie und Stichprobe führen. Zur Einordnung von Response und Remission existieren der *Reliable Change Index (RCI)* nach Jacobson & Truax (1991) – als Maß der absoluten Veränderung – und die *Methode der prozentualen Symptomreduktion bzw. Verbesserung* – als Maß der relativen Veränderung –, welche beide Vor- und Nachteile aufweisen (Hiller & Schindler, 2011; Hiller, Schindler, & Lambert, 2012). Zur Beurteilung der absoluten Veränderung wurde ein zweiteiliges Kriterium vorgeschlagen: Patient\*innen zeigen eine statistisch reliable Verbesserung als Ergebnis der Behandlung anhand des RCI (Jacobson & Truax, 1991) und die behandelten Patient\*innen sind statistisch nicht unterscheidbar von einer „gesunden“ Vergleichsgruppe nach der Behandlung (Kendall, Marrs-Garcia, Nath, & Sheldrick, 1999). Auch wenn eine Studie darauf hindeutet, dass der RCI verwendet werden soll (Bauer, Lambert, & Nielsen, 2004), so hat sich im Vergleich zwischen RCI, Effektstärken (ES) und Standardmessfehler gezeigt, dass der RCI die konservativste Schätzung des Outcomes ist (Eisen, Ranganathan, Seal, & Spiro, 2007). Eine aktuelle Studie hat zudem gezeigt, dass die Klassifikation eines „guten Outcomes“ anhand des RCI weder die eher pessimistischen Einschätzungen in den Outcomeerfahrungen erklärt noch die multidimensionale Natur eines „guten Outcomes“ aus Patient\*innensicht erfasst, wodurch solche Kennwerte ohnehin vorsichtig zu bewerten sind (De Smet et al., 2020). Auf der anderen Seite zeigt sich jedoch auch, dass die ES gemäß Cohen’s d zu einer Überschätzung der Effekte führen könnte (Shimokawa et al., 2010), daher müsste dafür in naturalistischen Studien korrigiert werden (z. B. mit Hilfe des sog. „Last observation carried forward“, kurz: LOCF).

Zumindest in klinischen Studien wird jedoch vom LOCF abgeraten (Siddiqui, Hung, & O'Neill, 2009). Effektstärken werden in der Literatur mit am häufigsten angegeben.

In der Literatur werden dabei u. a. auch andere Formen der Effektstärkenberechnung zur Darstellung der klinischen Signifikanz vorgeschlagen, z. B. die Anzahl der notwendigen Behandlungen (engl.: „Number needed to treat (NNT)“<sup>11</sup>), Vergleich der T- und C-Response bei der Fläche unter der ROC-Kurve, und die Unterschiede in den Erfolgsraten (Kraemer & Kupfer, 2006). Als Vorteil der NNT wird genannt, dass die – als zusätzliche nicht alternative Angabe zum Therapieeffekt – dazu dient, die Kommunikation zu Patient\*innen, zu Familien, Interessensvertreter\*innen und Entscheidungsträger\*innen zu erleichtern (Citrome, 2008). In der Literatur zu chronischen Schmerzen ist die Angabe von NNT inzwischen üblich, wobei jedoch auch hier bei der Berechnung, Verwendung und Interpretation Vorsicht geboten ist (Katz, Paillard, & Van Inwegen, 2015; Moore, 2015). Zur Interpretation wird auf folgendes Paper verwiesen (Andrade, 2015). Sie ist insgesamt in RCTs zu Psychotherapien bisher selten berichtet (Shearer-Underhill & Marker, 2010).

Eine weitere Überlegung, die von einigen Forscher\*innen angestellt wird, ist der Einbezug von Outcomes, die von mehreren Interessensvertreter\*innen wichtig sind, wie z. B. Wartezeit, Festlegung eines Behandlungsplans mit gemeinsam vereinbarten Zielen oder Zufriedenheit mit der Psychotherapie (Barkham et al., 2010). Eine weitere Methode zur Operationalisierung klinisch-bedeutsamer Veränderungen ist das Erfüllen einer Diagnose, d.h. wenn Personen, die früher die diagnostischen Kriterien für eine psychische Störung erfüllten, dies nach der Behandlung nicht mehr tun. Diese Methode birgt jedoch aufgrund der arbiträren Natur der Diagnosekriterien der unterschiedlichen klassifikatorischen Diagnosesysteme Nachteile (Roth & Fonagy, 2005; Westen & Morrison, 2001). Auch die Kostenverrechnung könnte berücksichtigt werden, d.h. der Einfluss auf das reale Leben mitsamt ökonomischer Konsequenzen durch Psychotherapie kann die Bedeutung für die Patient\*in, Therapeut\*in und die Gesellschaft herstellen (Lambert, 2013). In diesem Kontext wird auch die Frage wichtig, wie Patient\*innen die Erfolge aufrechterhalten können. Weitere wichtige Fragestellungen betreffen die optimale Therapielänge (auch: Therapiedosis) (Lutz, Neu, et al., 2019), was sich z. B. im Rahmen einer datengetriebenen empirischen Analyse der Dosis-Wirkungskurve realisieren lässt (Falkenström, Josefsson, Berggren, & Holmqvist, 2016; Hansen, Lambert, &

---

<sup>11</sup> <https://ktclearinghouse.ca/> bzw. <https://ebm-tools.knowledgetranslation.net/>

Forman, 2006; Harnett, O'Donovan, & Lambert, 2010; Howard, Kopta, Krause, & Orlinsky, 1986; Robinson, Kellett, & Delgado, 2020). Ein aktuelles systematisches Review stellt fest, dass eine kurvilineare (bzw. logarithmisch-lineare oder kubische) Beziehung zwischen der Behandlungsdauer und dem Outcome konsistent gefunden wird mit Ausnahme von Essstörungen und schweren psychiatrischen Erkrankungen: die optimale Dosis der Psychotherapie in Routinesettings liegt dabei zwischen 4 und 26 Sitzungen (4-6 Sitzungen für die geführte Selbsthilfe bei niedriger Intensität) und variiert je nach Setting, klinischer Population und Outcomemaßen (Robinson, Delgado, & Kellett, 2020). In diesem Zusammenhang sind ‚Sudden Gains‘ und verwandte Veränderungsmuster relevant. Ein Sudden Gain ist ein dramatischer Behandlungseffekt im Kontrast zur Sitzung-zu-Sitzungs-Veränderung. In der Literatur finden sich Begriffe wie „Sudden Gains“ (Collins & Coles, 2017; Lutz et al., 2013; Tang & DeRubeis, 1999), „Rapid Response“ (Fennell & Teasdale, 1987), „Early Response“ (Ilardi & Craighead, 1994, 2006; Tang & DeRubeis, 2006), „Large Sudden Improvements“ (Gaynor et al., 2003), „Rapid Early Response“ (Haas, Hill, Lambert, & Morrell, 2002), „Early Sudden Gains“ (Stiles et al., 2003).

Insgesamt kann daher entweder ein kontinuierliches, kategoriales oder dichotomes bzw. binäres Outcome gewählt werden. Ein dichotomes Outcomemaß (z. B. reliabel verbessert vs. nicht reliabel verbesserte Patient\*innen) ist nötig, um beispielsweise Prädiktoren für Non-Response zu identifizieren. Auch wenn binäre Outcomes aufgrund der einfachen und augenscheinlichen Einordnung und Interpretation der Forschungsergebnisse sehr populär in der Literatur sind, ist von einer statistischen Perspektive jedoch ein kontinuierliches Outcome (z. B. Angabe von Effektstärken) zu bevorzugen, da diese in den statistischen Analysen mehr Power haben - ebenso wie geordnete Outcomes (Steyerberg, 2019).

Die Wahl eines geeigneten Outcomes für eine gegebene Fragestellung sollte daher sowohl diese klinisch-inhaltlichen als auch methodisch-statistischen Überlegungen mit einbeziehen. Auch wenn Prädiktionsmodelle mit pragmatisch definierten Prädiktoren entwickelt werden, um zukünftig genutzt zu werden, so sollte das Outcome mit ähnlicher Strenge wie in klinischen Studien bestimmt werden, da zukünftige Entscheidungen auf der Vorhersage von diesem Modell abhängt und diese Vorhersagen daher auf robusten statistischen Zusammenhängen mit einem a priori bestimmten Outcome erfolgen soll (Steyerberg, 2019).

### 2.3.2. Absolute vs. Relative Wirksamkeit: Dodo-Bird-Verdict oder die Frage nach spezifischen und allgemeinen Effekten

Einige Autor\*innen argumentieren, dass in der Psychotherapieforschung die Unterscheidung von relativer und absoluter Wirksamkeit zur Einordnung und Interpretation von Studienergebnissen zwingend benötigt wird (Munder et al., 2019).

Die absolute Wirksamkeit wird darüber festgestellt, ob eine Psychotherapie gegenüber keiner Behandlung wirksam ist. Durchschnittliche Effektstärken von Psychotherapie gegenüber keiner Behandlungen belaufen sich auf  $ES = 0.80$  über verschiedene psychische Störungen hinweg (Wampold & Imel, 2015). Kritische Forscher\*innen zweifeln jedoch aufgrund von fundamentalen methodischen Problemen die Höhe der Effekte an und berichten beispielsweise kleine Effekte für Psychotherapie bei depressiven Störungen (Cuijpers et al., 2019), wobei diese Befunde intensiv diskutiert werden (Cuijpers, Karyotaki, Reijnders, & Ebert, 2019; Munder et al., 2019b). Im Kontext dieser Psychotherapieforschung wurden Begriffe definiert wie empirisch-gestützte Behandlungen, evidenzbasierte Behandlungen oder psychologische Behandlungen mit Forschungsunterstützungen. Ursprünglich wurden sie als empirisch-validierte Behandlungen bezeichnet. Wenn man bedenkt, dass die Effektstärken zumindest für die Psychotherapie depressiver Störungen aufgrund von methodischen Problemen wohl überschätzt sein könnten (Cuijpers et al., 2019; Cuijpers, van Straten, Bohlmeijer, Hollon, & Andersson, 2010; Cuijpers et al., 2013), ergibt sich erheblicher Optimierungsbedarf.

Die relative Wirksamkeit hingegen wird darüber festgestellt, ob es Unterschiede zwischen Psychotherapieverfahren gibt bzw. geht der Frage nach, ob ein Psychotherapieverfahren einer anderen Psychotherapie überlegen ist. Hier werden daher differentielle Effekte von Psychotherapie untersucht. Eine Netzwerk-Metaanalyse konnte zeigen, dass KVT die am meisten getestete Therapie für depressive Störungen darstellt (also über eine robuste Evidenz verfügt), die Unterschiede zwischen Therapieverfahren hingegen gering ausfallen (Barth et al., 2013). Einige andere meta-analytische Studien deuten hingegen auf eine Überlegenheit der KVT insbesondere bei Depressionen und Angststörungen hin (Marcus, O'Connell, Norris, & Sawaqdeh, 2014; Mayo-Wilson et al., 2014; Tolin, 2010). Gleichfalls zeigen andere Studien jedoch auch, dass die Unterschiede zwischen KVT und anderen Therapien bei ca.  $.20$  liegen und nicht signifikant sind (Baardseth et al., 2013; Wampold et al., 2017), wobei dieser Befund auch auf methodische Mängel und Fehlervarianz zurückgeführt werden kann (Tolin, 2014, 2015).

Die Ergebnisse einer groß angelegten Vergleichsstudie zwischen psychoanalytischer und kognitiv-behavioraler Langzeitbehandlung für chronisch depressive Patient\*innen zeigen für beide Psychotherapieformen große Effekte ohne signifikante Unterschiede (Leuzinger-Bohleber, Hautzinger, Fiedler, et al., 2019; Leuzinger-Bohleber, Hautzinger, Keller, et al., 2019), wobei deren vergleichbare Wirksamkeit aufgrund von methodischen Problemen, signifikanten Abweichungen vom Studienprotokoll sowie starken Dosis-Response-Unterschieden kritisch diskutiert wird (Hautzinger et al., 2019; Kaiser et al., 2019). Im Hinblick auf die relative Wirksamkeit impliziert eine aktuelle Übersichtsarbeit, dass wohl die Kombination aus Psychotherapie und pharmakotherapeutischer Behandlung die beste Wahl für Patient\*innen mit einer moderaten Depression ist (Cuijpers et al., 2020).

Insgesamt herrscht in der Forschung bzgl. der relativen Wirksamkeit aus diversen Gründen (z. B. Befürworter\*innen einer einzelnen Psychotherapieschule) eine kontroversere Diskussion als über die allgemeine Wirksamkeit, die auch als Schulenstreit bezeichnet wird. Die Debatte über die relative Wirksamkeit wird oft unter dem Deckmantel des *Dodo-Bird-Verdicts* oder „Äquivalenzparadoxon“ in der Psychotherapieforschung geführt (de Felice et al., 2019; Marcus et al., 2014; Wampold & Imel, 2015; Wampold et al., 1997). Bereits Rosenzweig (1936) behauptete, dass alle Therapiemethoden, die kompetent angewendet werden, gleich erfolgreich sein können. Er führte die allgemeinen Effekte von Psychotherapie auf den allen Therapien zugrundeliegenden allgemeinen Wirkfaktoren zurück – ein Phänomen, was im Anschluss der Dodo-Bird-Effekt genannt wurde. Die Gegenseite, Vertreter\*innen einzelner Psychotherapieschulen, glauben jedoch, dass einige Verfahren, die spezifische wissenschaftlich fundierte Elemente enthalten, wirksamer sind als andere. Die wissenschaftliche Analyse dieser Diskussion und der Etablierung von relativer Wirksamkeit benötigt komparative Outcomestudien (Kazdin, 1994) im Sinne von Superior bzw. Non-inferior-Studien. Die Fragen nach einem geeigneten Komparator – also eine geeignete Vergleichsbedingung – verbleibt, die meisten Studien verwendeten Wartelistenkontrollgruppen, TAU (manchmal auch „usual care“ genannt: UC) oder Supportive Therapie (Wampold et al., 2011). Im englischsprachigen Raum werden folgende Therapien als „Placebo-Psychotherapiebedingungen“ gehandelt, woran man die Schwierigkeiten bei der Bestimmung einer passenden Kontrollbedingung erkennt: „alternate treatment“, „nonspecific treatment“, „attention control“, „minimal treatment“, „supportive counseling“, „nondirective counseling“, und die „supportive therapy“. Ein geeigneterer Komparator wäre allerdings hingegen eine bona fide Psychotherapie, die durch



den jeweiligen Vertreter\*in unter Kontrolle der Behandlungsallianz stattfindet. Die erste größere Studie in diesem Rahmen war das „Treatment of Depression Collaborative Research Program“ (Elkin, 1994). Oft werden bei der Dodo-Bird-Debatte dabei allgemeine und spezifische Faktoren als voneinander unabhängige Faktoren gewichtet, was eine aktuelle Studie widerlegte (de Felice et al., 2019).

Die Diskussion über allgemeine und spezifische Wirkfaktoren scheint das Feld der Psychotherapieforschung wenig zielführend voranzubringen (Mulder, Murray, & Rucklidge, 2017). Aus Sicht einiger Autor\*innen sollte daher die weitere Forschung darauf fokussieren, wie Psychotherapie wirkt und wie der Zugang zu Psychotherapie verbessert werden kann (Munder et al., 2019). Zukünftige Studien sollten daher unabhängig von Debatten über die relative Wirksamkeit eher darauf abzielen, die vorhandenen Effekte von verschiedenen Psychotherapien noch weiter zu optimieren.

### 2.3.3. „End of average“ oder warum nicht nur Effektstärken zählen

Trotz aller Errungenschaften und breiten positiven Basis an Aussagen über Psychotherapie, die mit mehr Konfidenz als je zuvor gemacht werden kann, ist es weiterhin wichtig herauszustellen, dass gemittelte positive Effekte eine erhebliche Variabilität im Outcome maskieren können (Lambert, 2013). Zur Überwindung dieses Umstands wird populärwissenschaftlich auch oft von „End of Average“ gesprochen (Rose, 2016). Wie im vorherigen Abschnitt erwähnt (siehe Abschnitt 2.3.1.), ist die Berechnung von Effektstärken oft das Mittel der Wahl zur Interpretation der Effekte in klinischen Studien, welche die Wirksamkeit von Psychotherapie betrachten. Effektstärken geben hingegen den Effekt einer Behandlung gegenüber der mittleren Patientin an. Aus dieser Logik heraus resultieren verschiedene Probleme: Wenn man Effektstärken umrechnet, z. B. in eine Erfolgsrate von unbehandelten vs. behandelten Personen, dann eröffnet sich ein differenziertes Bild bzgl. der Wirksamkeit (Lambert, 2013). Bei einer Effektstärke von Cohens  $d = 1.0$  ist die Erfolgsrate von unbehandelten zu behandelten Personen bei .276 zu .724., der Anteil der erklärten Outcomevariabilität aufgrund der Behandlung liegt dann bei .200 und die NNT liegt bei 2. Diese Problematik wird umso deutlicher bei der Betrachtung von anderen Kennwerten als Effektstärken, d.h. Raten zum Ansprechen auf die Psychotherapie („Response-Rate“), Raten zur Symptombefreiheit („Remission-Rate“), Raten zur Restsymptomatik („Residualsymptom-Rate“), Raten zur Symptomverschlechterung („Deterioration-Rate“), Raten zum Therapieabbruch („Dropout-

Rate“), Raten zur Therapieverweigerung- bzw. ablehnung („Refusal-Rate“) und Raten zum Rückfall („Relapse-Rate“).

Bei der Untersuchung der Response- und Verschlechterungs-Raten von Psychotherapien für verschiedene psychische Störungen wird deutlich, dass diese vergleichsweise hoch ausfallen: bei vielen psychotherapeutischen Verfahren liegt die Nonresponse-Rate bei bis zu 35-40%, ca. 5-10% aller Patient\*innen verschlechtern sich (Lambert, 2013). Eine Behandlungsverweigerung und vorzeitige Beendigung sind auch Probleme der Psychotherapie (Swift, Greenberg, Tompkins, & Parkin, 2017), wobei die mittlere gewichtete Dropout-Rate von KVT vor der Behandlung bei 15.9% und während der Behandlung bei 26.2% liegt (Fernandez, Salem, Swift, & Ramtahal, 2015).

Bei Patient\*innen mit PTBS liegt beispielsweise die Nonresponse-Rate in vielen Studien über 50%, die Dropout-Raten kann je nach der untersuchten Stichprobe zwischen 0 und 54% ausfallen (Bradley, Greene, et al., 2005; Schottenbauer et al., 2008), wobei die Dropout-Rate in naturalistischen Studien höher als in klinischen Studien ist (Najavits, 2015). Bei Patient\*innen mit Zwangsstörungen liegt die mittlere Dropout-Rate der Behandlung in klinischen Studien bei etwa 15% (Eddy, Dutra, Bradley, & Westen, 2004), wobei die kognitive Therapie die niedrigste Rate (11.4%), die Expositionstherapie mit Reaktionsverhinderung (ERP) mit 19.1% eine höhere Rate und die kombinierte Therapie aus ERP und antidepressiver Medikation die höchste Rate (32.0%) aufweist (Öst et al., 2015). Die Refusal-Raten variieren je nach Studie sehr stark (Öst et al., 2015). Selbst dort, wo Patient\*innen den ‚Goldstandard‘ erhalten, scheinen etwa 30 % nicht von der Behandlung in Form von Nonresponse zu profitieren (Schruers, Koning, Luermans, Haack, & Griez, 2005). Beispielsweise treten nur bei 50% - 60% aller Patient\*innen mit Zwangsstörungen klinisch bedeutsame Veränderungen auf, während etwa 75% über Residualsymptome berichten (Fisher & Wells, 2005). Ein ähnliches Bild zeichnet sich für depressive Störungen. Ein Drittel bis die Hälfte aller Patient\*innen mit depressiven Störungen profitieren nicht von der Behandlung in Form von Non-Response (Lambert, 2017; van Bronswijk, Moopen, Beijers, Ruhe, & Peeters, 2019), die mittlere Dropout-Rate liegt für die KVT bei Depressionen vor der Behandlung bei 21.6% und während der Behandlung bei 36.4% (Fernandez et al., 2015), für jegliche Art von individueller Psychotherapie mit 17.5% etwas niedriger (Cooper & Conklin, 2015). Letztlich sind auch die Relapse-Raten nicht unerheblich. Beispielsweise erscheint die Relapse-Rate mehr als zwei Jahre nach einer Psychotherapie für

depressive Störungen mit 53.1% relativ hoch, jedoch im Vergleich zu nicht-psychotherapeutischen Behandlungen niedrig (Steinert, Hofmann, Kruse, & Leichsenring, 2014). So berichten beispielsweise etwa 65 - 80% der Patient\*innen mit depressiven Störungen nach der Behandlung (in dieser Studie: KT und IPT) immer noch über klinisch signifikante Residualsymptome (Lemmens et al., 2015). Bei Patient\*innen mit Borderline-Persönlichkeitsstörungen ist die Dropout-Rate besonders hoch (Barnicot, Katsakou, Marougka, & Priebe, 2011). In einer Studie betrug sie 37% (McMurrin, Huband, & Overton, 2010). Zählt man nur die Therapieabbrüche, die nicht in gegenseitigem Einverständnis von Patient\*in und Team erfolgten, so beträgt die Dropout-Rate allerdings nur noch ca. 10% (Kröger, Harbeck, Armbrust, & Kliem, 2013). Auch bei der Behandlung anderer psychischer Störungen sowie mit anderen Therapieformen decken sich diese Raten: Beispielsweise betrug die gewichtete mittlere Dropout-Rate für eine individuelle Psychotherapie bei Patient\*innen mit generalisierter Angststörung 16.99%, d.h. etwa jede sechste Patientin brach die Therapie ab (Gersh et al., 2017). Ebenso zeigte sich, dass die gewichtete Dropout-Rate bei Acceptance and Commitment Therapy (ACT) 15.8% betrug (Ong, Lee, & Twohig, 2018). Diese vergleichsweise hohen Raten deuten auf einen Optimierungsbedarf von evidenzbasierten Therapien hin.

### 2.3.4. Verfahrensinflation

Die Psychotherapieforschung birgt verschiedene Probleme, welche zum Teil zu einer Stagnation von Effekten in der Psychotherapie (siehe Abschnitt 2.3.5.) trotz intensiver Forschung (z. B. mangelnde Befunde zur relativen Wirksamkeit) und der Entwicklung von unzähligen neuen Psychotherapieansätzen (vgl. „Dritte Welle“ der Verhaltenstherapie) geführt haben (Churchill et al., 2013; Hunot et al., 2013). Der Zeitgeist in den letzten 20 Jahren hat in der Psychotherapieforschung neben wirksamen Behandlungsverfahren für eine Vielzahl von psychischen Störungen auch zu einer Verfahrensinflation (84 gelistete Verfahren gemäß APA) sowie einem sog. „Branding“ von Therapien (bzw. Therapieschulen) geführt (Goldfried, 2019; Rubel & Lutz, 2016). Schon Anfang der 80er Jahre stellt Goldfried fest: “The popularity of a therapy school is often a function of variables having nothing to do with the efficacy of its associated procedures. Among other things, it depends on the charisma, energy level, and longevity of the leader; the number of students trained and where they have been placed; and the spirit of the times.” (Goldfried, 1980).

Um dem Branding von Psychotherapie zu entgehen, der Komplexität von psychischen Störungen (z. B. Heterogenität der Symptome über psychische Störungen hinweg, hohe Raten an Komorbidität) besser gerecht zu werden sowie die Wirksamkeit (nicht nur für die „Durchschnittspatientin“) zu erhöhen, haben sich zwei andere Strömungen entwickelt: transdiagnostische Behandlungsprotokolle sowie individualisierte bzw. personalisierte Interventionen, die sich gegenseitig nicht ausschließen (Holmes et al., 2018). Die transdiagnostischen Behandlungsprotokolle zielen darauf ab, mit evidenzbasierten Verfahren diejenigen Mechanismen zu adressieren, die allen psychischen Störungen zugrundeliegen. Das prominenteste dieser Art ist das „Unified Protocol“ (Barlow et al., 2015) für emotionale Störungen basierend auf den Befunden der Emotionspsychologie (Boettcher, Schaeuffele, & Renneberg, 2018; Moses & Barlow, 2006; Sakiris & Berle, 2019). Als aktueller transdiagnostischer Ansatz kann die prozessbasierte Psychotherapie (Hayes & Hofmann, 2018; Hofmann & Hayes, 2019) erwähnt werden, welche die zugrundeliegenden Prozesse bzw. Mechanismen (z. B. Mediatoren und Moderatoren) für die Veränderung und letztlich Wirkung einer Psychotherapie anzielt. Die individualisierten bzw. personalisierten Psychotherapieansätze zielen darauf ab, Psychotherapien besser auf die individuellen Probleme und Verläufe anzupassen und werden insbesondere durch methodologische und methodische Errungenschaften aus anderen Disziplinen (z. B. ‚Computational Psychiatry‘) inspiriert (siehe Abschnitt 2.4).

### 2.3.5. Stagnation der Effekte von Psychotherapie

Die Effekte von Psychotherapie scheinen über die letzten Jahrzehnte eher stagniert zu haben, was auch daran liegen mag, dass neben spezifischen Effekten durch die Therapie auch andere Variablen eine Rolle beim Outcome einer Psychotherapie zu spielen scheinen (Wampold & Imel, 2015): Allgemeine Effekte wie die therapeutische Beziehung (Flückiger, Del, Wampold, & Horvath, 2018; Friedlander, Escudero, Welmers-Van De Poll, & Heatherington, 2018; Horvath, Del Re, Flückiger, & Symonds, 2011; Karver, De Nadai, Monahan, & Shirk, 2018; Norcross & Lambert, 2018; Norcross & Wampold, 2011), andere Beziehungsfaktoren wie z. B. Empathie (Elliott, Bohart, Watson, & Greenberg, 2011; Elliott, Bohart, Watson, & Murphy, 2018), Zielkonsens/Zusammenarbeit bzw. Kollaboration (Shick Tryon, Birch, & Verkuilen, 2018; Tryon & Winograd, 2011), Erwartungen (Constantino, Arnkoff, Glass, Ametrano, & Smith, 2011; Constantino, Visla, Coyne, & Boswell, 2018; Doering, Glombiewski, & Rief, 2018; Glass, Arnkoff,

& Shapiro, 2001; Greenberg, Constantino, & Bruce, 2006; Price, Anderson, Henrich, & Rothbaum, 2008; Price & Anderson, 2012; Rief & Glombiewski, 2017; Rief & Glombiewski, 2016; Strauß & Linden, 2013), Forscher\*innen- und Therapeut\*innenalleganz und kulturelle Adaptationen (Benish, Quintana, & Wampold, 2011) sowie Therapeut\*inneneffekte (Berglar et al., 2016; Goldberg, Hoyt, Nissen-Lie, Nielsen, & Wampold, 2018; Imel, Sheng, Baldwin, & Atkins, 2015; Johns, Barkham, Kellett, & Saxon, 2019). Viele Autor\*innen plädieren für ein kontextuelles Modell, da das Modell der evidenzbasierten Medizin (auch: medizinisches Modell) nur unzureichend auf die Psychotherapie übertragbar zu sein scheint und das Forschungsfeld nicht voranbringt. Es ist daher wahrscheinlicher von einer Therapeut\*in-Technik-Patient\*in-Interaktion auszugehen, welche den Psychotherapieerfolg zu bestimmen scheint.

Im Folgenden soll beispielhaft auf den Einfluss der Alleganz eingegangen werden, da diese eine wichtige Rolle bei der Interpretation von Psychotherapiestudien spielt. Sie bezieht sich auf den Grad, in dem die Therapeutin oder Forscherin glaubt, dass die Therapie wirksam ist und ist ein zentraler Mechanismus, um positive Behandlungs- und somit Veränderungserwartungen zu schürren. Alleganz (insbesondere von Therapeut\*innen) ist schwierig zu untersuchen, aber aktuellen Schätzungen zufolge beläuft sich die gemittelte Forscher\*innen- und Therapeut\*innenalleganz auf eine Effekstärke von ca.  $d = .65$  (Wampold & Imel, 2015). Beispielsweise zeigt eine Untersuchung, dass die Studien aus der Arbeitsgruppe von Adrian Wells, dem „Erfinder“ der Meta-Kognitiven Therapie (MKT), signifikant höhere Effekstärken für die MKT aufweisen als Studien, welche von anderen Arbeitsgruppen durchgeführt wurden (Normann & Morina, 2018).

Lambert (2013) hat die prozentuale Verteilung der Verbesserung bei Patient\*innen durch Psychotherapie als Funktion von verschiedenen therapeutischen Variablen basierend auf der Studienlage aufgeschlüsselt. Diese Verteilung ist in Abbildung 6 zu sehen. Aus der Abbildung 6 ist ersichtlich, dass sich der Großteil der Verbesserung der Symptomatik bei Psychotherapiepatient\*innen auf allgemeine Faktoren und auf Ereignisse im Leben der Patientin zurückführen lassen.

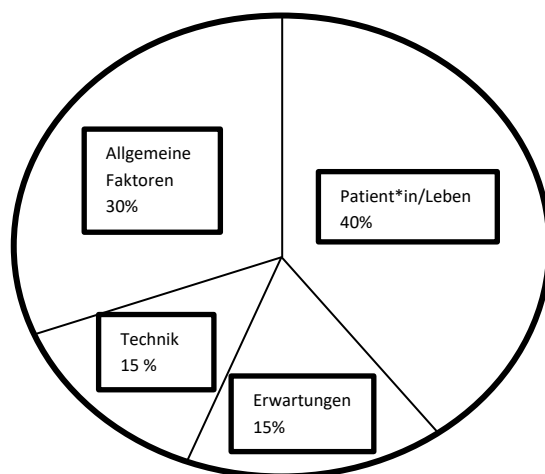


Abbildung 3. Schematische Darstellung der prozentualen Verteilung der Verbesserung in Psychotherapiepatient\*innen als Funktion von therapeutischen Faktoren (Lambert, 2013).

### 2.3.6. Research-Practice-Gap<sup>12</sup> oder warum die Elfenbeintürme weiter im Treibsand sinken

Ein zentrales Problem für die gegenwärtige Praxis und Forschung besteht darin, dass es eine deutlich zeitliche Verzögerung bei der Integration und Umsetzung von aktuellen Forschungsergebnissen in die alltägliche klinische Praxis gibt (Wampold & Imel, 2015), sodass eine große Kluft zwischen den Erkenntnissen aus der Forschung und der Implementierung im Anwendungssetting herrscht (Bohus, 2015). Dieser Umstand wird oft als "Research-Practice-Gap" bezeichnet (Fourie, 1996; Goldfried, 2010, 2019; Teachman et al., 2012). In der Forschung wird ebenso das Phänomen des „therapeutischen Drifts“ diskutiert. Es wird definiert als Phänomen, was auftritt, wenn Praktiker\*innen scheitern an der Bereitstellung einer optimalen evidenzbasierten Therapie trotz der nötigen Tools, wodurch Effektstärken in der Praxis oft geringer ausfallen (Waller & Turner, 2016). In dem Review (Waller & Turner, 2016) werden Gründe dafür benannt (z. B. Angst und Sicherheitsverhalten der Therapeutin insbesondere bei Verhaltensübungen) und Lösungsvorschläge gemacht.

Besonderes gravierend kommt dieses Problem beispielsweise zum Tragen bei der Integration von KVT bzw. im Spezifischen der ERP in der Behandlung von Zwangsstörungen. In diesem Kontext wird der "Research-Practice-Gap" durch Studien belegt, die zeigen, dass weniger als die Hälfte der Therapeut\*innen regelmäßig ERP verwenden (Külz et al., 2010) und sogar etwa 21% der in KVT ausgebildeten Psychotherapeut\*innen niemals oder nur selten ERP in der ambulanten Behandlung von Zwangsstörungen einsetzen (Roth, Siegl, Aufdermayer, &

<sup>12</sup> Manchmal wird diese Kluft auch „Science-Practice-Gap“ genannt (Lilienfeld, 2013; Sobell, 1996). In der vorliegenden Arbeit wird jedoch fortlaufend der Begriff „Research-Practice-Gap“ verwendet, da er in der einschlägigen Literatur häufiger eingesetzt wird.

Reinecker, 2004). In Übereinstimmung damit geben ca. zwei Fünftel der Patient\*innen mit Zwangsstörungen in einer Studie an, dass sie nie eine KVT mit ERP erhalten haben (Voderholzer et al., 2015). Die Integration in die tägliche klinische Praxis wird durch verschiedene Probleme erschwert, z. B. durch die mangelnde Motivation der Patient\*innen, die begrenzte Anzahl von Kliniker\*innen, die für die Bereitstellung der Behandlung qualifiziert sind, und durch Herausforderungen bei der Durchführung der Behandlung, wie z. B. organisatorischen Schwierigkeiten (McKay et al., 2015; Moritz et al., 2019).

Einige Forscher\*innen (Bohus, 2015) plädieren für bidirektionale Forschungs- und Entwicklungsansätze, in denen eine strukturierte Zusammenarbeit zwischen Praktiker\*innen und Forscher\*innen im Sinne von einer angewandten Wissenschaft stattfindet. Praxis-Forschungsnetzwerke (siehe Abschnitt 2.2) können als Initiativen dienen, die Brücke im Kontext eines Bottom-up-Forschungsansatzes zu schlagen (Castonguay et al., 2013; Lutz, Neu, et al., 2019). Durch Praxisorientierte Forschung kann die Generierung von Praxisbasierter Evidenz die Kluft zwischen Praxis und Forschung verringern bei gleichzeitiger Verbesserung des Therapieoutcomes individueller Patient\*innen. Im Rahmen solcher Initiativen können große Datensätze generiert werden, wobei sowohl die Datenqualität als auch die Bereitstellung der Behandlung als Vulnerabilitäten bedacht werden müssen. Daher kann dieses Paradigma auch als Ergänzung zur Evidenzbasierten Praxis verstanden werden (Barkham & Mellor-Clark, 2003).

Um den genannten Problemen der Psychotherapieforschung entgegenzuwirken, scheint es zunehmend wichtiger, vorhandene evidenzbasierte Psychotherapieverfahren (z. B. KVT, ERP, DBT) weiterzuentwickeln, indem Ansatzpunkte zur Optimierung identifiziert werden. Solche Ansatzpunkte könnte die Identifikation von Prädiktoren für die Response-, Remission-, Deterioration-, Dropout-, und Relapseraten sein (z. B. Kegel & Flückiger, 2015), um die Behandlungsauswahl zu optimieren, die Behandlung maßzuschneidern (im Sinne von „tailored treatments“) und die Effekte zu maximieren. Zur Umsetzung dieses Forschungsstranges bedarf es möglicherweise einem Wechsel im Forschungsparadigma. Im Folgenden wird ein möglicher Ansatz eingeführt, der zur Optimierung von Effekten in der Psychotherapie dienen kann.

### 2.4 Personalisierte Medizin als Anhaltspunkt zur Optimierung von Effekten in der Psychotherapie: Ein möglicher Wechsel im Forschungsparadigma

Analog zum psychopharmakologischen Vorgehen wird Psychotherapie größtenteils nach dem Trial-and-Error-Prinzip in der Praxis verschrieben. Es mangelt an Leitlinien, welche psychotherapeutische Methode für welchen Patient\*in die größte Wirksamkeit verspricht. Um solche Entscheidungsprozesse zu optimieren und so das Outcome zu maximieren, bedarf es eines neuen Forschungsparadigmas. Mit Blick auf andere verwandte Disziplinen (z. B. Medizin) zeigt sich, dass insbesondere personalisierte Behandlungsvorhersagen eine immer größere Rolle einnehmen. Eine aktuelle Studie zeigt, dass auf der Basis von Patient\*innencharakteristika vor Behandlungsbeginn eine personalisierte Vorhersage auf die antidepressive vs. Placebo-Response bei Patient\*innen mit depressiven Störungen sinnvoll sein kann (Webb et al., 2019). In der personalisierten Medizin soll jede Patientin unter weitgehender Einbeziehung individueller Gegebenheiten, über die funktionale Krankheitsdiagnose hinaus, behandelt werden (Simon & Perlis, 2010). Das schließt auch das fortlaufende Anpassen der Therapie an den Gesundheitsfortschritt an. Dieser Ansatz wird in der Literatur häufig auch Präzisionsmedizin (Collins & Varmus, 2015) oder auch personalisierte evidenzbasierte Medizin genannt (Kent, Steyerberg, & Van Klaveren, 2018). Solche Therapieansätze haben sich in den letzten Jahren großer Beliebtheit erfreut, z. B. zeigte ein solcher Ansatz basierend auf Markov-Entscheidungsprozessen im Sinne einer simulierten ‚Artificial-Intelligence‘-Studie, dass dieser Algorithmus durchaus in der Lage ist, optimale Entscheidungen in komplexen und unsicheren Umgebungen zu treffen (Bennett & Hauser, 2013). Der Ansatz der personalisierten Medizin könnte möglicherweise auf die Psychotherapie übertragen werden (vgl. Delgadillo & Lutz, 2020) und dazu dienen, die vorher genannten Probleme der Psychotherapieforschung zu überwinden.

Ähnlich zu diesen Strömungen innerhalb der Medizin scheint das Matching von Patient\*innen zu Psychotherapien bzw. Psychotherapeut\*innen (Delgadillo, Rubel, & Barkham, 2020) zurzeit ebenso ein guter Ansatz zur empirischen Evidenz darzustellen und die Abkehr von vorherigen Trends zu rigider Therapieschulenallegianz weiter zu fördern (Lambert, 2013). Eine offene Studie deutet darauf hin, dass eine personalisierte und modularisierte Behandlung für Depression und Angststörungen sehr vielversprechend sein kann (Fisher et al., 2019). Praktiker\*innen behandeln nicht die durchschnittliche Patientin, sondern Patient\*innen mit



Komorbiditäten, wenden oft nicht strikte Behandlungsmanuale an, arbeiten eklektisch und integrativ, und passen ihr Vorgehen auf die individuellen Charakteristika des Patient\*innen an (Cook, Biyanova, Elhai, Schnurr, & Coyne, 2010). Die Mehrheit aller Therapeut\*innen bietet daher zwar laut einem Selbstbericht „individuelle“ Psychotherapie an und geht nach einem „personalisierten“ Ansatz vor (Cook et al., 2010). Praktiker\*innen können allerdings ihre eigenen Misserfolge nur schlecht erkennen und vorhersagen (Hannan et al., 2005). Dadurch besteht ein erhöhter Bedarf an klinischen Unterstützungstools bei der Entscheidung und Anpassung von Behandlungen.

Die Psychotherapieforschung kann sich dabei an aktuellen Entwicklungen aus dem biomedizinischen Feld und der sog. „Computational Psychiatry“ bedienen, um die Perspektive zu erweitern (Moutoussis, Shahar, Hauser, & Dolan, 2018; Stephan, Siemerikus, Bischof, & Haker, 2017). In der „Computational Psychiatry“ ist das Ziel, ein mathematisch-informiertes Verständnis von psychischen Störungen zu erlangen und dieses Verständnis in computationaler Terminologie auszudrücken (Friston et al., 2014). Die „Computational Psychiatry“ nutzt dafür formale Modelle von Gehirnfunktionen zur Charakterisierung von psychopathologischen Mechanismen, welche durch computationale und mathematische Begriffe beschrieben werden kann (Friston et al., 2014). Das Gehirn kann aus dieser Perspektive heraus als Organ gesehen werden, welches Inferenzprobleme löst (Adams, Huys, & Roiser, 2016). Dazu zählen z. B. Verstärkungslernen, welche die Kernaspekte der Wahl formalisiert, und das sog. „Predictive Coding“, welche im Rahmen einer Bayesschen Perspektive die Kernaspekte von Inferenz formalisiert und die Bedeutung von Unsicherheit unterstreicht. In der „Computational Psychiatry“ werden z. B. Machine Learning-Algorithmen (ML-Algorithmen) verwendet, um diese Probleme zu analysieren (Huys, 2017; Stephan et al., 2017). Die „Computational Psychiatry“ umfasst insgesamt zwei Ansätze, sowohl (1) datengetriebene, theoretisch agnostische Datenanalysemethoden aus ML-Algorithmen, die breit angelegt sind (einschließlich erweiterten statistischen Standardmethoden), als auch (2) theorie-getriebene Modelle, die mechanistisch interpretierbare Beziehungen zwischen Variablen mathematisch spezifizieren (oft einschließlich sowohl beobachtbarer Variablen als auch postulierter, theoretisch sinnvoller verborgener Variablen, z. B. durch Bayessche Statistik) (Huys, Maia, & Frank, 2016). In dieser Dissertation soll sowohl ein datengetriebenes Vorgehen mit ML-Algorithmen als auch ein theorie-getriebenes Vorgehen mit Bayesscher Statistik verfolgt werden, um die Vor- und Nachteile beider Ansätze je nach Fragestellung zu kombinieren.

### 2.4.1. Klinische multivariate Prädiktionsmodelle: Einbezug von Machine Learning und Bayesscher Statistik

Machine Learning (ML) zielt darauf ab, Methoden zu entwickeln, die automatisch die Struktur in Daten entdeckt und dann die aufgedeckten Muster nutzt, um zukünftige Daten oder Outcomes vorherzusagen, oder andere Formen der Entscheidungsfindung unter Unsicherheit auszuführen (Murphy, 2012). Die systematische Anwendung der probabilistischen Argumentation auf alle schlussfolgernden Probleme, einschließlich der Ableitung von Parametern statistischer Modelle, wird häufig als Bayesscher Ansatz bezeichnet (Murphy, 2012). Im Sinne der probabilistischen Theorie können ML-Methoden auf jedes Problem angewandt werden, welches mit Unsicherheiten einhergeht. Die bekanntesten Anwendungsformen sind Mustererkennung (Bishop, 2006) und Merkmalsauswahl. Es kann insgesamt in zwei Formen durchgeführt werden: der prädiktive oder supervidierte Lernansatz, bei dem in ein Training- und ein Testset (bzw. Validierungsset) unterschieden wird, und auf Klassifikation und Mustererkennung abzielt (z. B. Dokumentenklassifikation und Spamfilter im Emailprogramm, Klassifikation von Blumen, Bildklassifikation, Handschrifterkennung, Gesichtserkennung), als auch der deskriptive oder unsupervidierte Lernansatz zur Entdeckung von Wissen (z. B. Entdecken von Clustern im E-Commerce, Entdecken von latenten Faktoren, Entdecken von graphischen Strukturen, Vervollständigung einer Matrix bei Imputationsverfahren von plausiblen Werten für Missings). Es existieren demnach parametrische und non-parametrische Modelle (Murphy, 2012).

Im Rahmen einer Prognoseforschung können klinische multivariate Prädiktionsmodelle entwickelt werden, um die Effekte von evidenzbasierten Psychotherapien wie der KVT noch zu optimieren und den Weg zu einer personalisierten evidenzbasierten Psychotherapie zu ebnen. Die jährliche Anzahl an Publikationen mit den Begriffen „prediction“ und „model“ ist in den letzten 25 Jahren exponentiell gestiegen auf über 50.000 Treffen in der Datenbank *Web of Science*, vor allem im Bereich der Naturwissenschaften, allen voran: Ingenieurwissenschaften, Mathematik, Computerwissenschaften/Informatik und Physik mit jeweils über 100.000 Treffern (Steyerberg, 2019). Solche Prädiktorenmodelle können durch ML-Ansätze (insbesondere durch prädiktive bzw. supervidierte Lernansätze) generiert werden, welche einige Vorteile gegenüber klassischen statistischen Auswertungsverfahren insbesondere bei großen Datensätzen (sog. „big data“) aufweisen (Bzdok, Karrer, Habel, & Schneider, 2017; Dwyer, Falkai, & Koutsouleris, 2018), auch wenn einige Forscher\*innen behaupten, dass ML-

Ansätze insgesamt keinen Leistungsvorteil gegenüber logistischen Regressionen bei der Entwicklung von klinischen Prädiktionsmodellen haben (Christodoulou et al., 2019). Der größte Vorteil ist jedoch deren Robustheit gegenüber der Verwendung einer großen Anzahl von Variablen, Merkmalen oder Prädiktoren im Verhältnis zur Anzahl der Beobachtungen bzw. Fallzahlen. So ist beispielsweise ein lineares Standard-Regressionsmodell schnell von einer großen Anzahl von Prädiktoren überfordert, wohingegen ML-Modelle robuster gegenüber einem hohen Verhältnis von Prädiktoren zu Beobachtungen sind, selbst wenn die Anzahl der Prädiktoren die Anzahl der Beobachtungen übersteigt, wie es z. B. bei fMRT-Studien geschehen kann, bei denen Hunderte oder sogar Tausende von Voxeln als Prädiktoren von hunderten Beobachtungen für jede Bedingung verwendet werden (Coutanche & Hallion, 2019). Außerdem sind ML-Algorithmen robuster gegenüber dem sog. „Overfitting“, zumindest wenn wenig flexible bzw. einfache Modelle eingesetzt werden (Murphy, 2012). Der Nachteil besteht in der Modellauswahl, da eine Vielzahl an Modellen mit unterschiedlicher Komplexität bzw. Flexibilität generiert werden. Eine natürliche Lösung wäre die Berechnung der Fehlklassifizierungsrate auf einem komplett unabhängigen Testset, der durch den Generalisierungsfehler<sup>13</sup> ausgedrückt werden kann und zur Auswahl eines Modells mit einer adäquaten Komplexität führt (Murphy, 2012). Wenn kein Zugriff auf ein Testset besteht, kann ein Validierungset aus den Trainingsdaten partitioniert werden, worauf die Leistung des ausgewählten Modells evaluiert werden kann. Falls das Trainingsset dadurch zu klein werden sollte, wird oft eine Kreuzvalidierungsprozedur (engl.: „cross validation (CV)“) verwendet, v. a. das sog. „5-fold CV“ und das „leave-one out cross validation (LOOCV)“ (Murphy, 2012).

ML beschäftigt sich also mit dem Entwickeln von verschiedenen Modellen und verschiedenen Algorithmen, um diese wiederum kontinuierlich anzupassen und weiterentwickeln zu können. Box & Draper (1987) schreiben in ihrem einflussreichen Werk folgenden Satz: „All models are wrong, but some models are useful.“ und treffen damit in die Kernproblematik von ML-Ansätzen. Die Annahme, dass kein universell bestes bzw. gültiges Modell existiert, wird oft als sog. „No-free-Lunch“-Theorem bezeichnet (Wolpert & Macready, 1997). Als Konsequenz müssen Forscher\*innen verschiedene Modelle entwickeln und deren Einsatz genauestens spezifizieren, um die große Vielfalt von Daten in der realen Welt abzudecken. Für jedes Modell kann es viele unterschiedliche Algorithmen geben, mit denen wir

---

<sup>13</sup> Der Generalisierungsfehler ist der erwartete Wert der Fehlklassifizierungsrate, wenn dieser über zukünftige Daten gemittelt wird (Murphy, 2012).

das Modell mit unterschiedlichen Geschwindigkeit-Genauigkeit-Komplexität-Tradeoffs trainieren können. Die Kombination von Daten, Modellen und Algorithmen ist wichtig (Murphy, 2012).

Die erfolgreiche Anwendung von ML-Algorithmen zur personalisierten Vorhersage des Therapieoutcomes sowie Matching von Patient\*in zu Therapeut\*in gelang in letzter Zeit über verschiedene Kontexte hinweg (Delgadillo, Rubel, et al., 2020; Meinlschmidt et al., 2019; Rubel, Zilcha-Mano, Giesemann, Prinz, & Lutz, 2019; Schwarzmeier et al., 2019). Dabei können verschiedene Algorithmen bei verschiedenen Modellen angewendet werden und umgekehrt (Murphy, 2012). In der Psychotherapieforschung hat sich in den letzten Jahren insbesondere der *Personal Advantage Index (PAI)* (Cohen & DeRubeis, 2018; Cohen, Kim, Van, Dekker, & Driessen, 2017; Deisenhofer et al., 2018; DeRubeis et al., 2014; Huibers et al., 2015; van Bronswijk et al., 2019; Webb et al., 2019) für Studien mit mehreren Behandlungsbedingungen als geeignet erwiesen, um personalisierte Behandlungsentscheidungen zu unterstützen. Ebenso kann der „Nearest Neighbor“-Ansatz (Lutz et al., 2005) eingesetzt werden, um zu untersuchen, wie ähnliche Patient\*innen in der Vergangenheit auf die unterschiedlichen Behandlungen reagiert haben. Eine aktuelle Studie hat gezeigt, dass sich ML-Ansätze durchaus eignen, um prädiktive Modelle zur Prognose des Therapieoutcomes bei Patient\*innen mit depressiven Störungen zu generieren (Webb et al., 2020). Die Synthese multipler Variablen durch ML kann zudem dabei die Vorhersage von Behandlungsergebnissen verbessern (Iniesta et al., 2016).

### **2.4.2. Vorhersage von Effekten in der Psychotherapie: Prognostische vs. preskriptive Prädiktoren**

Ein Prädiktor ist eine Variable, die weitgehend mit dem Therapieoutcome über alle Behandlungsbedingungen hinweg in einem Zusammenhang steht (Kraemer, Wilson, Fairburn, & Agras, 2002). Es existieren verschiedene Kategorien von Prädiktoren, welche für Prädiktionsmodelle vorgeschlagen wurden. Dazu zählen folgende Variablen, die als potentielle Prädiktoren dienen können (Harrell, 2015): demographische Variablen (z. B. Alter, Geschlecht, Ethnizität, Bildung, sozioökonomischer Status), Art und Schweregrad der Erkrankung (z. B. Hauptdiagnose, vorliegende Charakteristika), vorgeschichtliche Charakteristika (z. B. vorherige Erkrankungsepisoden, Risikofaktoren), Komorbiditäten (bzw. begleitende Erkrankungen), physischer Funktionsstatus (z. B. Karnofsky-Index, WHO Performance-Score) sowie subjektiver

Gesundheitsstatus und Lebensqualität (psychologische, kognitive und psychosoziale Funktionsfähigkeit). Bei der Entwicklung von Modellen, die sich auf die tägliche Praxis anwenden lassen, kann die Wahl von Prädiktoren pragmatischer getroffen werden (Moons et al., 2015). In diesem Fall sollten Definitionen und Scorings verwendet werden, die in Übereinstimmung mit der täglichen Praxis sind, i.e. Prädiktoren beinhalten, die ziemlich leicht verfügbar, nicht zu teuer in der Beschaffung und mit angemessener Genauigkeit messbar sind. Es hat sich in der Forschung gezeigt, dass man vorsichtig sein sollte, wenn ein Prädiktionsmodell in einem Kontext mit unterschiedlicher Messgenauigkeit angewandt wird, da eine solche Differenz einen Einfluss auf die Generalisierbarkeit des Modells hat (Luijken, Groenwold, Van Calster, Steyerberg, & van Smeden, 2019). Die Anwendung der Modelle, die in den Studien der vorliegenden Dissertation generiert werden, sollte sich daher explizit auf psychosomatische Kliniken beziehen, da dort von einer ähnlichen Messgenauigkeit auszugehen ist. Die Wahl der Prädiktoren deckt sich mit der Erhebung der Routinedaten während der Aufnahme in die Klinik und wurde daher pragmatisch getroffen (Moons et al., 2015).

Prognostische Variablen im Rahmen von sog. „Single-Treatment“-Studien können als generelle Prädiktoren aufgefasst werden und dienen daher zur Vorhersage, ob eine Patientin auf eine Behandlung anspricht, unabhängig von der Behandlung. Solche prognostischen Baseline-Charakteristika zu identifizieren ist wichtig, da sie für RCTs zur Randomisierung relevant sein können (Steyerberg, 2019). Sie können dabei helfen, den Einschluss und die Stratifizierung von Patient\*innen in prospektive Studien zu leiten und so die statistische Analyse einer RCT verbessern. Die Stärke der Randomisierung führt dazu, dass die Vergleichbarkeit zwischen den behandelten Gruppen sowohl in Bezug auf beobachtete als auch auf nicht beobachtete Ausgangsmerkmale erhöht wird. In RCTs können zwar keine systematischen Konfundierungen auftreten, aber einige beobachtete Baselinemerkmale können stark mit dem Outcome zusammenhängen. Die Anpassung für solche Kovariaten hat mehrere Vorteile (Gail, Wieand, & Piantadosi, 1984; Hauck, Anderson, & Marcus, 1998; Hernández, Eijkemans, & Steyerberg, 2006; Hernández, Steyerberg, & Habbema, 2004; Pocock, Assmann, Enos, & Kasten, 2002; Robinson & Jewell, 1991), z. B. jegliche Verzerrungen in der Schätzung des Behandlungseffekts zu reduzieren, die durch ein zufälliges Ungleichgewicht zwischen den Gruppen entstehen, und um die statistische Aussagekraft zur Erkennung eines Behandlungseffekts zu erhöhen. Die Bestimmung von Patient\*innencharakteristika zur Profilerstellung kann schon vor der Behandlung Aufschluss über Patient\*innen geben, die

wahrscheinlich nicht auf die Behandlung ansprechen. Somit können sie eine Prognose bzw. Vorhersage über individuelle Therapieverläufe geben (Delgadillo, Moreea, & Lutz, 2016) und als Red- bzw. Yellow-Flags für Behandler\*innen dienen. Es ermöglicht das Bilden von Subgruppen und Cluster, für die entweder neue Interventionen entwickelt oder einzelne als wirksam erwiesene Therapieelemente hinzugenommen werden müssen, z. B. im Sinne einer modularen Psychotherapie (vgl. Bohus, Herpertz, & Falkai, 2012; Brakemeier & Herpertz, 2019).

Zwei Formen von allgemeinen Prädiktoren zur Prognose der Behandlungs-Response sind in der Abbildung 4 dargestellt, adaptiert nach Simon & Perlis (2010). Die erste Form versteht sich als allgemeine Prädiktoren einer Symptomveränderung, ungeachtet einer Behandlung; die zweite Form als allgemeine Prädiktoren einer Response auf eine Behandlung, unabhängig der spezifischen Behandlung.

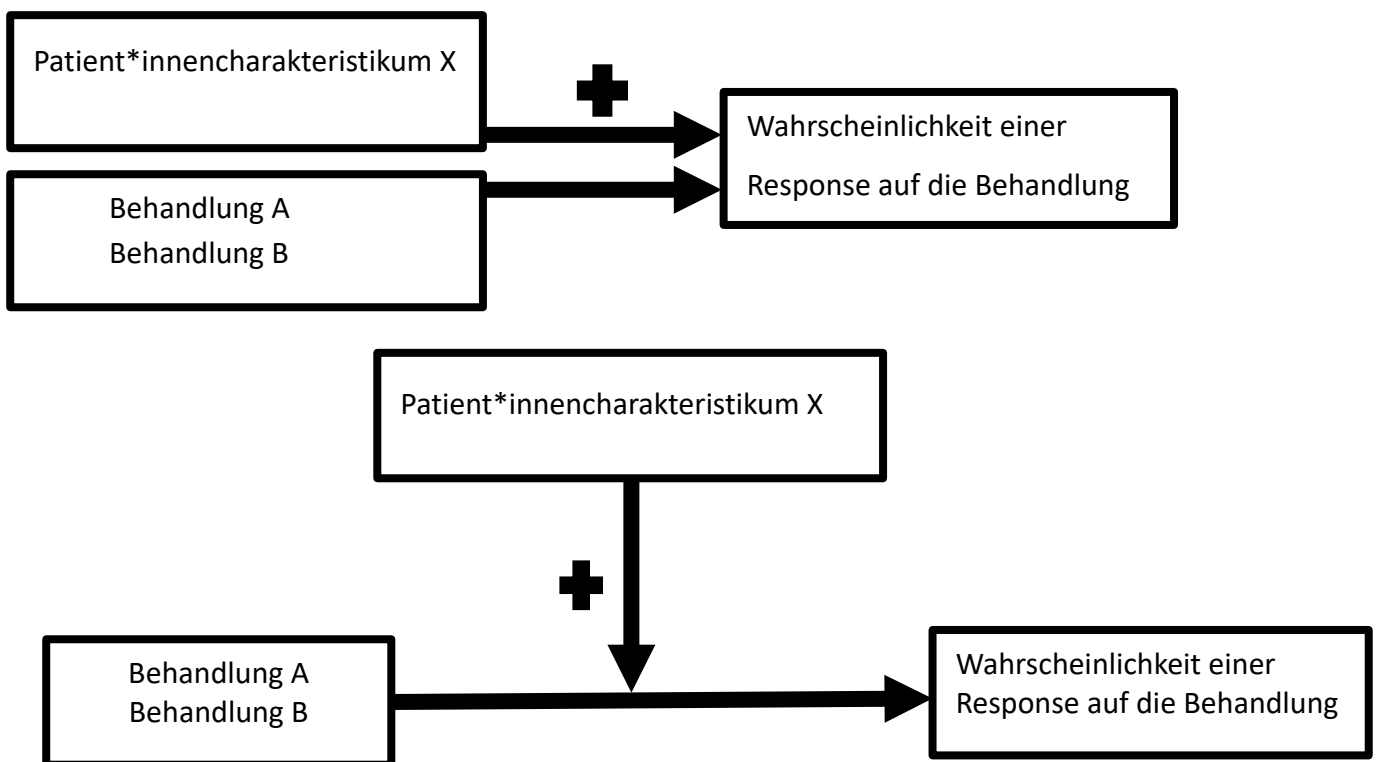


Abbildung 4. Obere Abbildung: Patient\*innencharakteristik X sagt ein besseres Outcome vorher, ungeachtet von der Behandlung. Untere Abbildung: Patient\*innencharakteristik X sagt einen größeren Benefit von jeder Behandlung vorher, adaptiert nach Simon & Perlis (2010).

Einige allgemeine Prädiktoren als prognostische Variablen haben die Forschung bisher geleitet und die identifizierten Variablen könnten als Indikatoren für allgemein schwer zu behandelnde Patient\*innen nützlich sein (Frank et al., 2011). Insgesamt scheint die Vorhersage des individuellen Behandlungsergebnisses durch hypothetische Prädiktorvariablen bis heute

allerdings schwierig zu sein (z. B. Falkenstein et al., 2019). So finden sich für verschiedene Störungen heterogene Ergebnisse zu den potentiellen Prädiktoren, so z. B. bei Patient\*innen mit PTBS (Karatzias et al., 2019; Lonergan, 2014; Richardson et al., 2014; Rizvi, Vogt, & Resick, 2012; Sonne et al., 2016; Van Minnen, Arntz, & Keijsers, 2002); bei Patient\*innen mit Zwangsstörungen (Keeley, Storch, Merlo, & Geffken, 2008; Knopp, Knowles, Bee, Lovell, & Bower, 2013); bei Patient\*innen mit Borderline-Persönlichkeitsstörungen (Barnicot et al., 2011; Barnicot et al., 2012; De Panfilis et al., 2012; Kliem & Kröger, 2013; Kröger, Harbeck, Armbrust, & Kliem, 2013; McMurrin et al., 2010; Rüsck et al., 2008); und bei Patient\*innen mit depressiven Störungen (Addis & Jacobson, 1996; Blom et al., 2007; Carter et al., 2015, 2011; Constantino et al., 2008; Enns, Cox, & Pidlubny, 2002; E. Frank et al., 2011; Gaston, Marmar, Gallagher, & Thompson, 1989; Groves, Douglas, & Porter, 2018; Hundt, Calleo, Williams, & Cully, 2016; Jarrett, Eaves, Grannemann, & Rush, 1991; Kertz, Koran, Stevens, & Björgvinsson, 2015; Mynors-Wallis & Gath, 1997; Ravitz, McBride, & Maunder, 2011; Rief et al., 2015; Stiles-Shields, Corden, Kwasny, Schueller, & Mohr, 2015; Thase et al., 1994).

Diese uneinheitlichen Ergebnisse lassen sich vor allem auf die geringe Power wegen einer zu geringen Stichprobengröße, die Verwendung unterschiedlicher Outcomemaße, den Einschluss einer geringen Anzahl an potentiell relevanten Vorbehandlungsvariablen sowie die häufige Verwendung von ungenügenden statistischen Ansätzen zurückführen (Lee et al., 2018; Simon & Perlis, 2010). Neue statistische Modelle, die durch die Kombination von klinischen und demographischen Daten mehrere Datenquellen nutzen, können hilfreich sein (Chekroud et al., 2016; Cipriani & Geddes, 2016; Høifødt et al., 2015).

Präskriptive Variablen im Rahmen von sog. „Multiple-Treatment“-Studien können als differentielle Prädiktoren aufgefasst werden und den differentiellen Therapieerfolg bestimmen. Solche Prädiktoren können letztlich auch als Moderatoren für das Ansprechen auf eine Therapie verstanden werden (Simon & Perlis, 2010), z. B. konnten im Rahmen von aktuellen Moderatoranalysen solche Variablen zur Unterscheidung zwischen CBASP und einer supportiven Gruppe für die Therapie der chronischen Depression (Serbanescu et al., 2020) und zur Unterscheidung zwischen KVT und IPT bei episodischen Depressionen (Whiston, Bockting, & Semkovska, 2019) bestimmt werden. In der Abbildung 5 sind differentielle Prädiktoren (Moderatoren) schematisch dargestellt.

## Optimierung der Effekte von Psychotherapie

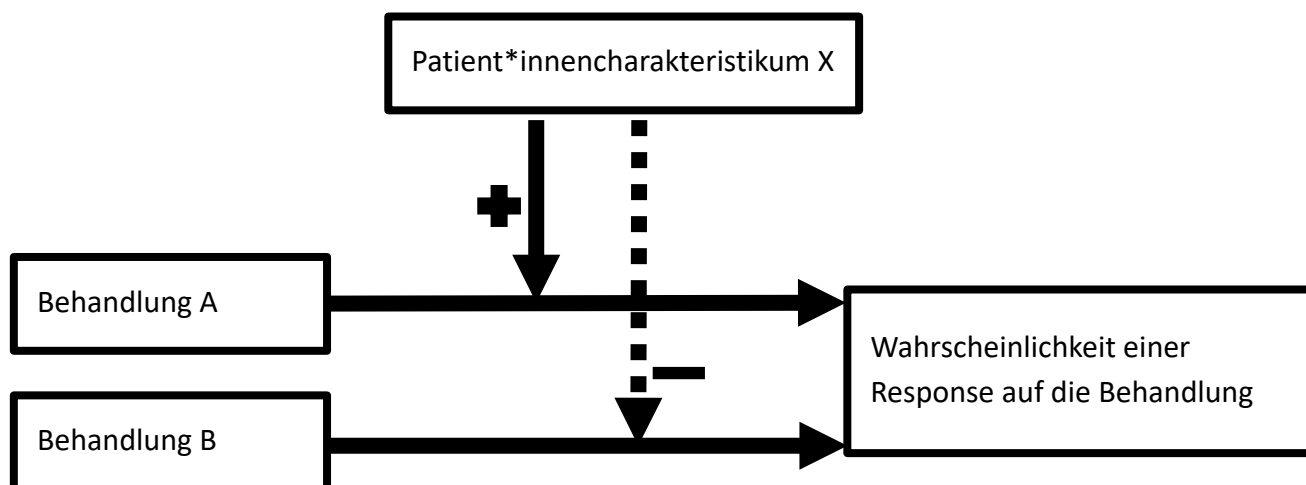


Abbildung 5. Patient\*innencharakteristik X sagt einen größeren Benefit für die Behandlung A als von Behandlung B vorher (d.h., moderiert den Vergleich in der Response auf Behandlung A und Behandlung B), adaptiert nach Simon & Perlis (2010).

Diese Variablen können daher bei der Auswahl einer optimalen Behandlung im Sinne von „Treatment selection“ helfen (Cohen & DeRubeis, 2018; Keefe et al., 2020) und dadurch Personalisierte Behandlungen (Grant et al., 2014; Simon & Perlis, 2010) ermöglichen. Ein zurzeit beliebtes modernes statistisches Verfahren ist, wie bereits erwähnt, die Nutzung des PAI (DeRubeis et al., 2014). Eine aktuelle Studie zeigte, dass der PAI zur Vorhersage von langfristigem Outcome bei Patient\*innen mit depressiven Störungen durch die Auswahl einer optimalen Behandlung im Sinne der „Precision Medicine“ das Therapieoutcome erhöht (van Bronswijk et al., 2019). Die Zuweisung zu therapeutischen Interventionen basierend auf ähnlichen Fällen verbessert das Therapieoutcome um ca. 30% (DeRubeis et al., 2014; Huibers et al., 2015). Eine frühe Identifizierung von Patient\*innen-Subgruppen über ihre vorhergesagten Behandlungsergebnisse können die Wahrscheinlichkeit der Verschreibung einer korrekten Behandlung zu einem frühen Zeitpunkt im Verlauf der Behandlung erhöhen und das Phänomen des Trial-and-Error verhindern. Eine Studie identifiziert zum Beispiel einen solchen differentiellen Prädiktor für psychologische Interventionen: mehr komorbide Persönlichkeitsstörungssymptome sind mit einem geringeren Ansprechen auf IPT, aber nicht auf KVT verbunden (Carter et al., 2011), andere Studien vergleichen Antidepressiva und KVT auf symptomspezifischer Ebene (Boschloo et al., 2019) oder eine antidepressive Behandlung vs. IPT mit jedoch tendenziell wenigen robusten Ergebnissen (Frank et al., 2011). Interessanterweise wurde auch bereits ein Instrument zur klinischen Risikostratifizierung für die Vorhersage von Behandlungsresistenzen bei Depressionen entwickelt, das die grundlegenden soziodemographischen und klinischen Merkmale berücksichtigt, aber noch weiter ausgearbeitet werden muss (Perlis, 2013; Trivedi, 2013).



### 2.4.3. Untersuchung von Veränderungsmechanismen: Mediatoren von Effekten in der Psychotherapie

Für die Psychotherapieforschung ist es essentiell wichtig zu verstehen, wie und warum Psychotherapie wirkt (Kazdin, 2007, 2009). Neben der bereits erwähnten Forschung, die unter dem Oberbegriff der Outcomeforschung betrieben wird, unterscheidet man noch die Prozess- sowie die Prozess-Outcome-Forschung. In der Psychotherapieforschung ist daher ebenso wichtig, (1) die Bedingungen herauszufinden, unter denen eine Behandlung mehr oder weniger effektiv ist, als auch (2) die Prozesse herauszufinden, durch welche eine Behandlung eine Veränderung erzeugt (Baron & Kenny, 1986; Holmbeck, 1997). Letzteres ist mit Mediatoranalysen möglich.

Während Moderatoren definiert werden als Variablen, welche feststellen, auf wen und unter welchen Umständen Behandlungen differentielle Effekte haben, werden Mediatoren als Variablen definiert, die identifizieren, warum und wie Behandlungen bestimmte Effekte haben (Kraemer et al., 2002). Ein Mediator ist also eine Variable, die dazu dient, den Prozess zu erklären, durch den eine Psychotherapie sich auf ein Outcome auswirkt (Lambert, 2013). Der Mediatoreffekt verdeutlicht dabei den Mechanismus, durch den die unabhängige Variable (z. B. Behandlung) mit dem Outcome in Beziehung steht (z. B. behandlungsbedingte Veränderungen). Mediationsmodelle sind also von Natur aus kausale Modelle. Daher sagen uns signifikante Mediationspfade im Kontext eines experimentellen Designs (d.h. einer zufälligen Zuweisung) etwas über kausale Zusammenhänge. Zur Beantwortung solcher Fragestellung erfordert es Prozessdaten. Um Prozesse der therapeutischen Veränderung besser abbilden zu können, können jedoch Moderatoren dazu genutzt werden, um Prozessvariablen als Mediatoren zu identifizieren, z. B. die therapeutische Allianz (Zilcha-Mano, 2019). Da ein Moderator demnach ein Merkmal, welches die Richtung oder die Größe der Beziehung zwischen einer unabhängigen (z. B. die Psychotherapie) und einer abhängigen Variablen (z. B. das Outcome) beeinflusst (Kazdin, 2009), stehen sie somit in Beziehung zu Mediatoren und (Veränderungs-)Mechanismen, weil sie vermuten lassen, dass unterschiedliche Prozesse beteiligt sein könnten, z. B. bei Männern oder Frauen (Lambert, 2013). Langfristig können solche Erkenntnisse Forschungsbemühungen im Rahmen der Prozessbasierten Therapie als transdiagnostischen Ansatz fördern (Hayes & Hofmann, 2018; Hofmann & Hayes, 2019).

Kazdin (2009) unterscheidet noch einmal deutlich genauer zwischen Ursache, Mediator und Veränderungsmechanismus. Eine Ursache definiert er als Variable oder Intervention, die zum Ergebnis oder zur Veränderung führt und für diese verantwortlich ist. Diese Ursache kann, aber muss nicht zwingend, eine Behandlung sein. Ein Mediator ist eine dazwischenliegende Variable, die (statistisch) für die Beziehung zwischen den unabhängigen und abhängigen Variablen aufkommt. Dabei muss etwas, das Veränderung mitteilt, nicht unbedingt erklären können, wie die Veränderung zustande kam. Auch könnte der Mediator ein Stellvertreter für eine oder mehrere andere Variablen oder ein allgemeines Konstrukt sein, das nicht notwendigerweise dazu da ist, die Mechanismen der Veränderung zu erklären. Ein Mediator kann ein Hinweis sein, der auf die möglichen Veränderungsmechanismen verweist, aber muss nicht notwendigerweise ein Mechanismus sein. Ein Veränderungsmechanismus hingegen ist die Grundlage für die Wirkung (d.h. die Prozesse oder Ereignisse, die für die Änderung verantwortlich sind; die Gründe, warum eine Veränderung eingetreten ist, oder auf welche Weise der Wechsel zustande kam). Kazdin (2009) formulierte gewisse Anforderungen, um sowohl Mediatoren als auch Veränderungsmechanismen aufzuzeigen (z. B. starke Assoziation, Spezifität, Konsistenz, experimentelle Manipulation, zeitlicher Verlauf, Gradient und Plausibilität oder Kohärenz).

Beispielsweise geht der aktuelle Forschungsstand davon aus, dass kognitive Veränderungen zur kurzfristigen Symptomveränderung beitragen, es allerdings unabhängig davon zu sein scheint, wie bzw. ob sie durch kognitive Techniken erreicht werden (Lorenzo-Luaces, German, & DeRubeis, 2015; Lorenzo-Luaces, Keefe, & DeRubeis, 2016). Im Rahmen der Depressionsbehandlung wird die kognitive Veränderung als allgemeiner anstatt eines spezifischen Veränderungsmechanismus gehandelt (Hofheinz, Reder, & Michalak, 2019). Durch das Adressieren von solchen Veränderungsmechanismen können Behandlungen auch im Hinblick des Langzeitoutcomes optimiert werden. Zum Beispiel können sie dabei helfen, das Potenzial eines chronischen Verlaufs der Depression zu verringern und gleichzeitig Behandlungsresistenzen und die damit verbundenen Faktoren zu verhindern, d.h. die Morbidität und Mortalität reduzieren, das Leiden verkürzen, die Rate der Behandlungsabbrucher\*innen senken sowie die Kostenbelastung für den Patient\*innen und die Gesellschaft reduzieren (Bergfeld et al., 2018; Cepeda, Reys, & Ryan, 2018; DiBernardo et al., 2018; Johnston, Powell, Anderson, Szabo, & Cline, 2019; Kautzky et al., 2019; Lex et al., 2019; Reutfors et al., 2018). Mit Hilfe von Bayesschen Modellen zur Abbildung von

Therapieeffekten können zwar nur bedingt Aussagen über Veränderungsmechanismen getroffen werden, jedoch kausale Zusammenhänge über bspw. sowohl kurz- als auch langfristige Veränderungen der Symptomatik während und nach der Psychotherapie abgebildet werden, was u.a. nützlich für die zukünftige Patient\*innenbehandlung als auch die ambulante Nachbehandlung nach der stationären Psychotherapie in den psychosomatischen Kliniken sein kann.

In den Studien 1-4 sollen die Effekte der stationären Psychotherapie für verschiedene Störungsbilder im Kontext von prognostischen Variablen vorhergesagt werden, um Ansatzpunkte zur Optimierung dieser Effekte zu identifizieren. In der Studie 4 werden explizit auch komplexe Prädiktorenmuster für Veränderungen untersucht. Unter Zuhilfenahme von ML-Algorithmen und Bayesscher Statistik sollen klinische multivariate Prädiktionsmodelle generiert werden, welche den Praktiker\*innen bei der Personalisierung ihrer Behandlungen in den psychosomatischen Kliniken behilflich sein können.

### 2.5 Kehrseite der Medaille: Von positiven Effekten zu negativen Effekten von Psychotherapie

Die große Mehrheit von klinisch-psychologischen Studien im Zusammenhang mit dem Outcome von Psychotherapien konzentriert sich auf die Wirksamkeit (im Sinne von „efficacy“, „effectiveness“, „cost-effectiveness“, etc.) im Kontext von positiven Effekten. Wenn auch andere Gesundheitswissenschaften (z. B. Pharmazie, Medizin) die Schäden oder Nebenwirkungen durch die jeweiligen Behandlungen untersuchen und erheben, ist dieses Forschungsfeld in der Psychotherapieforschung noch eher in den Kinderschuhen und ein jahrelang stiefmütterlich behandeltes Thema (Lilienfeld, 2007). Um die Position von Psychotherapie in ihrer Beziehung zur Medizin zu etablieren, lag der Fokus über Dekaden hinweg auf der Bestimmung ihrer Wirksamkeit (Barlow, 2010). Das Auftreten von negativen Effekten in der Psychotherapie wird insbesondere im Vergleich zur Pharmakotherapie noch immer selten berichtet (Vaughan, Goldstein, Alikakos, Cohen, & Serby, 2014). Nur ungefähr 25% aller RCTs erwähnen das Auftreten von sog. „Schäden“ in der Psychotherapie (Jonsson, Alaie, Parling, & Arnberg, 2014). Die CONSORT-Gruppe hat festgestellt, dass das Monitoring von Schäden in RCTs im Bereich Verhaltensmedizin limitiert ist (Ioannidis et al., 2004). Bis heute gibt es wenig systematische und methodisch ausgereifte Forschung zur Erfassung und Berichten von negativen Effekten in der Psychotherapie (Berk & Parker, 2009; Lilienfeld, 2007; Parry, Crawford, & Duggan, 2016).

Keinen Schaden anzurichten ist im Sinne des Leitsatzes „Primum non nocere“ die erste ethische Maxime bei Mediziner\*innen und Psychotherapeut\*innen. In diesem Kontext wird kritisch eine vorläufige Liste von zehn wahrscheinlich schädlichen und zwei möglicherweise schädlichen psychologischen Behandlungen diskutiert, welche auf der Basis von Verschlechterungsraten von Psychotherapie erstellt wurde (Lilienfeld, 2007; Weiss et al., 2005), welche von negativen Effekten und Nebenwirkungen im engeren Sinne unterschieden werden müssen. Um Behauptungen über die Schädlichkeit von Psychotherapien zu untersuchen, muss zwischen Verschlechterung und Schaden unterschieden werden (Wampold & Imel, 2015). Eine Verschlechterung für eine Patientin wird festgestellt, wenn die Werte am Ende der Behandlung ein schlechteres Funktionslevel auf einem oder mehreren Outcomemaßen als zu Beginn der Behandlung aufweisen. Um jedoch ein Outcome als schädlich zu charakterisieren, muss gezeigt werden, dass diese Verschlechterung tatsächlich iatrogen ist, d.h. dass die Verschlechterung

durch die Behandlung verursacht wurde. Die Zuschreibung eines Schadens erfordert logischerweise den Ausschluss der plausiblen konkurrierenden Hypothesen für die beobachtete Verschlechterung (z. B. durch den natürlichen Verlauf ohne Behandlung, durch externe Ereignisse wie Todesfälle, Verlust von finanziellen Ressourcen, Beziehungsabbrüche, Messfehler) und muss darin resultieren, dass ein verringertes Funktionslevel nach der Behandlung ohne diese Behandlung nicht beobachtet worden wäre. In diesem Fall wird auch von behandlungsinduzierten Verschlechterungseffekten gesprochen (Lilienfeld, 2007).

In den letzten Jahren ist die Aufmerksamkeit an negativen Effekten und Nebenwirkungen von psychologischen Interventionen systematisch gestiegen (Crawford et al., 2016; Scott, 2017; Scott & Young, 2016). Mit der Beachtung von negativen Effekten in der klinisch-psychologischen Forschung sind auch zunehmend definitorische, klassifikatorische und terminologische Fragen aufgeworfen worden. Linden et al. (2018) definiert negative Effekte als unerwünschte Ereignisse, die in einem Zusammenhang zur Therapie stehen und Nebenwirkungen, Kunstfehler und ethisches Verhalten beinhaltet. Nebenwirkungen werden dabei durch eine ‚lege artis‘ durchgeführte Psychotherapie verursacht und können verschiedene Lebensbereiche betreffen (z. B. vorübergehende Symptomverschlechterung, Konflikte in interpersonellen Beziehungen, Stigmatisierung). Negative Effekte scheinen über verschiedene Kontexte hinweg aufzutreten (z. B. Boettcher, Rozentel, Andersson, & Carlbring, 2014; Ladwig, Rief, & Nestoriuc, 2014; Schneibel et al., 2017). Inzidenzraten, d.h. ein Patient\*in berichtet von mind. einem negativen Effekt während der Behandlung, bewegen sich zwischen 45.2% und 93.8% je nach Behandlungssetting (Ladwig et al., 2014; Rheker, Beisel, Kräling, & Rief, 2017). Diese hohen Raten sprechen für eine regelmäßige Erhebung von negativen Effekten während, am Ende und nach der Behandlung. Eine aktuelle Studie (Schermuly-Haupt & Linden, 2020) zeigt zudem: Psychotherapeut\*innen in Ausbildung rechnen in etwa jedem zweiten Fall mit dem Auftreten von Nebenwirkungen. Sie erwarten vor allem Symptomverschlechterungen und negative Entwicklungen im sozialen Netz. Es zeigt sich eine Tendenz, dass sie die eigenen Verfahren und die eigenen Behandlungen für weniger riskant halten als die der Kolleg\*innen. Ursachen von Nebenwirkungen sehen die Psychotherapeut\*innen vor allem in der therapeutischen Beziehung sowie Therapeut\*innenvariablen und weniger in Patient\*innencharakteristika.

Reviews zeigen, dass die Instrumente zur Erfassung von negativen Effekten in RCTs heterogen und nicht systematisch berichtet werden (Jonsson et al., 2014), z. B. in Studien zur chronischen Depression (Meister et al., 2016). Dieser Umstand lässt sich auf verschiedene Gründe zurückzuführen: Das Spektrum an potenziellen negativen Effekten ist größer als bei Pharmakotherapie, da nicht nur Symptome sondern auch soziales Verhalten beachtet wird (Szapocznik & Prado, 2007). Oft ist unklar, was mit negativ gemeint ist und auch die Herstellung des kausalen Zusammenhangs, d.h. dass man ein unerwünschtes Ereignis auch zweifelsfrei auf die Therapie zurückführen kann, ist schwierig (May, 1971). Darüber hinaus geht man davon aus, dass vermutlich auch ein Wahrnehmungsbias seitens der Psychotherapeut\*innen hin zu einem positiven Outcome vorhanden ist (Hatfield et al., 2010). Außerdem werden inhaltsähnliche Begriffe nicht einheitlich verwendet (Linden, 2013) und es besteht kaum eine trennscharfe Begriffsdefinition (Parry et al., 2016). Im englischsprachigen Wissenschaftsraum existieren Begriffe wie „deterioration“ (Bergin, 1963), „side effects“ (Berk & Parker, 2009; Schmoll, 2012), „negative effects“ (Ladwig et al., 2014; Suh, Strupp, & O’Malley, 1986), „negative outcome“ (Mohr, 1995), „unwanted/undesired effects“ (Lieberei & Linden, 2008; Wittmund, 2007), „adverse effects/events/outcomes“ (Jonsson, Johanson, Nilsson, & Lindblad, 2016; Klatte, Strauss, Flückiger, & Rosendahl, 2018; Moritz et al., 2015, 2018), „harm“ (Lilienfeld, 2007), „iatrogenic effects“ (Parry et al., 2016; Weiss et al., 2005) und „mistakes“ (Bienenstein & Rother, 2007; Riedler-Singer, 2005). Diese Gründe führen zu Schwierigkeiten bei Entwicklung von Messinstrumenten zur Erfassung von negativen Effekten bzw. Nebenwirkungen, sodass es bis heute keine allgemein akzeptierten Messinstrumente zur Erfassung von negativen Effekten bzw. Nebenwirkungen von Psychotherapie gibt und Regeln zur Planung wissenschaftlicher Studien sowie des Monitorings bei RCTs vorhanden sind. Um die aufgetretenen Probleme bzgl. der Relevanz und standardisierten Erfassung zu überwinden, fordern internationale Forscher\*innen, dass die Erhebung von negativen Effekten in RCTs zu psychologischen Therapien mit geeigneten Evaluationsmethoden stattfinden soll (Guidi et al., 2018).

Die Relevanz von negativen Effekten kann durch den Schweregrad und die Dauer bestimmt werden, indem der Zusammenhang zum kurz- und langfristigen Therapieoutcome hergestellt wird. Eine erste Studie zeigt, dass das Auftreten von negativen Effekten in einem negativen Zusammenhang zum Outcome bei Patient\*innen mit Zwangsstörungen stehen kann (Moritz et al., 2015). Die Bestimmung über verschiedene Kontexte, wie z. B. stationäre

Psychotherapieprogramme, hinweg kann zudem aufschlussreich sein, da spezifische Effekte aufgrund der kurzen und intensiven Behandlung in stationären Konzepten getriggert werden können (Schneibel et al., 2017). In einem stationären Setting werden oft schwer erkrankte Patient\*innen (z. B. behandlungsresistente chronisch depressive Patient\*innen) behandelt, sodass durch verschiedene Therapieansätze unterschiedliche Nebenwirkungsprofile auftreten können (Brakemeier et al., 2018; Meister et al., 2019).

Die Identifikation und Minimierung von negativen Effekten im Allgemeinen und Nebenwirkungen im Spezifischen kann somit auch im Rahmen der Optimierung von Behandlungseffekten von Psychotherapien verstanden werden. Um negative Effekte valide identifizieren zu können, bedarf es geeigneter Messinstrumente, die auf der Grundlage einer konsensualen Definition und Konzeptualisierung basieren. In der vorliegenden Disseration soll sich ebenso dieser Fragestellung gewidmet werden. Die Erstellung von spezifischen Profilen zu negativen Effekten kann darüber hinaus Aufschluss geben, welche Psychotherapie welche Art von negativen Effekten auslöst. Insofern das Auftreten von negativen Effekten im Zusammenhang zum Behandlungserfolg stehen, sollten diese im Rahmen der Psychotherapie frühzeitig adressiert und reduziert werden. Die Bestimmung der Relevanz von solchen negativen Effekten, die im Rahmen von spezifischen Psychotherapien auftreten, ist ein weiteres Vorhaben in der vorliegenden Dissertation.

### 3 Darstellung des Dissertationsvorhabens

#### 3.1 Relevanz und Herleitung der Fragestellungen

Die Erforschung der Behandlungseffekte, welche die Ergebnisse der RCTs zur Wirksamkeit von Psychotherapien mit den Komplexitäten, die von Praktiker\*innen festgestellt werden, integriert, ist für die Entwicklung optimaler psychischer Gesundheitsdienstleistungen von wesentlicher Bedeutung (Wampold & Imel, 2015). In Deutschland wird die psychotherapeutische Versorgung von schwer psychisch erkrankten Patient\*innen meistens in psychosomatischen oder psychiatrischen Kliniken durchgeführt, weshalb die Daten der Studien in der vorliegenden Dissertation hauptsächlich aus diesem Kontext stammen. Für die Psychotherapieforschung werden Phase IV-Studien gemäß des Phasenmodells nach Buchkremer & Klingberg (2001) zunehmend wichtiger. Solche Studien untersuchen die Anwendung von Psychotherapie unter Routinebedingungen und bewerten die klinische Nützlichkeit durch nicht-randomisierte Wirksamkeitsstudien im Sinne der „effectiveness“ mit erweiterten Einschlusskriterien und unter zunehmend naturalistischen Bedingungen, wodurch sie eine hohe externe und ökologische Validität vorweisen. Diese Studien können somit als ein Indikator für die Implementierung und Dissemination von evidenzbasierten Psychotherapien (gemäß EBP und nationaler Leitlinien) fungieren. Durch die Berechnung von Effekten können Benchmarks generiert werden, welche ein besserer Schätzer für die absolute Wirksamkeit (im Sinne der „clinical effectiveness“) darstellen können und als internes Benchmarking innerhalb einer Klinik bzw. in der vorliegenden Dissertation innerhalb eines Klinikverbunds von psychosomatischen Kliniken eingeordnet werden kann (Camp, 1989). Solche patient\*innenfokussierte Benchmarkingstudien (Barkham et al., 2010) dienen daher ebenso der Qualitätsverbesserung und sind Teil des Qualitätsmanagements.

Einige Forscher\*innen argumentieren, dass die ES zumindest für die Psychotherapie von depressiven Störungen in RCTs deutlich überschätzt sind (Cuijpers et al., 2019; Cuijpers et al., 2010; Cuijpers et al., 2013). Das Programm IAPT in Großbritannien zeigt hingegen beispielhaft, dass es möglich ist, entsprechende ES auch in der Routineversorgung durch eine gute Implementierung zu etablieren (Clark, 2018; Clark et al., 2018; Gyani et al., 2013; Radhakrishnan et al., 2013; Zala et al., 2019). Um die Implementierung von Psychotherapie im stationären Setting als wichtige Versorgungseinrichtung in Deutschland zu beurteilen, sind



daher Benchmarkingstudien sinnvoll. Im Kontext der Praxisorientierten Forschung (Castonguay & Muran, 2016) können in derartigen Benchmarkstudien zugleich praxisbasierte Evidenz generiert sowie Praxis-Forschungsnetzwerke etabliert und gefördert werden (Castonguay et al., 2013), wodurch einige der Probleme der Psychotherapieforschung (z. B. ‚Research-Practice-Gap‘) adressiert werden (Barkham & Margison, 2008).

Darüber hinaus kann die Übertragung eines Forschungsparadigmas im Sinne der personalisierten Medizin in die Psychotherapie sinnvoll sein, um andere Probleme der Psychotherapieforschung (insbesondere die Verfahrensinflation, Mangel an Befunden zu relativer Wirksamkeit, Stagnation von Effekten) zu überbrücken und die Effekte von Psychotherapie noch weiter zu optimieren. In diesem Rahmen erscheinen klinische Prädiktionsmodelle sinnvoll, welche u.a. ML-Ansätze und Bayessche Statistik verwenden, um robuste und stabile Prädiktoren zu identifizieren. Im Kontext von solchen ‚Single-Treatment‘-Studien können prognostive Variablen als allgemeine Prädiktoren entdeckt werden. Bei der Generierung von klinischen multivariaten Prädiktionsmodellen erscheint die Orientierung an den TRIPOD-Richtlinien sinnvoll (Collins, Reitsma, Altman, & Moons, 2015; Moons et al., 2015). Es ist wichtig zu erwähnen, dass es, wie von einigen Forscher\*innen beanstandet, nicht darum geht, eine Monokultur einer integrierten KVT-basierten Psychotherapie und somit einen „one size fits all“-Ansatz zu schaffen, sondern der Pluralität und Diversität gerecht zu werden, indem man vorhandene Verfahren noch optimiert (Leichsenring, Abbass, Hilsenroth, et al., 2018). Als Schattenseite der Psychotherapie ist kaum noch abzustreiten, dass evidenzbasierte Verfahren auch negative Effekte haben können. Die Erforschung von solchen Effekten ist jedoch noch nicht systematisch und methodisch ausgereift, was u.a. an einem fehlenden Überblick über vorhandene Messinstrumente sowie deren zugrundeliegenden Arbeitsdefinitionen und konzeptuellen Rahmen liegt. Außerdem fehlen noch Studien zur Identifikation von Nebenwirkungsprofilen von spezifischen Therapieverfahren, wodurch sich ein neues Forschungsfeld eröffnet.

Die vorliegende Dissertation versucht durch ein Benchmarkingansatz die Behauptung, dass die ES aus klinischen Studien überschätzt sein könnten, zu prüfen und legt den Fokus darauf, welche positiven und negativen Effekte in der stationären Routineversorgung erzielt werden, wodurch praxisbasierte Evidenz generiert wird. Die Dissertation beschäftigt sich daher mit folgenden Leitfragen:

1. Lassen sich die Ergebnisse aus der Wirksamkeitsforschung auf ein angewandtes stationäres Setting generalisieren?
2. Durch welche Variablen und Faktoren lassen sich unterschiedliche Outcomemaße zur Bestimmung des Psychotherapieerfolgs bei verschiedenen psychischen Störungen vorhersagen?
3. Wie können negative Effekte von Psychotherapien definiert, erfasst und deren Relevanz zum Therapieoutcome beurteilt werden?

Dieser Ansatz kann durch Rückmeldeschleifen eine Optimierung der Effekte von Psychotherapie in der stationären Routineversorgung ermöglichen und dadurch Praxis-Forschungsnetzwerke in Deutschland festigen (Castonguay et al., 2013).

### 3.2 Fragestellungen des Dissertationsvorhabens

Basierend auf der bisherigen Forschungslage wurden dem Dissertationsvorhaben folgende studienspezifische Fragestellungen zu Grunde gelegt:

**Studie 1:** Wie wirksam ist die Psychotherapie von Borderline-Persönlichkeitsstörungen in der stationären Routineversorgung und welche Prädiktoren des Behandlungserfolgs lassen sich identifizieren?

**Studie 2:** Wie wirksam ist die Psychotherapie von Posttraumatischen Belastungsstörungen in der stationären Routineversorgung und welche Prädiktoren des Behandlungserfolgs lassen sich dafür identifizieren?

**Studie 3:** Wie wirksam ist die Psychotherapie von Zwangsstörungen in der stationären Routineversorgung und welche Prädiktoren des Behandlungserfolgs lassen sich identifizieren?

**Studie 4:** Wie wirksam ist die Psychotherapie von Depressionen in der stationären Routineversorgung und welche Prädiktoren des Behandlungserfolgs lassen sich identifizieren?

**Studie 5:** Welche Messinstrumente zur Erfassung von negativen Effekten in der Psychotherapie existieren und wie ist deren Güte zu beurteilen?

**Studie 6:** Wie ist der Zusammenhang zwischen dem Auftreten von negativen Effekten mit dem Therapieoutcome eines stationären CBASP-Programms für chronische Depression?

## 4 Zusammenfassung der Studien

### 4.1 Studie 1: Wirksamkeit und Prädiktoren einer stationären Psychotherapie für Borderline-Persönlichkeitsstörungen

**Zitation:** Herzog, P., Feldmann, M., Voderholzer, U., Gärtner, T., Armbrust, M., Rauh, E., Doerr, R., Rief, W., & Brakemeier, E.-L. (accepted). Drawing the Borderline: Predicting treatment outcomes in patients with borderline personality disorder, *Behaviour Research and Therapy*.

**Hintergrund:** Patient\*innen mit einer Borderline-Persönlichkeitsstörung (BPS) werden in Deutschland häufig in psychosomatischen Kliniken behandelt, wobei Response-Raten von 44% berichtet werden. Um Behandlungspläne für die spezifischen Charakteristika und Probleme von Patient\*innen mit BPS maßzuschneidern, erscheint die Identifikation von Prädiktoren des Behandlungserfolgs anhand von Patient\*innenmerkmalen zu Beginn der Behandlung vielversprechend. Bisherige Studien mit diesem Ziel sind häufig durch zu geringe Stichprobengrößen für Prädiktorenanalysen limitiert und konzentrieren sich entweder auf den vorzeitigen Behandlungsabbruch oder Symptomveränderungen als Therapieoutcome, was zu heterogenen Ergebnissen führt. Daher verfolgt diese Studie das Ziel, die Wirksamkeit einer stationären DBT-Behandlung in der stationären Routineversorgung zu bestimmen und ein multivariates klinisches Prädiktionsmodell auf der Basis von verschiedenen Patient\*innencharakteristika zur Vorhersage des Therapieoutcomes zu generieren.

**Methode:** In einer naturalistischen Stichprobe von routinemäßig erhobenen Daten wurden 878 Patient\*innen mit BPS (M = 31.1 Jahre [SD = 10.3], 81.2% weiblich) mit nicht-manualisierter Dialektisch-Behavioraler Therapie (DBT) für durchschnittlich 66 Tage in einer psychosomatischen Klinik behandelt. Effektstärken werden mittels Hedges' g für verschiedene Outcomevariablen berechnet. Zusätzlich werden 91 potentielle Prädiktoren in einer Bootstrapped gesteigerten regularisierten Regressionsanalysen eingeschlossen, um stabile Prädiktoren für eine vorzeitige Therapiebeendigung sowie residualisierte symptom-spezifische (Brief Symptom List, BSL-95) und funktionale Veränderungen (Short Form 36, SF-36) zu ermitteln. Die Prädiktoren, die in einem Trainingset identifiziert werden, werden anschließend in einem Testset sowie in einem Design mit wiederholter fünffacher Kreuzvalidierung (5-fold CV) getestet.

**Ergebnisse:** Die Prä-Post-Effektstärken sind insgesamt klein (SF-36, BSL) bis mittel (BDI-II) und reichen von Hedges'  $g = .28$  bis  $.51$ . Ein höherer Schulabschluss und eine komorbide rezidivierende depressive Störung können als Hauptprädiktoren für Therapieabschluss identifiziert werden. Niedrige Affektregulationsfähigkeiten und keine vorherige ambulante Psychotherapie sagen residualisierte symptomsspezifische Veränderungen vorher. Ein jüngeres Alter, vorhandene Arbeitsfähigkeit, hohe Werte in emotionalen und physischen Rollenlimitationen und geringerer körperlicher Schmerz stehen in prädiktiven Zusammenhängen mit starken Verbesserungen im funktionellen Outcome. Die Vorhersage im Testsample erreicht einen AUC von 63.1% für vorzeitigem Therapieabbruch, 5.6% erklärte Varianz für residualisierte funktionale Veränderung und 0.6% für residualisierte symptomsspezifische Veränderung. Die prädiktive Qualität des Modells variiert, wobei das beste Modell für das symptombezogene Outcome gefunden wurde ( $R^2=18\%$ ).

**Diskussion:** Die Effekte der stationären DBT-Behandlung liegen etwas unterhalb von denjenigen aus vergleichbaren Studien, was auf eine kürzere Behandlung und eine schwerer erkrankte Patient\*innenstichprobe zurückgeführt werden kann. Die Ergebnisse, dass ein niedrigerer Schulabschluss und ein niedrigeres Alter einen vorzeitigem Therapieabbruch vorhersagen, replizieren Befunde von vergleichbaren deutschen Studien mit großen Stichproben. Die positiven Effekte niedrigerer Affektregulationsfähigkeiten auf residualisierte symptomsspezifische Veränderung liefern Hinweise auf die Wichtigkeit von Affektregulationsfähigkeiten als ein Veränderungsmechanismus von DBT. Die niedrige Vorhersagekraft, die die Ergebnisse limitiert, könnte auf die Heterogenität der Störung und untersuchten Stichprobe zurückzuführen sein. Obwohl der explorative Variablenauswahlprozess bisherige Forschung untermauert, legen die Ergebnisse aus der Validierungsprozedur nahe, dass es ratsam ist, sich mit vorschnellen Urteilen über die Behandlungsverläufe einzelner Patient\*innen basierend auf klinischen Daten und Selbstauskünften zurückzuhalten. Ferner weisen die Ergebnisse darauf hin, dass die Personalisierung von Behandlungsangeboten auf die individuellen Bedürfnisse einer Patientin mit BPS nicht ausschließlich auf soziodemographischen, klinischen und psychologischen Baseline-Daten beruhen sollte. Zukünftige Studien sollten sich aufgrund der Heterogenität des Störungsbildes auf Prädiktorenanalysen im Rahmen von Subgruppen unter Einbezug von Follow-up-Daten zur Bestimmung des langfristigen Therapieoutcomes konzentrieren.

## 4.2 Studie 2: Wirksamkeit und Prädiktoren einer stationären Psychotherapie für Posttraumatische Belastungsstörungen

**Zitation:** Herzog, P., Voderholzer, U., Gärtner, T., Osen, B., Svitak, M., Doerr, R., Rolvering-Dijkstra, M., Feldmann, M., Rief, W., & Brakemeier, E.-L. (in revision). Predictors of outcome during inpatient psychotherapy for posttraumatic stress disorder: A single-treatment, multi-site, practice-based study, *Psychotherapy Research*.

**Hintergrund:** Die Wirksamkeit von trauma-fokussierter kognitiver Verhaltenstherapie zur Behandlung der Posttraumatischen Belastungsstörung (PTBS) wurde in randomisiert-kontrollierten Studien und in ambulanten Therapieprogrammen nachgewiesen. An Studien zur Bestimmung der Wirksamkeit unter stationären Routinebedingungen mangelt es hingegen. Zudem profitieren nicht alle Patient\*innen von diesen Psychotherapien, einige zeigen substanzielle Residualsymptome, brechen die Therapie vorzeitig ab oder erleben Rückfälle. Die Forschung zur Bestimmung von stabilen Prädiktoren des Behandlungserfolgs bei PTBS zeigt insgesamt nur heterogene Ergebnisse. Daher besteht das Ziel der vorliegenden Studie darin, die Wirksamkeit von stationärer Psychotherapie der PTBS in der Routineversorgung zu bestimmen und Prädiktoren des Behandlungsergebnisses zu identifizieren.

**Methode:** Ein routinemäßig erhobener Datensatz aus 612 Patient\*innen mit der Hauptdiagnose der PTBS (M = 42.3 Jahre [SD = 11.6], 75.7% weiblich) wird analysiert. Die Patient\*innen wurden mit traumafokussierter Psychotherapie mit Schwerpunkt auf kognitiv-behavioralen Prinzipien in verschiedenen deutschen psychosomatischen Krankenhäusern behandelt. Das primäre Outcome wird als symptombezogene Veränderung (IES-R) festgelegt, als sekundäre Outcomes werden Veränderungsmaße in der Funktionsfähigkeit (GAF), Ängstlichkeit und Depression verwendet. Hedges' g-korrigierte Prä-post-Effektstärken werden für alle Outcomes berechnet. „Elastic net“-Regularisierung als datengetriebener, stabilitätsbasierender Machine-Learning-Ansatz wurde verwendet, um ein stabiles klinisches Vorhersagemodell auf dem klinischen und dem funktionellen Outcome zu identifizieren.

**Ergebnisse:** Die Prä-Post-Effektstärken reichen von Hedges' g = .58 - 1.42 und deuten auf mittlere bis große Effekte auf allen Outcomes hin. Die Ergebnisse der Prädiktoranalysen verdeutlichen, dass ein kombiniertes Vorhersagemodell aus soziodemographischen, klinischen und psychometrischen Variablen zur Vorhersage von verschiedenen Outcomes beiträgt. Über

das klinische und funktionelle Outcome hinweg zeigt sich, dass Psychotizismus, die Anzahl an Diagnosen und eine Asthma-Diagnose konstant einen negativen prädiktiven Zusammenhang zum Behandlungsergebnis aufweist. Auf seiner klinischen Signifikanz erscheint die erklärte Varianz durch das Modell mit der besten Passung zufriedenstellend.

**Diskussion:** Traumafokussierte Psychotherapie basierend auf kognitive-behavioralen Prinzipien kann effektiv in die stationäre Routineversorgung in Deutschland implementiert werden. Einige prognostisch relevante Variablen können dabei identifiziert werden, welche das Behandlungsergebnis vorhersagen, z. B. die Anzahl an Diagnosen und eine Asthma-Diagnose. Sollte das vorgeschlagene klinische multivariate Prädiktorenmodell in konfirmatorischen kontrollierten Studien (z. B. RCTs) repliziert werden, können diese Variablen helfen, die Behandlung noch mehr auf die Bedürfnisse und Probleme der individuellen Patientin zu personalisieren, um so langfristig das Behandlungsergebnis zu optimieren. Zukünftige Studien sollten hingegen der Unterteilung zwischen PTBS und komplexer PTBS (kPTBS) z. B. durch den Einbezug einer umfassenderen störungsspezifischen Diagnostik Rechnung tragen.

### 4.3 Studie 3: Wirksamkeit und Prädiktoren einer stationären Psychotherapie für Zwangsstörungen

**Zitation:** Herzog, P., Osen, B., Stierle, C., Middendorf, T., Voderholzer, U., Koch, S., Feldmann, M., Rief, W., & Brakemeier, E.-L. (under review). Determining prognostic variables of treatment outcome in obsessive-compulsive disorder: Effectiveness and its predictors in routine clinical care, *Behavior Therapy*.

**Hintergrund:** Trotz hoher Effektstärken in randomisiert-kontrollierten Studien v.a. bzgl. der Expositionstherapie mit Reaktionsmanagement profitieren nicht alle Patient\*innen mit Zwangsstörungen von diesen Psychotherapien (z. B. substanzielle Residualsymptome, Therapieabbruch, Rückfall). Die Ergebnisse von Studien zu Prädiktoren des Therapieoutcomes sind zudem uneinheitlich. In der vorliegenden Studie werden daher sowohl die naturalistische Wirksamkeit untersucht als auch Prädiktoren des Therapieoutcomes bei Patient\*innen mit Zwangsstörungen identifiziert.

**Methode:** Ein routinemäßig erhobener Datensatz aus 1.595 Patient\*innen mit der Hauptdiagnose einer Zwangsstörung (M = 33.9 Jahre [SD = 11.7], 60.4% weiblich) wird analysiert. Die Patient\*innen wurden mit Psychotherapie mit Schwerpunkt auf kognitiv-behavioralen Prinzipien in verschiedenen deutschen psychosomatischen Krankenhäusern behandelt. Die Prä-Post-Effektstärken (ES) werden für verschiedene Outcomevariablen (Y-BOCS, OCI-R, BSI, BDI-II, PHQ-9, GAD-7, PHQ-15, GAF, SWLS) in der Gesamtstichprobe (N = 1.595) und einer Substichprobe (N = 514) berechnet. Die Reduktion auf diese Substichprobe basiert auf der Verfügbarkeit von Daten zu potentiellen Prädiktoren (z. B. OCI-R), die für die Fragestellung relevant sind und welche nur in einer Klinik verwendet wurden. In dieser Substichprobe werden Prädiktoranalysen auf ein klinisches und funktionelles Outcome durch regressionsanalytische Ansätze durchgeführt. Der potentielle Pool an Prädiktoren wird vorher faktorenanalytisch reduziert.

**Ergebnisse:** Die Prä-Post-Effektstärken in unterschiedlichen Outcomevariablen können insgesamt als groß eingeordnet werden (z. B. ES auf primären Outcome Y-BOCS: Hedges'  $g = 1.34$ ). Als Prädiktoren für die Symptomveränderung wird die Symptomschwere (Aufnahme) und psychopathologische Belastung identifiziert. Es zeigt sich zudem, dass insbesondere Patient\*innen mit einer höheren sozialen Unterstützung und mehr Waschzwängen eine



größere Symptomveränderung zeigen. Die Subanalysen zeigen ein distinktes Prädiktorenprofil in Bezug auf die Veränderung von Zwangsgedanken im Vergleich zu Zwangshandlungen.

**Diskussion:** Die Ergebnisse deuten darauf hin, dass eine evidenzbasierte Psychotherapie basierend auf kognitiv-behavioralen Prinzipien der Zwangsstörung in der stationären Routineversorgung wirksam implementiert werden kann. Neben bekannten Prädiktoren des Therapieoutcomes (Symptomschwere bei Aufnahme, psychopathologische Belastung) konnten soziale Unterstützung und Waschzwänge als wichtige positive Prädiktoren identifiziert werden. Die Identifizierung solcher prognostischer Variablen und die Ermittlung von spezifischen Prädiktorenprofilen bergen die Chance, bereits zu Beginn der Psychotherapie spezifische Interventionen auf die Bedürfnisse der Patient\*innen abzustimmen, um den Behandlungserfolg zu optimieren. Zukünftige Studien sollten preskriptive Variablen im Sinne von Moderatoren des Therapieoutcomes für verschiedene Psychotherapien (z. B. Kognitive Therapie vs. Expositionstherapie mit Reaktionsverhinderung) unter standardisierteren Bedingungen untersuchen.

#### 4.4 Studie 4: Wirksamkeit und Prädiktoren einer stationären Psychotherapie für depressive Störungen

**Zitation:** Herzog, P., Feldmann, M., Kube, T., Langs, G., Gärtner, T., Rauh, E., Doerr, R., Hillert, A., Voderholzer, U., Rief, W., Endres, D., & Brakemeier, E.-L. (under review). Inpatient psychotherapy for depression in a large routine psychosomatic care sample: A Bayesian approach to examining effectiveness and predictors of change, *Journal of Consulting and Clinical Psychology*.

**Hintergrund:** Obwohl es viele Belege für die Wirksamkeit von Psychotherapie bei Depressionen unter randomisiert-kontrollierten Bedingungen gibt, untersuchen nur wenige große Phase IV-Studien die Wirksamkeit einer Depressionsbehandlung sowie die zugrundeliegenden Prädiktoren einer Veränderung unter naturalistischen Bedingungen. Daher verfolgt die vorliegende Studie die Ziele (1) die Wirksamkeit einer nicht-manualisierten Psychotherapie für Depression in einem naturalistischen stationären Setting zu untersuchen, und (2) potentielle Prädiktoren für diese Veränderung zu identifizieren.

**Methode:** In einer großen Stichprobe von  $N = 22.681$  stationären Patient\*innen mit der Hauptdiagnose einer Depression ( $M = 47.3$  Jahre [ $SD = 12.7$ ], 58.1% weiblich) werden Prä-Post- und Prä-Follow-up-Effektstärken (Hedges'  $g$ ) für verschiedene Outcomemaße berechnet. Die Patient\*innen wurden mit KVT-basierter Psychotherapie in verschiedenen deutschen psychosomatischen Krankenhäusern behandelt. Als primäres Outcome wird der BDI-II bzw. PHQ-9, als sekundäre Outcomes der BSI, die SF-36 und die SWLS verwendet. Zur Bestimmung eines probabilistischen Modells von Prädiktoren für diese Veränderung werden die Daten in mehreren Teilschritten analysiert. Um statistische Artefakte durch die Struktur der fehlenden Daten aus allen Kliniken zu vermeiden, werden nur Daten aus einer Klinik verwendet. Anhand einer Substichprobe von  $N = 6.377$  Patient\*innen mit Depression wird zuerst eine unabhängige Komponentenanalyse berechnet, um Komponenten aus den demographischen und klinischen Daten zu bestimmen. Anschließend wird eine Bayessche exploratorische Faktorenanalyse durchgeführt, um Faktoren aus den verfügbaren psychometrischen Fragebogendaten vom Prätest, Posttest und Follow-up-Zeitpunkt zu generieren. Zur Auswahl des Modells mit der besten Anpassungsgüte wurden verschiedene Bayessche Pfadmodelle berechnet. Schließlich

wurde eine Bayessche Pfadanalyse durchgeführt, um die wichtigsten Faktoren zur Vorhersage von Veränderungen zu identifizieren.

**Ergebnisse:** Für das primäre Outcome können große ( $g = 1.04 - 1.32$ ) und für die sekundären Outcomes mittlere ( $g = .40 - .99$ ) Prä-Post-Effektstärken identifiziert werden. Prä-Follow-up-Effektstärken sind für das primäre Outcome groß ( $g = .74 - 1.25$ ) und klein für das sekundäre Outcome ( $g = .31$ ). Fast alle Faktoren vor Behandlungsbeginn (mit Ausnahme des "aktuellen Rentenwunsches") beeinflussen zu Beginn der Behandlung die psychopathologische Gesamtbelastung auf unterschiedlichen Faktoren signifikant, während sich ihr direkter Einfluss im Verlauf der Behandlung abschwächt. Verschiedene Faktoren erweisen sich als veränderungsresistent während der Behandlung, wie z. B. Suizidalität, Agoraphobische Ängste, Lebensunzufriedenheit, physische Beeinträchtigung und Schmerzen, was auf stabile psychopathologische Faktoren hinweist. Die stärksten Einflüsse können von Suizidalität zu Behandlungsbeginn auf negative Kognitionen bei Behandlungsende, von agoraphobischen Symptomen bei Behandlungsbeginn auf die Ängstlichkeit bei Behandlungsende und von der körperlichen Beeinträchtigung bei Behandlungsbeginn auf die wahrgenommene Beeinträchtigung bei Behandlungsende beobachtet werden. Die Auswirkungen der klinischen Faktoren zu Behandlungsende auf die klinischen Faktoren im Follow-up sind insgesamt eher gering.

**Diskussion:** Die Ergebnisse deuten auf große Effekte von Psychotherapie bei Depression in der stationären Routineversorgung in Deutschland hin. Unsere Analysen decken dabei Faktoren, die während der Psychotherapie schwierig zu verändern sind und so den Therapieerfolg beeinträchtigen (z. B. Suizidalität), auf. Darüber hinaus deuten die Ergebnisse auf die gegenseitige Beeinflussung zwischen gewissen Symptombereichen über die Zeit hin (z. B. Suizidalität auf negative Kognitionen). Der Einbezug von solchen Faktoren in probabilistischen Modelle zur Vorhersage von kurz- und langfristigen Veränderungen durch die Depressionsbehandlung könnte die Personalisierung von Interventionen an die spezifischen Probleme und Bedürfnisse von Patient\*innen erleichtern und so die Schwierigkeiten bei der Optimierung des Therapieoutcomes in der stationären Routineversorgung überwinden. Zukünftige Studien sollten differentielle Prädiktoren für verschiedene Psychotherapien (z. B. KVT vs. IPT) unter kontrollierten Bedingungen untersuchen, um so durch die Erstellung von Entscheidungsbäumen die Behandlungsempfehlung zu erleichtern.

#### 4.5 Studie 5: Erfassung von negativen Effekten in der Psychotherapie

**Zitation:** Herzog, P., Lauff, S., Rief, W., & Brakemeier, E.-L. (2019). Assessing the unwanted: A systematic review of instruments used to assess negative effects of psychotherapy. *Brain and Behavior*, 9(12). <https://doi.org/10.1002/brb3.1447>

**Hintergrund:** Während die Wirksamkeit der Psychotherapie bei der Behandlung psychischer Störungen bereits seit Jahrzehnten gut untersucht wird, sind negative Effekte der Psychotherapie vergleichsweise selten betrachtet worden. Dies könnte unter anderem daran liegen, dass es kein anerkanntes Instrument zur Erfassung der negativen Effekte gibt, welches weltweit als ‚Goldstandard‘ genutzt wird. Daher wurden im Rahmen eines systematischen Reviews zunächst die existierenden Instrumente zur Erfassung negativer Effekte der Psychotherapie hinsichtlich ihrer Gemeinsamkeiten und Unterschiede beleuchtet, um anschließend einen gemeinsamen konzeptuellen Rahmen zu definieren und Empfehlungen zur Verbesserung der Erfassung abzuleiten.

**Methode:** Das Studienauswahlverfahren folgte den aktuellen Best-Practice-Richtlinien für die Durchführung systematischer Übersichtsarbeiten. Eingeschlossen wurden Studien, die auf Englisch und Deutsch zum Thema veröffentlicht wurden. Ausgeschlossen wurden Studien, deren vollständiger Text nicht verfügbar war. Um eine möglichst ausführliche Literatursuche durchzuführen, wurden Studien nicht wegen fehlender psychometrischer Gütekriterien, dem Herkunftsort der Studie, dem Jahr der Publikation oder der Stichprobe bzw. der Art der untersuchten psychischen Störung ausgeschlossen. In drei Datenbanken (PsycInfo, PubMed, Web of Science) konnten 10 Studien identifiziert werden. Die neun darin beschriebenen Instrumente wurden jeweils kritisch auf ihre theoretische Ausrichtung (bzw. die erfassten Bereiche), ihre psychometrischen Eigenschaften (Validität, Reliabilität) und ihre diagnostischen Überlegungen (z. B. Item-Sensitivität, Anzahl der Items) hin überprüft. Alle eingeschlossenen Studien wurden von den ersten beiden Autoren unabhängig voneinander kodiert, um eine vollständige Extraktion relevanter Merkmale der einzelnen Instrumente zu ermöglichen und eine Gegenprüfung zu gewährleisten.

**Ergebnisse:** Innerhalb der neun Instrumente können siebzehn Bereiche, welche mit negativen Effekten der Psychotherapie assoziiert sind, identifiziert werden, die jedoch im Rahmen der neun Instrumente uneinheitlich berücksichtigt werden. Die meisten Instrumente

liefern lediglich erste Daten über ihre psychometrischen Eigenschaften. Im Hinblick auf diagnostische Überlegungen werden verschiedene Item-Response-Formate verwendet, oft jedoch wird die Attribution zur Psychotherapie erfragt, welche einen wichtigen Faktor darzustellen scheint.

**Diskussion:** Das systematische Review zeigt, dass die bestehenden Instrumente zur Messung der negativen Auswirkungen der Psychotherapie zwar ein breites Spektrum relevanter Bereiche abdecken, es jedoch keinen Konsens über die wichtigsten Bereiche gibt. Ihre psychometrischen Eigenschaften sind meist unbefriedigend. Als Ergebnis des Reviews werden eine konsensuale Definition und Konzeptualisierung negativer Effekte vorgeschlagen. Negative Effekten werden in diesem Rahmen definiert als unerwünschte Ereignisse, die durch die Psychotherapie verursacht werden, und in Nebenwirkungen und therapeutisches Fehlverhalten bzw. unethisches Verhalten unterteilt werden können. Nebenwirkungen im Speziellen sind negative Effekte, die durch eine korrekt ‚lege artis‘ durchgeführte Psychotherapie ausgelöst werden. Daraus können Empfehlungen zur Verbesserung der Erfassung von negativen Effekten in der Psychotherapie abgeleitet werden, die für die zukünftige Forschung nützlich sein könnten. Beispielsweise sollte die Erfassung von Nebenwirkungen verschiedene wichtige Bereiche (z. B. Stigmatisierung, therapeutische Beziehung) umfassen, das Ausmaß der Belastung dadurch bestimmen, die Attribution zur Psychotherapie explizit herstellen und Items mit positiven Effekten beinhalten.

#### 4.6 Studie 6: Zusammenhang zwischen dem Auftreten von negativen Effekten und dem Outcome eines stationären CBASP-Programms für chronische Depression

**Zitation:** Herzog, P., Häusler, S., Normann, C., & Brakemeier, E.-L. (under review). Negative effects of an inpatient CBASP program: Frequencies and their impact on treatment outcome in chronic depression, *Frontiers in Psychiatry (Psychological Therapies)*.

**Hintergrund:** Eine immer größer werdende Anzahl an Studien deutet daraufhin, dass das Cognitive Behavioral Analysis System of Psychotherapy (CBASP) eine effektive Form der Psychotherapie in der Behandlung von chronischer Depression darstellt. Hingegen gibt es noch wenige Studien, welche die negativen Effekte von CBASP beleuchten. Daher ist das Ziel dieser Studie, die Häufigkeit von negativen Effekten eines stationären CBASP-Programms zu untersuchen und deren Zusammenhang zum Behandlungsergebnis herzustellen.

**Methode:** 52 Patient\*innen mit chronischer Depression (CD) und Behandlungsresistenz (M = 48.1 Jahre [SD = 10.1], 61.5% weiblich) wurden in einem strukturierten 12-wöchigen stationären CBASP-Programm behandelt. Während der Follow-up-Datenerhebung füllten die eingeschlossenen Patient\*innen retrospektiv das „Inventory for the Assessment of Negative Effects of Psychotherapy“ (INEP) aus. Der Schweregrad der depressiven Symptomatik wurde mittels Selbst- und Fremdrating (BDI-II und HRDS-24) während der Aufnahme, Entlassung und im 6-Monats-Follow-up erfasst. Die Häufigkeiten von negativen Effekten wurde berechnet und eine binäre logistische Regressionsanalyse wurde durchgeführt, um den Zusammenhang zum Behandlungsergebnis herzustellen.

**Ergebnisse:** 92.3% aller Patient\*innen mit chronischer Depression berichtet von mindestens einem negativen Effekt durch das stationäre CBASP-Programm und 45.2% berichten von einer Abhängigkeit von der Therapeutin. Stigmatisierung und finanzielle Sorgen sowie intrapersonelle Veränderungen wurden von ca. einem Drittel berichtet. Nur eine wahrgenommene Abhängigkeit von der Therapeutin steht in einem negativen Zusammenhang zum Behandlungsergebnis in beiden Outcomemessungen.

**Diskussion:** Fast alle Patient\*innen berichten von negativen Effekten ihrer Psychotherapie, deren Häufigkeit höher als in anderen Studien in einem ähnlichen stationären Setting ist. Dieser Unterschied könnte auf die spezifischen Charakteristika von Patient\*innen mit behandlungsresistenter CD, d.h. schwere Symptome, frühes Onset (Alter < 21 Jahre),

Suizidalität und hoher Prozentsatz an Misshandlungserfahrungen in der Kindheit, sowie das hochdosierte kurze stationäre CBASP-Programm mit einem starken Fokus auf negative Beziehungserfahrungen in der Kindheit und die Therapeut\*-Patient\*in-Beziehung zurückgeführt werden. Die meisten aller berichteten negativen Effekte scheinen sich jedoch nicht negativ auf das Behandlungsergebnis auszuwirken. Nichtsdestotrotz scheint eine wahrgenommene Abhängigkeit von der Therapeut\*in einen negativen Einfluss auf das Behandlungsergebnis auszuüben. Insofern diese Befunde in randomisiert-kontrollierten Studien repliziert werden können, sollten Therapeut\*innen sich dessen im Rahmen von CBASP bewusst sein und in der Therapie adressieren (z. B. durch die gezielte Förderung der Autonomie frühzeitig in der Therapie).

### 5 Zusammenfassende Diskussion und Ausblick

Das Ziel der vorliegenden Dissertation besteht darin, im Rahmen von Phase IV-Studien sowohl die Wirksamkeit (im Sinne der „effectiveness“) von Psychotherapie bei verschiedenen psychischen Störungen in der stationären Routineversorgung, d.h. unter naturalistischen Bedingungen im alltäglichen Arbeitskontext, zu untersuchen als auch durch Prädiktorenanalysen Ansatzpunkte zur Optimierung dieser gemittelten Effekten zu identifizieren. Darüber hinaus sollen Risiken und Nebenwirkungen von Psychotherapie näher beleuchtet werden.

Die Schätzung von Prä-Post-Effektstärken anhand von großen Datensätzen in der stationären Routineversorgung ermöglicht es, eine Art Benchmark für stationäre Psychotherapie von Menschen mit psychischen Erkrankungen zu erhalten sowie Hinweise darüber zu bekommen, wie gut eine evidenzbasierte Psychotherapie im naturalistischen Setting implementiert und disseminiert wurde (Cook, Schwartz, & Kaslow, 2017; McHugh & Barlow, 2010)<sup>14</sup>. Daher leisten die Ergebnisse der Studien 1-4 einen wichtigen Beitrag zur absoluten Wirksamkeitsforschung von stationärer Psychotherapie in Deutschland im Sinne einer Phase IV-Studie.

Trotz zum Teil hoher Effektstärken für Psychotherapie – insbesondere KVT – bewegen sich die Nonresponse-Raten, d.h. eine nicht klinisch signifikante Verbesserung der Symptomatik, im Bereich bis zu 50% aller behandelten Patient\*innen (Lambert, 2013; vgl. Abschnitt 2.3.3.). Daher wurden in der vorliegenden Dissertation in einem zweiten Schritt Prädiktoren für den Behandlungserfolg mit Hilfe verschiedener statistischer Verfahren identifiziert, um den Behandlungserfolg für solche Patient\*innensubpopulationen, die bisher auf diese evidenzbasierten Psychotherapieverfahren noch nicht ausreichend ansprechen, zu optimieren. Durch die Identifikation robuster Prädiktoren soll ermöglicht werden, die Effekte der angebotenen stationären Psychotherapie durch eine bessere Abstimmung der Psychotherapie auf die einzelne Patientin zu erhöhen.

Die Ergebnisse der Studien 1-4 sollen langfristig zur praxisbasierten Evidenz beitragen (Barkham et al., 2010; Barkham & Margison, 2008; Holmqvist et al., 2015). Durch

---

<sup>14</sup> Als Implementierung wird der Transfer von EBP in das klinische Alltagssetting bezeichnet, als Disseminierung die Streuung von EBP über das Land hinweg.



Rückmeldeschleifen an die klinisch-tätigen Praktiker\*innen, was durch Publikationen, Kongressbeiträge und gezielte Vorträge in den psychosomatischen Kliniken erreicht werden soll, wird der Forderung nach einer praxisorientierten Forschung Rechnung getragen. Langfristig ist geplant, dass die Ergebnisse der Analysen dieser großen Datensätze zu einem evidenzbasierten Outcome-Monitoring für zukünftige Patient\*innen beitragen, wodurch auch das Praxis-Forschungsnetzwerk in Deutschland gestärkt wird.

Zusätzlich zu positiven Effekten und dem Ausbleiben positiver Effekte werden in dieser Dissertation negative Effekte von Psychotherapie untersucht. Dazu wird im Rahmen eines systematischen Reviews zunächst die Erfassung von negativen Effekten in der Psychotherapie beleuchtet, indem auch definitorische und konzeptuelle Fragen aufgeworfen werden sowie als Resultat ein konzeptueller Rahmen geschaffen und Ansatzpunkte zur Verbesserung der Messung von negativen Effekten in der Psychotherapie vorgeschlagen werden. Basierend auf diesen Ergebnissen werden Daten analysiert, welche das Auftreten von negativen Effekten mit dem Therapieoutcome eines stationären CBASP-Programms für chronische Depression in Verbindung bringen.

### 5.1 Zusammenfassung der Ergebnisse im Rahmen der Leitfragen

Die vorliegende Dissertation hat sich im Rahmen einer praxisorientierten Forschung (Castonguay & Muran, 2016) dabei konkret an den folgenden Leitfragen orientiert (vgl. Abschnitt 3.1.):

1. Lassen sich die Ergebnisse aus der Wirksamkeitsforschung auf ein angewandtes stationäres Setting generalisieren?
2. Durch welche Variablen und Faktoren lassen sich unterschiedliche Outcomemaße zur Bestimmung des Psychotherapieerfolgs bei verschiedenen psychischen Störungen vorhersagen?
3. Wie können negative Effekte und Nebenwirkungen von Psychotherapien definiert, erfasst und deren Relevanz zum Therapieoutcome beurteilt werden?

Im Folgenden sollen die wichtigsten Befunde in integrierter Form zusammenfassend dargestellt werden.

Bezüglich der Frage 1 deuten die Studien 1-4 in der Zusammenschau darauf hin, dass spezifische evidenzbasierte stationäre Psychotherapieprogramme wirksam implementiert werden können, jedoch in ihrer naturalistischen Wirksamkeit (im Sinne der „effectiveness“) Unterschiede aufweisen. Während die Effektstärken für die Behandlung von Zwangsstörungen (Studie 3) und depressiven Störungen (Studie 4) – vergleichbar mit anderen ähnlichen Studien für Depression (Gyani et al., 2013; Karlin et al., 2012; Knapstad, Nordgreen, & Smith, 2018) und Zwangsstörungen (Gönner et al., 2012; Grøtthe et al., 2018) - als groß beurteilt werden können, sind sie für die Behandlung von PTBS (Studie 2) mittel bis groß. Diese Befunde sind dennoch vergleichbar mit anderen PTBS-Studien (Eftekhari et al., 2013) und sogar z. T. größer als in einer aktuellen Benchmark-Studie bei Jugendlichen mit PTBS (Rudd et al., 2019). Die starke psychopathologische Beeinträchtigung und die hohe Anzahl und Schwere der komorbiden Störungen haben möglicherweise eine größere symptom-spezifische Veränderung in dieser Stichprobe verhindert, was darauf hindeutet, dass die Stichprobe eine große Anzahl von Patienten mit komplexer PTBS (kPTBS) umfassen könnte. Im Einklang damit zeigen psychologische Interventionen bei kPTBS geringere Effekte als bei PTBS (Karatzias et al., 2019). Zudem ist die durchschnittliche Aufenthaltsdauer mit 54.3 Tagen deutlich kürzer als in anderen Studien, z. B. durchschnittlich 74.8 Tage in einer anderen stationären Routinebehandlung (Kratzer et al., 2018). Für die Behandlung von Borderline-Persönlichkeitsstörungen (Studie 1) werden kleine bis mittelgroße Effekte gefunden, wodurch das stationäre nicht-manualisierte DBT-Programm etwas weniger effektiv erscheint im Vergleich zu einer Studie von Kröger et al. (2013), welche im Durchschnitt mittelgroße Effekte identifiziert hat. Gleichfalls haben nur 22.9% der Patient\*innen die Behandlung vor dem geplanten Ende des Behandlungsprogramms verlassen, was unter der von McMurrin et al. (2010) berichteten mittleren Dropout-Rate von 37% liegt. Zählt man nur die Abbrüche, die nicht in gegenseitigem Einverständnis von Patient\*in und Team erfolgten, beträgt die Dropout-Rate 12.2%, was mit der von Kröger et al. (2013) berichteten Rate von 10% vergleichbar ist.

Bei den jeweiligen Prädiktorenanalysen können unterschiedliche Variablen identifiziert werden, die im Zusammenhang mit dem Therapieoutcome zu stehen scheinen: Bei der DBT-Behandlung von Patient\*innen mit Borderline-Persönlichkeitsstörung werden in Studie 1 unterschiedliche Prädiktoren für unterschiedliche Outcomemaße identifiziert. Ein höherer Schulabschluss und eine komorbide rezidivierende depressive Störung sind die Hauptprädiktoren für einen regulären Behandlungsabschluss. Niedrige

Affektregulationsfähigkeiten und keine vorherige ambulante Psychotherapie sagen dabei größere symptomspezifische Veränderungen vorher. Was die Vorhersage des symptomspezifischen Outcomes betrifft, so scheint der Ausgangswert dieses Outcomes wichtiger zu sein als jede andere routinemäßig erhobene Variable. Ein jüngeres Alter, vorhandene Arbeitsfähigkeit, hohe Werte in emotionalen und physischen Rollenlimitationen und geringerer körperlicher Schmerz stehen in prädiktivem Zusammenhang mit starken Verbesserungen im funktionellen Outcome. Im Vergleich zu Patient\*innen in einer anderen Studie ist die Stichprobe aus Studie 1 hingegen in mehrfacher Hinsicht stärker beeinträchtigt (z. B. höhere durchschnittliche Anzahl psychiatrischer Diagnosen, höhere Anzahl komorbider rezidivierender depressiver Störungen, niedrigeres allgemeines Funktionsniveau zu Behandlungsbeginn) und die Behandlungsdauer ist in Studie 1 etwa 2,5 Wochen kürzer als in anderen vergleichbaren Studien (Kröger et al., 2013).

Bei der Behandlung von PTBS zeigt sich in Studie 2 über das klinische und funktionelle Outcome hinweg, dass insbesondere Psychotizismus, die Anzahl an Diagnosen und eine Asthma-Diagnose konstant einen negativen prädiktiven Zusammenhang zum Behandlungsergebnis darstellen. Die Stichprobe ist in Bezug auf das Alter und den Schweregrad der depressiven Symptome mit einer Stichprobe einer ähnlichen Studie vergleichbar (Deisenhofer et al., 2018). Der störungsspezifische Symptomschweregrad ist nur geringfügig niedriger als bei einer Stichprobe von stationären Patient\*innen mit komplexeren Formen von PTBS (IES-R; Median = 90.9, Kratzer et al., 2018).

Bei der Behandlung von Zwangsstörungen können in Studie 3 neben bekannten Prädiktoren des Outcomes wie z.B. die Symptomschwere bei Aufnahme und psychopathologische Belastung (Kohls, Bents, & Pietrowsky, 2002; Kyrios, Hordern, & Fassnacht, 2015; Raffin, Guimarães Fachel, Ferrão, Pasquoto de Souza, & Cordioli, 2009; Steketee, Eisen, Dyck, Warshaw, & Rasmussen, 1999), auch soziale Unterstützung und Waschzwänge als wichtige Prädiktoren identifiziert werden. Die Stichprobe in dieser Studie scheint zudem typisch für eine Klinik in Deutschland zu sein: In einer weiteren Stichprobe aus zwei deutschen stationären Behandlungseinrichtungen wird ein ähnlicher Mittelwert für Alter, Geschlechtsverteilung und Y-BOCS berichtet, allerdings mit einem etwas höheren OCI-R-Score und einem höheren Anteil komorbider Persönlichkeitsstörungen (Voderholzer et al., 2013).

Im Hinblick auf die Prädiktoren von kurz- und langfristigen Veränderungen durch die Behandlung von Depressionen zeichnet sich in Studie 4 ein komplexeres Bild: Faktoren vor der Behandlung, wie soziodemographische und klinische Variablen im Sinne von Patient\*inneneigenschaften, beeinflussen die Symptomschwere auf verschiedenen Ebenen zu Beginn der Behandlung signifikant, aber ihr direkter Einfluss wird im Verlauf der Behandlung und darüber hinaus fast vollständig abgeschwächt. Zudem erscheinen während der Behandlung von depressiven Störungen einige Faktoren veränderungsresistent und tragen signifikant zur Stabilität der Psychopathologie bei, wie z. B. Suizidalität, agoraphobische Ängste, Lebensunzufriedenheit, physische Beeinträchtigung und Schmerzen. Weiterhin kann der Einfluss von agoraphobischen Tendenzen zu Behandlungsbeginn auf die Ängstlichkeit zu Behandlungsende, der Einfluss von körperlicher Beeinträchtigung bei Behandlungsbeginn auf die wahrgenommene Beeinträchtigung bei Behandlungsende sowie der Einfluss von Suizidalität auf die negativen Kognitionen nach der Behandlung gezeigt werden. Die Studie gibt erste Hinweise, was der „Motor der Psychopathologie“ bei depressiven Störungen sein könnte.

Zusammenfassend lässt sich in Bezug auf die Studien 1-4 festhalten, dass zum Teil dieselben Prädiktoren (z. B. Symptomschwere zur Aufnahme) wie in RCTs identifiziert werden können, zum Teil jedoch in diesen Stichproben von stationären Psychotherapiepatient\*innen auch andere Variablen eine Rolle zur Vorhersage des Therapieoutcomes spielen. Das liegt wahrscheinlich am stationären Behandlungssetting und seiner Spezifika (z. B. höhere Therapiedosis in kürzerer Zeitspanne) und der dortigen Patient\*innenpopulation, welche insgesamt eine höhere Krankheitsschwere aufweist. Weitere Erklärungen ergeben sich daraus, dass bestimmte Variablen (z. B. durch eine ausführliche medizinische Diagnostik) typischerweise bei der Planung, Konzeption und Durchführung von RCTs keine Rolle spielen bzw. gar nicht erst erhoben werden (z. B. aufgrund der hohen Zahl an Ein- und Ausschlusskriterien). Um die Wirksamkeit einer Intervention gut beurteilen zu können, sind RCTs zweifelsohne wichtig; die Patient\*innenpopulationen unterscheiden sich jedoch oft zu denen unter Routinebedingungen. Aus den Ergebnissen der vorliegenden Prädiktorenanalysen lässt sich somit ableiten, dass diese relevanten Faktoren zwar unter den gleichen stationären Routinebedingungen adressiert werden sollten, sie jedoch in zukünftigen RCTs mit erhoben, im besten Falle repliziert und anschließend genutzt werden können, z. B. können sie als Variablen zur Stratifizierung in RCTs dienen (Steyerberg, 2019). Die Ergebnisse der Studien 1-4 der vorliegenden Dissertation haben verschiedene Prädiktoren identifiziert, die unter realen

Versorgungsbedingungen den Behandlungserfolg einer stationären Psychotherapie für verschiedene psychische Störungen vorhersagen, welche jedoch in zukünftigen Studien bestätigt werden sollten.

Die Ergebnisse des systematischen Reviews in Studie 5 zeigen, dass die bestehenden Instrumente zur Messung der negativen Effekte von Psychotherapie zwar ein breites Spektrum relevanter Bereiche abdecken, jedoch aufgrund unterschiedlicher definitorischer und klassifikatorischer Aspekte keinen Konsens über die korrekte Erfassung von negativen Effekten erzielen und ihre psychometrischen Eigenschaften meist noch unbefriedigend sind. Basierend auf den Ergebnissen konnte eine konsensuale Definition von negativen Effekten formuliert werden. Demnach werden negative Effekte in Nebenwirkungen und therapeutisches Fehlverhalten bzw. unethisches Verhalten unterteilt, und im Allgemeinen definiert als unerwünschte Ereignisse während oder nach der Psychotherapie, die auf die Psychotherapie attribuiert werden können. Nebenwirkungen im Speziellen werden dabei definiert als negative Effekte, die durch eine ‚lege artis‘ (i.e. korrekt) durchgeführte Psychotherapie ausgelöst werden. Das Review leitet daraus Empfehlungen für die Erfassung von Nebenwirkungen in der Psychotherapie (z. B. wichtige Bereiche sowie den daraus resultierenden Belastungsgrad erfassen und die Attribution zur Psychotherapie herstellen) ab, welche zukünftige Forschungsbemühungen in diesem Feld stimulieren kann.

Um die Relevanz von negativen Effekten von Psychotherapie zu bestimmen, soll deren Auftreten zum Therapieoutcome in Zusammenhang gebracht werden. In Studie 6 wird gezeigt, dass zwar fast alle Patient\*innen eines stationären CBASP-Programms von negativen Effekten ihrer Psychotherapie berichten, die meisten aller berichteten negativen Effekte hingegen nicht in einem negativen Zusammenhang zum Behandlungoutcome stehen. Es gibt jedoch eine wichtige Ausnahme: eine wahrgenommene Abhängigkeit von der Therapeutin scheint einen signifikanten negativen Einfluss auf das Behandlungoutcome auszuüben. In einer anderen Studie konnte ebenso anhand einer studentischen Stichprobe der negative Zusammenhang zwischen einigen Facetten der Abhängigkeit und psychologischen Behandlungen hergestellt werden (Geurtzen, Keijsers, Karremans, & Hutschemaekers, 2019).

### 5.2 Stärken und Einschränkungen

#### Allgemeine Stärken

Die durchgeführten Studien weisen einige Stärken auf, nicht zuletzt da sie praxisbasierte Evidenz liefern, das Praxis-Forschungsnetzwerk stärken und so im Rahmen von Phase IV-Studien mit hoher externer und ökologischer Validität zur praxisorientierten Forschung in Deutschland beitragen. Die größte Stärke der Studien 1-4 besteht in der Stichprobengröße sowie der Anwendung statistisch komplexer Auswertungsstrategien, welche viele Vorteile gegenüber klassischen Analyseverfahren aufweisen. Vergleichsweise selten werden solche großen Stichproben – insbesondere im Verhältnis zur Prävalenz der jeweiligen psychischen Störung – mit detailliert beschriebenen Stichprobencharakteristika zur Analyse einbezogen: In die Studie 1 konnten Daten von 878 Patient\*innen mit Borderline-Persönlichkeitsstörungen, in die Studie 2 Daten von 612 Patient\*innen mit Posttraumatischer Belastungsstörung, in die Studie 3 Daten von 1.595 bzw. 514 Patient\*innen mit Zwangsstörungen und in die Studie 4 Daten von 22.681 bzw. 6.377 Patient\*innen mit depressiven Störungen genutzt werden. Dadurch können robuste und replizierbare Aussagen über die Wirksamkeit von stationärer Psychotherapie in der Routineversorgung in Deutschland getroffen werden, die als Benchmark für zukünftige Forschung herangezogen werden können. Um die komplexe Pathologie einer schwer beeinträchtigten, stationären Stichprobe von Patient\*innen mit diversen psychischen Störungen gerecht zu werden, werden neben einer Vielzahl an soziodemographischen und klinischen Daten auch alle berichteten medizinischen Diagnosen in die Analysen einbezogen. Vorteilhaft ist ebenso, dass die Outcomes in den Studien 1-4 unter Einbezug unterschiedlicher Symptombereiche und unterschiedlicher Bereiche des Funktionsniveaus gebildet werden (Roth & Fonagy, 2005).

#### Spezifische Stärken

Die Stärke von Studie 1 ist die Trennung eines explorativen Verfahrens zur Variablenselektion auf residualisierten Outcomes über den Baselinewert hinaus unter Nutzung einer Bootstrap-gesteigerten Regression sowie die anschließende Validierung der Ergebnisse in einer Holdout-Stichprobe mit wiederholter Kreuzvalidierung (Murphy, 2012). Ebenso vorteilhaft ist die Kombination aus verschiedenen Outcomemaßen, i.e. die Verwendung von vorzeitigem Therapieabbruch sowie symptom-spezifischen und funktionalen Maßen als

Outcome (Panos et al., 2014). Eine weiterer Stärke dieser Studie ist, dass die Qualitätssicherung der Behandlung in der betreffenden Klinik durch die DBT-Zertifizierung der Station geschieht, die ein jährliches DBT-Teamtraining durch eine offiziell anerkannte externe DBT-zertifizierte Supervisorin und die Einhaltung der DBT-Adhärenz durch die regelmäßige DBT-Supervision der Stationsleitung umfasst.

Die Stärke der Studie 2 besteht in der Verwendung eines Machine-Learning-Ansatzes – „Elastic net“-Regularisierung – um Prädiktoren zu identifizieren, wie es auch Dwyer et al. (2018) empfohlen haben und zurzeit vermehrt Anwendung findet (Webb et al., 2020). Zudem werden solide Module zur Vorverarbeitung verwendet, um die Daten so zu bereinigen, dass ein höherer Grad an Homogenität erreicht wird. Zudem erscheint die Aufteilung der Stichprobe in ein Trainings- und Testset bzw. Validierungsset zur Evaluierung der Leistung des ausgewählten Modells als ein Vorteil dieser Studie (Murphy, 2012). Die Tatsache, dass die erklärten Varianzen bei dem Trainings- als auch Testset vergleichbar sind, unterstützt die Annahme, dass das sog. „Overfitting“ kein Problem darstellt, und erhöht die Chancen, ähnliche Prädiktoren in einem vergleichbaren anderen Datensatz zu finden. Durch die Verwendung eines störungsspezifischen Messinstruments – dem IES-R (Maercker & Schützwohl, 1998) – können Limitationen vergleichbarer Studien (z. B. Deisenhofer et al., 2018) überwunden werden.

Die expliziten Stärken der Studie 3 ist die Reduktion von vielen potentiellen Prädiktoren auf wenige distinkte, gut interpretierbare Dimensionen mit Hilfe einer „state-of-the-art“ Faktorenanalyse (SPCA, vgl. Erichson et al., 2018) als datengetriebenes Vorgehen sowie einer anschließenden regressionsanalytischen Methode. Weiterhin vorteilhaft ist der Einsatz eines störungsspezifischen Selbstbeurteilungsinstruments für Zwangsstörungen – dem Y-BOCS-SR (Federici et al., 2010) – als primäres Outcome.

Die Studie 4 verwendetet Bayessche Statistik, um Prädiktoren für kurz- und langfristige Symptomveränderungen während und nach der Behandlung zu identifizieren. Ein großer Vorteil davon ist, dass die sog. „Bayesian inference“ es ermöglicht, die Unsicherheit in den Daten direkt abbilden zu können (z. B. Høifødt et al., 2015). In der Wissenschaft wird für die Verwendung von derartigen ausgereifteren statistischen Auswertungsmethoden in der klinischen Psychologie und Psychotherapie vermehrt geworben (Craske, 2019). Als Ergebnis der Studie 4 wird ein probabilistisches klinisches multivariates Modell der Prädiktoren für Symptomveränderungen in der Depressionsbehandlung vorgeschlagen, das die Heterogenität

und Komplexität der stationären Patient\*innen mit depressiven Störungen berücksichtigt, indem es mehrere Datenquellen zu potentiellen Prädiktoren (z. B. soziodemographische, klinische und psychometrische Daten) in die Analysen einbezieht (Iniesta et al., 2016).

Die Studien 5-6 untersuchen ein lang vernachlässigtes und bisher wenig beforschtes Thema in der Psychotherapie: Negative Effekte bzw. Nebenwirkungen. Studie 5 untersuchte im Rahmen eines systematischen Reviews Messinstrumente zur Erfassung von negativen Effekten in der Psychotherapie. Diese Studie ist die erste ihrer Art, welche im Rahmen eines Reviews negative Effekte von Psychotherapie beleuchtet, wodurch sich hier eine explizite Stärke der Studie 5 ergibt. Dabei stößt sie auf uneinheitliche Definitionen und Konzeptualisierungen und versucht, im Sinne eines „bottom-up“-Ansatzes einen Rahmen zur Klassifikation von negativen Effekten zu liefern, der zukünftige Forschung stimulieren kann. Gleichzeitig liegt die Stärke der Studie 5 in der Ableitung von Empfehlungen zur Verbesserung der Erfassung von negativen Effekten in der Psychotherapie, welche eine Stoßrichtung für zukünftige Forschungsbemühungen in diesem Bereich geben kann. Studie 6 verfolgte das Ziel, den Zusammenhang von negativen Effekten zum Therapieoutcome bei einem stationären CBASP-Programm für chronische Depression herzustellen. Ebenso als eine der ersten Studien ihrer Art (z. B. Meister et al., 2019), beleuchtet sie die Auswirkungen von aufgetretenen negativen Effekten während bzw. nach der Psychotherapie, worin eine weitere explizite Stärke gesehen werden kann.

### Allgemeine Einschränkungen

Eine allgemeine Einschränkung der Studien 1-4 besteht darin, dass für die Analysen ein naturalistischer retrospektiver Routinedatensatz von Patient\*innen mit psychischen Erkrankungen, die in psychosomatischen Kliniken behandelt wurden, genutzt wurde. Naturalistische Studien sind durch eine geringere interne Validität insbesondere im Vergleich zum Goldstandard der Therapiestudien – den RCTs (siehe CONSORT-Kriterien: Pittler et al., 2011) – gekennzeichnet. Generell ist daher anzumerken, dass jegliche Symptomveränderung außerhalb von RCTs nicht eindeutig kausal auf die psychotherapeutische Behandlung, in Abgrenzung zu Zeit- und Situationseinflüssen, zurückzuführen ist.

Zu den Limitationen der Studien 1-4 zählt demnach vor allem die fehlende Randomisierung bei Behandlungszuweisung (z. B. unbekannte aber relevante Patient\*innen-



und/oder Kliniker\*innenmerkmale könnten zur Wahl der Behandlung beigetragen haben), keine Verwendung einer Kontrollgruppe (z. B. eine Wartegruppe als Vergleich zur Überprüfung des natürlichen Verlaufs der Symptome), keine Verwendung eines einheitlichen Therapiemanuals in den verschiedenen Kliniken, nur eine teilweise Dokumentation der besuchten Therapiebausteine, angewandten Strategien und Medikamente, meist fehlende Überprüfung der Therapieadhärenz sowie nur teilweise Informationen zur Therapeut\*innenkompetenz. Es kann nicht direkt gesagt werden, welche Strategien und Techniken in welcher Intensität und in welchem Setting (Einzel- oder Gruppensetting) tatsächlich eingesetzt wurden. Streng genommen könnte daher nur von Prädiktoren der Veränderung von Symptomen über die Zeit gesprochen werden, wobei die Veränderung aus einer Kombination aus Behandlung und natürlichem Verlauf der Symptome resultiert. Die Symptome nehmen an Schwere zu und ab, oft als Reaktion auf Lebensstressoren. Der Rückgang der Symptome hat möglicherweise wenig mit der Behandlung zu tun. Es könnte theoretisch sein, dass die Symptome vorübergehend reduziert werden, indem die Patient\*innen aus ihrem alltäglichen Umfeld entfernt und in eine stressfreie stationäre Umgebung gebracht werden. Die Ergebnisse zeigen daher nicht explizit und zweifelsfrei, dass die Behandlung wirksam ist, da es keine Kontrollgruppe gibt. Man kann nur sagen, dass sich die Patient\*innen während der Therapie verändert haben, aber die Ursachen für die Veränderung bleiben zu einem gewissen Grad unklar. Das heißt wiederum, es ist auch bis zu einem gewissen Grad unsicher, ob die Prädiktoren zuverlässig die Ergebnisse für andere Behandlungsprotokolle vorhersagen, die in unterschiedlichen Formaten und in anderen Situationen angeboten werden.

Eine weitere Einschränkung der Studien 1-4 ist daher auch die reduzierte Übertragbarkeit bzw. Generalisierbarkeit auf Behandlungssettings außerhalb von Deutschland. Stationäre Psychotherapie wird vor allem in Deutschland angeboten, die meisten internationalen Psychotherapiestudien basieren auf der Untersuchung ambulanter Psychotherapie. Die Hospitalisierung von psychisch erkrankten Menschen liegt in der spezifischen Tradition des deutschen Gesundheitsversorgungssystems (Reimer, 2007). Dies kann als Einschränkung insofern angesehen werden, als dass es die internationale Relevanz und den Impact teilweise reduziert. Hingegen können wir von einer hohen externen Validität in den Studien 1-4 zumindest für ein stationäres Psychotherapiesetting aufgrund von sehr geringen Ein- und Ausschlusskriterien ausgehen. Dadurch können wir annehmen, dass der Heterogenität der diversen Störungsbilder besser Rechnung getragen wurde. Letztlich können die Ergebnisse der

Studien 1-4 zumindest als relativ aussagekräftig für eine KVT-orientierte stationäre psychotherapeutische Behandlung von schwer psychisch erkrankten Patient\*innen in einer psychosomatischen Klinik in Deutschland angesehen werden, als Benchmark der Wirksamkeit dieses Behandlungsformats dienen und somit unter Umständen einen ‚Incentive‘ geben, solche Behandlungsformen in anderen Ländern zu implementieren. Hierzu benötigt man insbesondere jedoch noch gesundheitsökonomische Maße, um auch die Kosten-Nutzen-Effektivität zu bestimmen. In diesem Zusammenhang kann kritisiert werden, dass mangelnde Klarheit über die Psychotherapie selbst herrscht. Die Patient\*innen erhielten ein Behandlungspaket anstatt eine spezifischen Behandlung, sodass die Effekte auch durch Gruppentherapien und mehrere Therapeut\*innen, die in der Gesamtbehandlung beteiligt waren, potentiell beeinflusst wurden. Ebenso sind keine Daten zur Medikation und anderen psychopharmakologischen Interventionen zur Aufnahme und im Verlauf verfügbar, wofür also nicht kontrolliert werden kann. Weiterhin gibt es in dem der Studien 1-4 zugrundeliegenden Datensatz keine Angaben zur Behandlungstreue, so dass die Therapeut\*inneneffekte nicht kontrolliert werden konnten. Es fehlen meistens Informationen über die Anzahl der Therapeut\*innen, ihre Kompetenz, ihr Ausbildungsniveau und ihre therapeutische Erfahrung sowie in den Studien 2-4 ebenso über das Monitoring der Therapie (z. B. durch Supervision), um die Therapiequalität noch besser zu bestimmen. Auch das Patient\*innenengagement und die Patient\*innencompliance zur Behandlung wird nicht erhoben. Es hat sich jedoch gezeigt, dass die Behandlungstreue durch Therapeut\*innenadhärenz<sup>15</sup> und -kompetenz<sup>16</sup> nur in geringem bis keinem Ausmaß zum Therapieoutcome steht (Webb, DeRubeis, & Barber, 2010).

Eine weitere Limitation der Studien 1-4 ist die Verwendung von Selbsteinschätzungsfragebögen als primäre Outcomemaße, welche aufgrund der Zeit- und Kosteneffizienz im täglichen Arbeitsalltag im Vergleich zu Fremdbeurteilungen genutzt werden. Das ist aus methodischen Gründen problematisch: die Messreliabilität ist methodenabhängig, d.h. Benchmarks sind immer instrumentenspezifisch. Da in internationalen klinischen Studien oft Fremdbeurteilungsinstrumente eingesetzt werden, wäre der Einbezug derselben für das Setzen von Benchmarks aufgrund der höheren Vergleichbarkeit besser. Die Outcomemessung kann allerdings in Reaktivität und Spezifität varriieren, so sind bspw. Fremdbeurteilungsinstrumente durch Kliniker\*innen in solchen Studien häufig reaktiver als

---

<sup>15</sup> Adhärenz: Einhaltung des Behandlungsprotokolls

<sup>16</sup> Kompetenzratings der Durchführung bestimmter Behandlungen

Selbstbeurteilungsinstrumente durch Patient\*innen und Instrumente mit einer höheren Spezifität (z. B. symptom-spezifische Messinstrumente ergeben größere Effekte; Barkham et al., 2010; Minami et al., 2007). Zudem sind die Diagnosen des Datensatzes der Studien 1-4 durch die jeweilige Bezugstherapeutin erhoben und nicht durch eine unabhängige, verblindete Raterin. Die Interrater-Reliabilität für die Diagnosen ist aufgrund des naturalistischen Designs nicht verfügbar. Durch die Verwendung von ausschließlich Selbst- statt Fremdbeurteilungsinstrumenten ist zusätzlich auch die Zuverlässigkeit, mit der die Diagnose untermauert wird (z. B. IES-R bei PTBS, BSL-95 bei Borderline-Persönlichkeitsstörungen, Y-BOCS-SR bei Zwangsstörungen und BDI-II bei depressiven Störungen), reduziert. Einige Forscher\*innen könnten bei den Studien 1-4 kritisch die Verwendung eines kontinuierlichen statt binären bzw. dichotomen Outcomes (z. B. klinisch signifikante Veränderungen; Patient\*innen reliabel verbessert) diskutieren. In der Tat ist das binäre bzw. dichotome Outcome oft besser interpretierbar und kommunizierbar. Aus statistischen Gründen (u.a. höhere Power) entscheiden wir uns im Hinblick auf unsere naturalistische Stichprobe jedoch für ein kontinuierliches Outcome (Steyerberg, 2019). Eine weitere Einschränkung der Studien 1-4 ist aufgrund von nur einer Behandlungsbedingung die Verwendung des LOCF als Prä-Post-Effektstärkenschätzung. Dieser Ansatz ist sehr konservativ und wird aufgrund seiner Nachteile normalerweise nicht empfohlen (Siddiqui et al., 2009).

Eine weitere Einschränkung der Studien 1-4 besteht darin, dass das Benchmarking nicht auf statistischer Ebene vollzogen wurde. Einige Forscher\*innen entwickelten eine Range-Null-Strategie (Minami, Serlin, et al., 2008; Minami, Wampold, et al., 2008; Minami et al., 2007), die auf Serlins „Good-Enough“-Methode beruht (Serlin & Lapsley, 1985, 1993), bei der festgestellt wird, ob ein erzielter Effekt innerhalb einer Range des Benchmarks liegt. Das Problem dabei ist jedoch auch, dass der Range arbiträr gewählt wird (liberales vs. konservatives Kriterium), d.h. die Definition eines klinisch unbedeutsamen Unterschieds ist von der Forscherin abhängig. Beispielsweise betrachteten die Forscher\*innen beim Vergleich einer Studie mit den gebildeten Benchmarks für die Wirksamkeit einer Psychotherapie die Schätzungen der Effektstärke, die maximal 10 Prozent unter den Benchmarks lagen, als klinisch äquivalent, wohingegen sie beim Vergleich mit den Benchmarks für den natürlichen Verlauf (d.h. Kontrollgruppe ohne Behandlung) die Effektstärken diese Bedingung um mindestens  $d = 0.20$  übertreffen musste (Minami et al., 2007). In unseren Studien kann kein Test auf solche klinische Äquivalenz (Rogers et al., 1993) bzw. Signifikanztest auf Unterschiede durchgeführt werden, da bisher noch keine

relevanten Vergleichsdaten aus RCTs zu stationärer Psychotherapie vorliegen, gegen die ein Äquivalenz- bzw. Signifikanztest sinnvoll gewesen wäre.

Insgesamt fehlen in den Studien 1-4 wichtige Daten zur besseren Modellierung (z. B. Anzahl an Therapeut\*innen, Anzahl an beteiligten Stationen, genaue Angabe aller tatsächlich durchgeführten Behandlungsmodalitäten, etc.), wodurch es nicht möglich ist, mehrere Analyselevel einzubeziehen und statistische Techniken im Bereich des sog. „Multilevel-Modelling“ anzuwenden, z. B. Analyse von Therapeut\*inneneffekte bei Case-Mix-Adjustment-Studien (Brown, Lambert, Jones, & Minami, 2005; Lutz, Martinovich, Howard, & Leon, 2002; Mullin, Barkham, Mothersole, Bewick, & Kinder, 2006). Auch bei der Effekstärkenberechnung wäre eine Nestung der Daten wichtig zu beachten (Wampold & Serlin, 2000). Es scheint nämlich so, dass einige Therapeut\*innen bessere Outcomes erzielen als andere Therapeut\*innen, wodurch sich die Frage nach der „richtigen“ Benchmark ergibt. Solche oft hierarchischen linearen Modelle könnten genesteten Daten besser gerecht werden und die Aussagekraft erhöhen, z. B. könnte ein 3-Level-Modell (Nestung von Erhebungen in Patient\*innen, Patient\*innen in Therapeut\*innen/Stationen/Teams und diese wiederum Kliniken) analysiert werden. Um den direkten kompetitiven Vergleich zwischen den Kliniken zu vermeiden, werden in den Studien 1-4 keine vergleichenden Analysen zwischen Kliniken durchgeführt, wodurch insbesondere in den Studien 2-4 nicht für den sog. „Study-site-Effekt“ kontrolliert werden konnte, was als ein Nachteil anzusehen ist (Barkham et al., 2010).

Die erklärte Varianz unserer klinischen Prädiktionsmodelle in den Studien 1-4 lässt noch Spielraum für Verbesserungen, was aufgrund der Stabilitätsanalysen unter Nutzung statistischer Replikation weniger ein statistisches Problem, als viel mehr ein inhaltliches Problem darzustellen scheint. Zwar kann auf einen großen Pool an soziodemographischen, klinischen und psychometrischen Variablen in den Studien 1-4 zurückgegriffen werden, die Analysen sind jedoch limitiert genau auf diesen Variablenpool und können somit keine Variablen, die sich als wichtige Prädiktoren in anderen empirischen Studien erwiesen, mit in die Berechnungen einschließen. Transdiagnostisch relevante Variablen, also über psychische Störungen hinweg, könnten Erwartungen (Kube et al., 2018; Kube, Herzog, et al., 2019; Rief et al., 2015) und Kindesmisshandlung (R.-Mercier, Masson, Bussièrès, & Cellard, 2018) sein. In Studie 1 könnten darüber hinaus auch Prozessvariablen und die therapeutische Beziehung (Barnicot et al., 2012) sowie die Anzahl an vorherigen Suizidversuchen (Rüsch et al., 2008), in

Studie 2 die Behandlungsresistenz und Chronizität (Dunlop, Kaye, Youngner, & Rothbaum, 2014) und in Studie 3 und 4 ebenso die therapeutische Beziehung (Vogel, Hansen, Stiles, & Göttestam, 2006), motivationale Prozesse (de Haan et al., 1997; Reid et al., 2017; Steketee et al., 2011) und Patient\*innenadhärenz (Maher et al., 2012; Simpson et al., 2011) eine Rolle gespielt haben. Da der Baseline-Wert des jeweiligen Outcomes in den Studien 2-3 ein wichtiger Prädiktor zu sein scheint, sind Decken- und Bodeneffekte sowie Effekte aufgrund der Regression zur Mitte schwer zu schätzen und könnten zu den Ergebnissen beigetragen haben. Die Ergebnisse der Studien 1-4 beruhen ausschließlich auf den Daten von Kliniker\*innen und Patient\*innen. Eine kürzlich veröffentlichte Studie (Lee et al., 2018) hat gezeigt, dass Modelle mit mehreren Datenquellen (d.h. auch Neuroimaging und genetische Daten) genauer und adäquater sind als Modelle mit einem einzigen niedrig-dimensionalen Datentyp (d.h. ausschließlich phänomenologische Daten). In diesem Zusammenhang behaupten die Autor\*innen der Studie, dass es unwahrscheinlich ist, dass nur klinische Marker - selbst in allen klinischen Beurteilungsbereichen - die Auswahl oder Anpassung der Behandlung lenken können, da die biologische Heterogenität, die zugrundeliegende Pathologie und die daraus resultierenden Phänomenologien, d.h. die Variabilität der Symptompräsentation, bei Depressionen sehr unterschiedlich sind (vgl. Trivedi, 2013).

Weitere Einschränkungen der Studien 1-3 bezieht sich auf den explorativen Charakter der Analysen und den fehlenden Follow-up-Messzeitpunkt. Es können daher keine Aussagen über langzeitige Symptomveränderungen (z. B. Rückfälle oder Stabilität der Behandlungseffekte) gemacht werden, die aufgrund des spezifischen Charakters der stationären Umgebung von Bedeutung sein könnten. Insbesondere bei der Behandlung von Zwangsstörungen im Rahmen der Expositionstherapie mit Reaktionsverhinderung in Studie 3, könnten Zwangshandlungen und Rituale einfach in Abhängigkeit von der kontrollierten Umgebung verhindert werden (ähnlich der Bedeutung der Abwesenheit von Alkoholkonsum als primäres Outcome bei der stationären Behandlung von Alkoholmissbrauch und -abhängigkeit).

### Spezifische Einschränkungen

Eine Einschränkung der Studie 1 ist die fehlende Information von nicht-suizidalem selbstverletzendem Verhalten, welches in Studien mit Borderline-Patient\*innen unter DBT-Behandlung oft als primäres Outcome verwendet wird. Dieses Outcome wäre auch durch Fremdbeobachtung relativ einfach sowie zeit- und kostenökonomisch zu generieren. Die

Verwendung anderer Outcomemaße, wie diejenigen, die in Studie 1 verwendet wurden, wurde jedoch kürzlich von mehreren Autor\*innen empfohlen (Panos et al., 2014).

Eine Einschränkung der Studie 2 bezieht sich auf die Verwendung der IES-R (Creamer, Bell, & Failla, 2003; A Maercker & Schützwohl, 1998) als primäres Outcome. Zum einen basiert diese Skala noch auf dem DSM-IV (und nicht auf dem DSM-5), zum anderen wird sie eher in Deutschland und selten in internationalen klinischen Studien eingesetzt. Um jedoch eine bessere internationale Vergleichbarkeit herzustellen und den aktuellen Entwicklungen im Zuge des DSM-5 gerecht zu werden, könnte sich die Verwendung der „Posttraumatic Diagnostic Scale (PDS-5)“ (Foa et al., 2016) oder die „PTSD Checklist for DSM-5 (PCL-5)“ (Blevins, Weathers, Davis, Witte, & Domino, 2015; Bovin et al., 2016; Wortmann et al., 2016) eignen. Es wäre wünschenswert, auch noch die „Clinician-Administered PTSD Scale for DSM-5 (CAPS-5)“ (Weathers et al., 2018) als Fremdbeurteilungsinstrument der Symptomschwere und insbesondere Absicherung der Diagnose mitzuerheben. Andererseits besitzt die IES-R (Creamer et al., 2003; Maercker & Schützwohl, 1998) eine hohe Sensitivität. Die Verwendung des GAF in Studie 2 kann zudem als ein Einzeliteminstrument kritisiert werden, da die Interrater-Reliabilität nicht verfügbar ist und die Reliabilität generell eher gering ist. Die Verwendung einfacher, einmaliger globaler Skalen kann jedoch nützlich sein, da die klinische Synthese heterogener Komponenten oft besser zu diskriminieren scheint als komplexe Multi-Item-Messungen (Fava, Evins, Dorer, & Schoenfeld, 2003). Trotz der Relevanz der ICD-10 ist der GAF besser in der Lage, relevante Bereiche des ICF zu erfassen, d.h. Beeinträchtigung der Funktionalität durch psychischen Störungen, welche insbesondere in diesem stationären Setting eine große Bedeutung zukommt (Linden, 2017).

Eine gemeinsame Einschränkung der Studie 1-2 ist das fehlende sog. „Downsizing“ des Prädiktorenpools a priori, da hier der gesamte Pool potentieller Prädiktoren verwendet wird. Studien deuten darauf hin, dass eine Reduzierung des Pools potenzieller Prädiktoren vor der Datenanalyse aus mehreren Gründen nützlich sein könnte: Das Risiko der Unsicherheit in Bezug auf einen einzelnen Prädiktor ist bei einer erhöhten Anzahl von potentiellen Prädiktoren höher, insbesondere in linearen Standard-Regressionsmodellen (Riley, Snell, Ensor, Burke, Harrell, et al., 2019; Riley, Snell, Ensor, Burke, Harrell Jr, et al., 2019). Wann immer möglich, sollte eine vorherige Reduzierung des Prädiktorenpools erfolgen, die jedoch auf einer überzeugenden theoretischen Grundlage beruhen muss, um willkürliche Entscheidungen zu vermeiden. Eine

empirische Kürzung des Prädiktorenpools könnte nämlich stichprobenspezifisch und ebenso atheoretisch sein und die Prädiktoren schwierig replizierbar machen. In dem Stadium des wissenschaftlichen Prozesses bei diesen beiden Studien können die generierten Prädiktoren nun teilweise als wissenschaftliche Grundlage für die Auswahl und strengere Prüfung von Prädiktorvariablen in der Zukunft dienen. Zudem hätte unter Umständen die Bildung von Subgruppen durch Stratifizierung für einige wichtige Variablen (z. B. komorbide Depression) auch den Vorhersagewert der klinischen Prädiktionsmodelle in Studie 2 verbessern können, während jedoch gleichzeitig die statistische Aussagekraft durch eine verringerte Stichprobengröße der einzelnen Untergruppen verringert werden würde. Aufgrund fehlender empirischer Robustheit für solche Variablen und dem Vorteil großer Stichproben wurde sich in Studie 2 dagegen entschieden.

Eine Einschränkung der Studie 3 ist die Verwendung des Selbstratings der Y-BOCS, auch hier wäre die Verwendung der Fremdratingversion wünschenswert (Goodman et al., 1989). Gleichfalls ist daher auch die ausschließliche Verwendung des BDI-II (Beck et al., 1996) bzw. PHQ-9 (Kroenke et al., 2001) ohne zusätzliches Fremdratinginstruments wie z. B. der Skala „Hamilton Rating Scale for Depression (HRSD)“ (Hamilton, 1960) bei Studie 4 eine Limitation. Es hat sich allerdings gezeigt, dass die Verwendung der HRSD signifikant größere Effekte als entsprechende Selbstbeurteilungsinstrumente wie z. B. der BDI-II produziert (Cuijpers, Li, Hofmann, & Andersson, 2010; Lambert, Hatch, Kingston, & Edwards, 1986), wodurch die Studie 4 die Effekte tendenziell eher unterschätzt hat. Eine andere Studie weist hingegen darauf hin, dass es zwischen Selbst- und Frembeurteilungsverfahren für Depression zwar Unterschiede bei der Response-Rate gibt, allerdings insgesamt große Effekte für alle Skalen mit geringer Variabilität zwischen den Skalen in der Routineversorgung gefunden werden (Zimmerman, Walsh, Friedman, Boerescu, & Attiullah, 2018).

Eine weitere Einschränkung der Studie 4 bezieht sich darauf, dass nur ‚Granger-kausale‘ Schlussfolgerungen aus den Analysen und aufgrund des ‚Cross-Lagged-Panel‘-Designs keine kausalen Schlussfolgerungen direkt aus den Ergebnissen gezogen werden können. Durch mehrfache wiederholte Messungen und deren Einbezug in das Cross-Lagged-Modell werden zwei Voraussetzungen für die Herstellung von Kausalzusammenhängen erfüllt: (1) Herstellung eines Zusammenhangs zwischen den untersuchten Variablen, und (2) die Berücksichtigung der zeitlichen Abfolge der Prozesse (z. B. die Ursache muss vor dem Ergebnis eintreten). Eine solche

Assoziation zwischen Variablen, bei der eine Variable X zukünftige Werte einer anderen Variable Y vorhersagt, wird als "Granger-Kausal" bezeichnet (Granger, 1969; Schuurman, Ferrer, de Boer-Sonnenschein, & Hamaker, 2016). Das bedeutet, dass nicht gemessene Konstrukte mit den in dieser Studie gemessenen Konstrukten in Beziehung stehen und als Konfundierungsvariable und wahre Ursachen für die Veränderung anderer Faktoren fungieren könnten.

Eine wesentliche Einschränkung von Studie 5 ist, dass die Studien- und Datenlage relativ begrenzt ist. Daher konnten nur wenig relevante Studien zur Synthese in das systematische Review miteinbezogen werden (d.h. seit 1986 nur 32 Publikationen mit standardisierten Umfragemethoden). Hinzu kommt die Heterogenität der Messinstrumente, wodurch sich die Aussagekraft zusätzlich eingeschränkt. Klinimetrische Eigenschaften wurden zudem nicht berücksichtigt, die bei der Evaluation von Messinstrumenten in der klinischen Psychologie und Psychotherapieforschung zusätzlich einbezogen werden sollten (Bech, 2016; Fava, Rafanelli, & Tomba, 2012).

Eine Einschränkung der Studie 6 ist neben der relativ kleinen Stichprobe und einer zusätzlichen algorithmus-basierten Pharmakotherapie die retrospektive Erhebung des INEP (Ladwig et al., 2014) in einem Follow-up-Zeitpunkt, der zwischen 6-12 Monaten variiert. Gedächtnisverzerrung, Vergesslichkeit, Wiederherstellungsfehler oder wichtige Erfahrungen, die nach Abschluss der Behandlung aufgetreten sein könnten, könnten die berichteten negativen Effekte unterschiedlich verzerrt haben. Darüber hinaus muss bei der Interpretation der Ergebnisse dieser Studie die Rückblickverzerrung berücksichtigt werden, d.h., dass Patient\*innen, die nicht auf die Behandlung angesprochen haben, ihre Behandlung in einem weniger positiven Licht sehen und mehr negative Effekte berichten. Zudem besitzt der INEP keine spezifischen Items für das stationäre Setting (z. B. Gruppentherapien) und erfasst auch keine positiven Effekte wie z. B. der PANEPS (Moritz et al., 2015, 2018). Auch der Mangel einer Kontrollgruppe lässt die Frage offen, ob die Ergebnisse wirklich CBASP-spezifisch sind.



### 5.3 Implikationen für die klinische Praxis

Aus den Ergebnissen der vorliegenden Dissertation lassen sich vielfältige Implikationen für die klinische Praxis ableiten. Die vorgeschlagenen klinischen Prädiktionsmodelle aus den Studien 1-4 könnten Praktiker\*innen in psychosomatischen Kliniken in Deutschland helfen, diejenigen Patient\*innen zu identifizieren, bei denen das Risiko eines geringeren Therapieoutcomes besteht, d.h. diese identifizierten Prädiktoren könnten bereits bei der Behandlungsplanung als Mittel zur Optimierung des Therapieoutcomes berücksichtigt werden. Im Folgenden werden die klinischen Implikationen der Studien 1-6 diskutiert, wobei insbesondere die Ergebnisse aufgegriffen werden, welche auch durch andere Studien unterstützt werden:

Die Ergebnisse bezüglich der stationären Psychotherapie von Borderline-Persönlichkeitsstörungen im Rahmen eines DBT-Programms deuten darauf hin, dass niedrige Affektregulationsfertigkeiten und keine vorherige ambulante Psychotherapie symptomspezifische Veränderungen positiv vorhersagen. In diesem Sinne kann es ratsam sein, dass Patient\*innen mit bereits stärker vorhandenen Affektregulationsfertigkeiten und Therapievorerfahrung anderen Psychotherapieverfahren als DBT zugewiesen werden. Diese Patient\*innen erscheinen bereits ihre Affekte regulieren zu können, was sie möglicherweise in vorherigen Therapien mit Skillstraining gelernt haben, da Skillstraining direkt Affektregulationsfertigkeiten adressiert (Swenson et al., 2001). Dennoch erscheinen sie zu einem gewissen Grad therapieresistent, so dass andere evidenzbasierte psychologische Therapien (Sollberger & Walter, 2010), wie die Mentalisierungsbasierte Therapie (MBT; Bateman & Fonagy, 2009; Bateman & Fonagy, 2009; Bateman & Fonagy, 2010; Jørgensen et al., 2013; Jørgensen et al., 2014) oder Schematherapie (ST; Giesen-Bloo et al., 2006; Kellogg & Young, 2006; Sempértegui, Karreman, Arntz, & Bekker, 2013; Young, Klosko, & Weishaar, 2003) hier indiziert sein könnten, welche beide auch ihre Wirksamkeit in der Behandlung von BPS gezeigt haben (siehe auch das aktuelle Cochrane-Review: Storebø et al., 2020), jedoch mehr die Kindheitsmisshandlungen adressieren und weniger die Affektregulationsfertigkeiten. Ein höherer Bildungsgrad und eine komorbide schwere, rezidivierende depressive Störung erwiesen sich in der vorliegenden Studie als stabile Prädiktoren für einen regulären Therapieabschluss. Eine andere Studie ergab ebenfalls, dass ein niedrigeres Bildungsniveau einen vorzeitigen Therapieabbruch vorhersagt (Perroud, Uher, Dieben, Nicastro, & Huguelet,

2010). In der Studie von Kröger et al. (2013) sagte ebenso ein niedrigeres Alter den vorzeitigen Therapieabbruch vorher. Im Einklang damit wurde bei einer kleinen Stichprobe von Frauen, die die Therapie abgeschlossen haben, eine höhere interne Motivation festgestellt als bei denjenigen, die die Therapie nicht abgeschlossen haben (Webb & McMurrin, 2009). Daher sollte den Patient\*innen möglichst transparent erklärt werden, welche Auswirkungen ein vorzeitiger Therapieabbruch haben kann. Ein jüngeres Alter sowie vorhandene Arbeitsfähigkeit standen in einem prädiktiven Zusammenhang mit starken Verbesserungen im funktionellen Outcome. Es sollte daher überlegt werden, wie vorhandene effektive Behandlungsangebote Patient\*innen mit Borderline-Persönlichkeitsstörungen noch früher erreichen, sodass man Chronifizierungsprozessen frühzeitig präventiv entgegenwirken kann.

Die Ergebnisse der Studie 2 legen nahe, dass bei der stationären Psychotherapie von Patient\*innen mit PTBS körperliche Erkrankungen bereits bei der Behandlungsplanung mit berücksichtigt werden sollten, da z. B. eine Asthma-Diagnose negative Auswirkungen auf das Behandlungsergebnis zu haben scheint. Asthma wurde mit einer höheren PTBS-Komorbidität in Verbindung gebracht (Shiratori & Samuelson, 2012), während Asthmapatient\*innen mit PTBS auch eine höhere Rate der Inanspruchnahme von stationären Gesundheitsleistungen aufwiesen (Mindlis et al., 2017). Eine PTBS kann auch ein Risikofaktor für die Entwicklung von Asthma (Hung et al., 2019) und für die Verschlimmerung von bereits bestehendem Asthma verantwortlich sein (Udner et al., 2019). Insgesamt könnte eine Asthma-Diagnose ein Warnzeichen für ein schlechteres Ansprechen zu Beginn der stationären Psychotherapie darstellen und sollte im Rahmen eines kontinuierlichen Monitorings beobachtet werden sowie anschließend ggf. zu einer Anpassung auf den individuellen Behandlungsplan führen.

In den Analysen zur stationären Psychotherapie von Zwangsstörungen hat sich neben replizierten Prädiktoren die soziale Unterstützung als ein positiver Prädiktor für den Therapieerfolg gezeigt. Dieser Befund weist einmal mehr darauf hin, dass Angehörige von Patient\*innen mit Zwangsstörungen aktiv mit in die Behandlung eingebunden werden sollten. Dieser Befund steht im Einklang mit den Ergebnissen einer kürzlich erschienenen systematischen Übersicht (Palardy, El-Baalbaki, Fredette, Rizkallah, & Guay, 2018), indem gezeigt wurde, dass negative soziale Unterstützung offenbar mit schwereren Symptomen verbunden ist, während positive soziale Unterstützung für Patient\*innen von Vorteil sein könnte. Dies steht auch im Einklang mit einer älteren Studie, welche ebenso die Qualität der

sozialen Unterstützung hervorhob (Van Noppen & Steketee, 2009). Daher unterstreichen diese Ergebnisse erneut die bedeutsame Rolle der Angehörigen bei der Unterstützung von Patient\*innen mit Zwangsstörungen. Es könnte daher vorteilhaft sein, entweder die Angehörigen von Patient\*innen mit Zwangsstörungen in den Behandlungsprozess einzubeziehen, indem ihnen nützliche Informationen, z. B. Psychoedukation zum Umgang mit der Zwangserkrankung und ihren Symptomen im Alltag, zur Verfügung gestellt werden, oder die Qualität der Beziehungen durch systemische oder familiäre Interventionen zu stärken (Sydow & Borst, 2018). Darüber hinaus ergeben sich durch die Zwangserkrankung oft Konflikte mit Angehörigen, die mit Strategien aus der IPT bearbeitet werden könnten (Schramm, 2019) zu stärken. Dies könnte auch zu einer Verringerung der komorbiden depressiven Symptome führen, da die vorliegenden Ergebnisse die positive Rolle der sozialen Unterstützung ebenso bei der Reduktion der depressiven Symptome zeigen konnte.

Bei den Analysen zur stationären Psychotherapie von depressiven Störungen zeigten sich verschiedene Faktoren veränderungsresistent im kurzfristigen Verlauf, wie z. B.: Suizidalität, agoraphobische Ängste, Lebensunzufriedenheit, physische Beeinträchtigung und Schmerzen. Unsere Ergebnisse weisen daraufhin, dass insbesondere das Ausmaß der Suizidalität im Verlauf der Behandlung ziemlich stabil zu bleiben scheint und nicht durch die Psychotherapie signifikant reduziert wurde. Diese Ergebnisse unterstreichen die Notwendigkeit, suizidales Verhalten bei Menschen mit Depressionen (und anderen psychischen Problemen) angemessen zu berücksichtigen, wie auch in Teil 9 des Leitfadens der *Lancet Psychiatry Commission* (Holmes et al., 2018) vorgeschlagen wird. Die Rolle von Suizidalität sollte daher unbedingt in der Therapie adressiert werden, z. B. – wie in der DBT – durch einen Non-Suizid-Entschluss (NSE) zu Behandlungsbeginn und durch weitere Strategien zur Behandlung suizidalen Erlebens und Verhaltens (Teismann, Koban, Illes, & Oermann, 2016). Ein NSE kann ein nützliche Alternative zum Non-Suizid-Vertrag (NSV) sein (Hohagen, 2011), deren Nutzen als klinisch wirksame Instrumente aufgrund des Mangels an quantitativen Beweisen hinterfragt wird (McMyler & Pryjmachuk, 2008). Die gegenseitigen starken Wechselwirkungen aus Interessenverlust, Rumination, Konzentrationsschwächen und suizidalen Tendenzen könnten bspw. durch ergänzende Interventionen adressiert werden wie beispielsweise durch die *Attention Training Technique* (ATT) der MKT (Wells & Fisher, 2016). Agoraphobische Ängste bei Menschen mit Depressionen könnten z. B. durch erwartungsfokussierte Verhaltensexperimente verändert werden (Kube et al., 2018). In der Tat ist die sog. „ängstliche Depression“ ein Konstrukt, das in

der Praxis als eine schwieriger zu behandelnde Untergruppe der Depressionspopulation bekannt ist (Ionescu, Niciu, Henter, & Zarate, 2013). Im Allgemeinen zeigten auch achtsamkeitsbasierte Interventionen positive Effekte auf Angst und Depression (Hofmann, Sawyer, Witt, & Oh, 2010), welche bei diesen Patient\*innen nützlich sein könnten. Das achtsamkeitsbasierte Behandlungsprotokoll mit der breitesten Evidenzbasis ist die sog. „Acceptance and Commitment Therapy (ACT)“, welche sich ebenso als wirksam erwies, um die Lebenszufriedenheit zu erhöhen (Forman, Herbert, Moitra, Yeomans, & Geller, 2007). Kürzlich hat sich eine achtsamkeitsbasierte KVT als wirksam erwiesen, um die Lebenszufriedenheit von Jugendlichen mit depressiven Symptomen und suizidalen Ideen zu erhöhen (Raj et al., 2019), welche aber auch in Form eines kurzen gruppenbasierten Formats in einer Gemeinschaftsstichprobe nützlich scheint (Harnett et al., 2010). Insbesondere könnte die Verstärkung von werteorientiertem Verhalten zur Steigerung der Lebenszufriedenheit führen. Die Ergebnisse dieser Studie 4 zeigen zudem, dass es ratsam sein kann, die Vorbehandlungsvariablen zwar zu berücksichtigen bei Aufnahme, ihnen aber im Verlauf der Psychotherapie nicht so viel Raum zu geben und zu gewichten, da ihr Einfluss nachlässt.

Auch wenn die Ergebnisse des systematischen Reviews in Studie 5 kein bereits bestehendes Instrument identifizieren konnte, welches zur einheitlichen Verwendung empfohlen werden, so geben sie dennoch Hinweise, welches Instrument zu welcher Fragestellung passt könnte (vgl. Guidi et al., 2018). Die Ergebnisse des Reviews können dazu dienen, mehr Klarheit in diesem unübersichtlichen Forschungsfeld zu gewinnen sowie Empfehlungen zur Verbesserung der Erfassung von negativen Effekten in der Psychotherapie zu geben. Dazu zählen insbesondere die Erhebung von verschiedenen Bereichen, in denen Nebenwirkungen auftreten können, die Unterscheidung zwischen Nebenwirkung und therapeutischem Fehlverhalten, die Messung des Ausmaßes an Belastung sowie der Relevanz zum Therapieoutcome, die Herstellung der Attribution zur Psychotherapie, die Unterscheidung verschiedener Settings, Perspektiven und therapeutischer Orientierung sowie die Kontrolle von verzerrenden Antwortmustern durch Erfassung von positiven Effekten. Außerdem können die vorgeschlagene Konzeptualisierung und Definition von negativen Effekten in der Psychotherapie dazu beitragen, Praktiker\*innen über dieses wichtige Thema aufzuklären und zu sensibilisieren.

In Studie 6 scheint die stationäre CBASP-Behandlung negative Effekte auszulösen, die sich durch die spezifischen Merkmale von Patient\*innen mit behandlungsresistenten chronischen Depressionen und die Fokussierung der CBASP-Techniken auf die Patient\*in-Therapeut\*in-Beziehung erklären lassen. Nur die selbstberichtete wahrgenommene Abhängigkeit von der Therapeutin als häufigste Dimension negativer Effekte hängt dabei negativ mit dem Therapieoutcome zusammen. Wenn große randomisiert-kontrollierte Psychotherapiestudien ergeben, dass das stationäre CBASP-Programm eher eine Abhängigkeit von der Therapeutin auslöst als andere Psychotherapiekonzepte und wenn diese wahrgenommene Abhängigkeit sich tatsächlich immer noch negativ auf das Therapieoutcome auswirkt, dann sollten klinische Implikationen wie die stärkere Konzentration auf Selbsthilfe und Autonomie der Patientin sowie die Verlängerung der Behandlung in Betracht gezogen werden. Darüber hinaus scheint es wichtig, die Angehörigen und weitere Bezugspersonen der Patientin in die Therapie einzubeziehen, sofern sie verfügbar bzw. klinisch nah und motiviert sind. Es gibt in der Tat erste Hinweise darauf, dass die familiäre Psychoedukation die Rückfallraten senken könnte (Frank et al., 2015; Shimazu et al., 2011). Bei sozial isolierten Patient\*innen sollte der therapeutische Schwerpunkt auf dem Aufbau verschiedener positiver Beziehungen im Alltag der Patientin liegen, um potentielle Abhängigkeiten zu verhindern.

### 5.4 Perspektiven für zukünftige Forschung

Aus den Ergebnissen der vorliegenden Dissertation lassen sich vielfältige Perspektiven für zukünftige Forschungsarbeiten ableiten, welche das Ziel einer Personalisierung von psychotherapeutischen Angeboten im Rahmen einer praxisorientierten Forschung verfolgen. Aufgrund der eingangs beschriebenen Problematiken (vgl. Abschnitt 2.3) setzen sich führende Forscher\*innen dafür ein, dass Neurowissenschaftler\*innen und klinische Psycholog\*innen wieder enger zusammenarbeiten, um die Grabenkämpfe zu überwinden – im Englischen rufen sie zu einer "Mental health science" auf (Holmes, Craske, & Graybiel, 2014). Zur Professionalisierung der „Mental health“-Wissenschaften benötigt es laut der Autor\*innen drei Schritte: 1) Erforschung von Wirkmechanismen von vorhandenen psychologischen Therapien, 2) Optimierung von bestehenden psychologischen Therapien einschließlich der Entwicklung von neuen Ansätzen und 3) die enge Verbindung zwischen klinischen Forscher\*innen und Laborforscher\*innen (Holmes et al., 2014). Auf Möglichkeiten zur Adressierung dieser drei Schritte wird im Folgenden eingegangen.

Neben Wirksamkeitsanalysen (im Sinne der „effectiveness“) wie in der vorliegenden Arbeit werden zukünftig auch Analysen zur Kosteneffektivität (im Sinne der „cost-effectiveness“) im naturalistischen Routinesetting zum Zwecke des Benchmarkings wichtig sein, da verschiedene Interessenvertreter\*innen mit einbezogen werden müssen (Barkham et al., 2010). Ein wichtiger Punkt bei der Verwendung von Benchmarking-Daten ist, dass eine Therapie im Laufe der Zeit und nicht zu einem einzigen Zeitpunkt untersucht werden sollte. Im Rahmen der vorliegenden Analysen wird es zukünftig interessant sein, zu erforschen, ob der Prozentsatz der Patient\*innen, welche ähnliche Outcomes erzielen, aufrechterhalten werden kann von Jahr zu Jahr oder ggf. noch gesteigert werden kann mit den vorgeschlagenen Implikationen. Es könnte auch interessant sein, ob die klinisch signifikanten Outcomes durch die Menge (d.h. die Dosis) der erhaltenen Therapie beeinflusst wird (Barkham et al., 2010). Um in naturalistischen Psychotherapiestudien besser nachvollziehen zu können, welches „Behandlungspaket“ erhalten wurde, könnte zur Bestimmung der Therapiedosis und Check der Leitlinien-treue – ähnlich wie es in den Rehabilitationswissenschaften üblich ist (z. B. Bethge, Markus, Streibelt, Gerlich, & Schuler, 2019) – neben der überbrachten Dosis, die durch Leitlinien vorgeschrieben ist, auch die erhaltene Dosis (z. B. im Patient\*innenbericht) strukturiert und standardisiert miterhoben und in die administrativen Aufnahmen überführt werden.

In klinischen Studien mit mindestens zwei Behandlungsarmen (z. B. RCTs) bieten zudem gemischte Modelle zur ES-Schätzung mehrere Vorteile gegenüber LOCF und in zukünftigen prospektiven Studien kann z. B. das „Mixed-Effect Model Repeated Measure (MMRM)“ verwendet werden (Siddiqui et al., 2009). Ebenso könnte neben Effekstärken zur Darstellung der klinischen Bedeutsamkeit eines Effekts auch noch NNT vermehrt eine Rolle spielen (Kraemer & Kupfer, 2006), welche bislang in der Psychotherapieforschung noch weniger Beachtung finden, aber in der medizinischen Forschung durchaus üblich sind.

Die Studien konnten zwar durch retrospektive Analysen wichtige prognostische Variablen im Sinne von allgemeinen Prädiktoren des Behandlungsergebnisses bei einer stationären Psychotherapie identifizieren. Die Befunde zu den prognostischen Variablen sollten allerdings in prospektiven Studien vor allem auch im ambulanten Setting und unter kontrollierten Bedingungen (z. B. im Rahmen von RCTs) und Meta-Analysen repliziert werden, um die Generalisierbarkeit dieser Befunde auf andere Kontexte und Länder zu ermöglichen (Buckman et al., 2019). Zudem könnte durch die Hinzunahme von zusätzlichen wichtigen Variablen als diejenigen, die in den vorliegenden Studien verwendet wurden, noch mehr Varianz in solchen heterogenen Stichproben aufgeklärt werden. Wie in der Einleitung erwähnt, ist die Durchführung der stationären Psychotherapie in deutschsprachigen Ländern relativ einzigartig. Neben der Verwendung von RCTs können sich auch neue methodische Designs in der Psychotherapieforschung zur Replikation der gefundenen Prädiktoren eignen. Zur schnelleren Entwicklung und Optimierung von Psychotherapie können in Zukunft insbesondere das sog. „Leapfrog Design“ verwendet werden, das einige Quellen der Ineffizienz in der Entwicklung von Behandlungen überwunden hat (Blackwell, Woud, Margraf, & Schönbrodt, 2019). Leapfrog-Designs basieren auf adaptiven Rollendesigns, welche ursprünglich aus der Krebsforschung heraus entwickelt wurde. In solchen Designs werden mehrere Behandlungsoptionen simultan gegeneinander getestet und diejenigen aus der fortlaufenden Studie entfernt, welche durch sequentielle Bayessche Analysen schlecht performen. Gleichzeitig können neue Behandlungsoptionen während der kontinuierlichen Durchführung aufgenommen werden. Die Darstellung eines prototypischen Leapfrog-Designs kann der Studie von Blackwell et al. (2019) entnommen werden. Es stellt ein integratives, kontinuierliches Forschungsdesign dar, welches gut in die Routineversorgung (z. B. Ambulanzen) implementiert werden kann und sich dadurch im Rahmen von praxisorientierter Forschung zur Generierung von praxisbasierter Evidenz sehr gut eignen würde.

Auf der Suche nach größeren Effekten von Psychotherapie könnten entweder im Rahmen von RCTs oder im Rahmen solcher Leapfrog-Designs komparative klinische Studien (z. B. TxA vs. TxB) durchgeführt werden, in denen sowohl das Therapieoutcome als auch Dosis-Wirkungszusammenhänge untersucht und ebenso Komponentendesigns (z. B. Dismantling<sup>17</sup> und Additive Designs) getestet werden können. So könnten die spezifischen Effekte einer psychologischen Therapie identifiziert werden. Eine aktuelle Studie fand heraus, dass distinkte Dosis-Response-Muster zwischen hoch- und niedrig-intensiver KVT vorhanden sind (Robinson, Kellett, et al., 2020). Eine ältere Komponentenanalyse zur KVT bei Depression ergab, dass einzelne Komponenten durchaus so effektiv wie die Gesamtbehandlung sein können (Jacobson et al., 1996), was einzelne Forscher\*innen zur Kritik an der Spezifität von Psychotherapie geführt hat (Ahn & Wampold, 2001). Eine neuere Meta-Analyse zu Komponentenstudien konnte zeigen, dass bisherige Dismantlingstudien keine spezifischen Effekte zeigen konnten, jedoch Studien mit additiven Designs einen kleinen signifikanten Effekt finden konnten (Bell, Marcus, & Goodlad, 2013). Immer mehr Studien deuten daraufhin, dass die Erhöhung der Therapieintensität auf 2 Sitzungen pro Woche einen positiven Effekt auf das Behandlungsergebnis bei Patient\*innen mit depressiven Störungen hat (Bruijniks et al., 2020). Beispielsweise könnte auch die Überprüfung von kreuzmodalen Behandlungen, bestehend aus psychologischen und pharmakologischen Therapien, zur Innovation in der Therapieforschung beitragen (Holmes et al., 2018). Die Implementierung in die Routineversorgung durch solche praxisbasierten Studien z. B. in Hochschulambulanzen oder Tageskliniken wäre eine Möglichkeit, um den ‚Research-Practice-Gap‘ durch Förderung von praxisbasierter Evidenz weiter zu verringern und praxisorientierte Forschung zu etablieren (Castonguay et al., 2013).

Diese in den Studien 1-4 identifizierten prognostischen Variablen sind als allgemeine Prädiktoren einer einzelnen Behandlung anzusehen, aber sagen noch nichts über differentielle Therapieoutcomes aus. Solche differentiellen Prädiktoren können sich als Moderatoren des Therapieoutcomes verstehen. Sie könnten mit Hilfe des PAI (DeRubeis et al., 2014) anhand einer RCT mit mind. zwei verschiedenen Behandlungsbedingungen identifiziert werden und Auskunft über den preskriptiven Wert einer Variable geben (Dwyer et al., 2018). Langfristig kann die Identifikation von solchen preskriptiven Variablen helfen, die Auswahl von

---

<sup>17</sup> Ein als kritisch definiert „Inhaltsstoff“ wird aus einer bestehenden Behandlung entfernt oder eine Komponente wird einer bestehenden Behandlung hinzugefügt, um die Auswirkungen der gelöschten oder hinzugefügten Komponente zu bestimmen.



Behandlungen zu leiten (Cohen & DeRubeis, 2018; Keefe et al., 2020). So zeigte eine aktuelle Studie zum Vergleich zwischen DBT und einer allgemeinen psychiatrischen Behandlung bei Borderline-Persönlichkeitsstörungen, dass die wichtigsten Moderatoren der Schweregrad der psychiatrischen Symptome, die borderline-spezifische Impulsivität (beide GPM > DBT), abhängige Persönlichkeitsmerkmale, emotionaler Missbrauch in der Kindheit und soziale Anpassung (alle DBT > GPM) sind (Keefe et al., 2020). Die Bestimmung von solchen Prädiktoren bereits zu Behandlungsbeginn ermöglicht langfristig eine Etablierung von Entscheidungsbäumen bzw. Klassifikationsbäume zur Entscheidungsfindung (Hannöver, Richard, Hansen, Martinovich, & Kordy, 2002), um die Entscheidung von Praktiker\*innen bei der Auswahl einer passenden und vielversprechenden Behandlung zu leiten. Sie kann somit als Entscheidungshilfe vor bzw. zu Beginn der Therapie dienen, woraus sich individuelle Behandlungsempfehlungen ableiten lassen.

Eine personalisierte Behandlungsempfehlung könnte aus zwei Vorgehensweisen resultieren: nomothetisch und idiographisch. Durch Cluster-Analysen im Rahmen von Netzwerkmodellen könnten nomothetisch Subtypen von sehr heterogenen Störungsbildern wie z. B. depressive Störungen, bei denen die ES für Psychotherapie auch deswegen deutlich unterhalb derer von anderen Störungsbildern liegt, identifiziert werden, die über klassische Diagnosen hinausgehen (Robinaugh, Hoekstra, Toner, & Borsboom, 2019). Das ist insbesondere in naturalistischen Studien mit TAU sinnvoll, da dort wenig Ein- und Ausschlusskriterien herrschen. Im Rahmen von symptom-spezifischen Wirksamkeitsanalysen (z. B. Netzwerkinterventionsanalysen) können dann Therapieoutcomes auf der Basis von den identifizierten Symptomprofilen vorhergesagt werden (z. B. Elliott, Jones, & Schmidt, 2019) und personalisierte Behandlungszuweisungen im Sinne der „Treatment selection“ stattfinden (z. B. „Leeds risk index“, Delgadillo et al., 2020). In zukünftigen Studien mit Patient\*innen mit PTBS könnte ebenso zwischen Patient\*innen mit PTBS und komplexer PTBS (kPTBS) mit Hilfe angemessener diagnostischer Hilfsmittel und datengetriebener Analysen unterschieden werden (Böttche et al., 2018; Karatzias et al., 2019; Maercker et al., 2013), um die Komplexität der spezifischen Symptomatik einer Person zu erfassen und das Behandlungsergebnis genauer zu bestimmen. Idiographisch könnten Einzelfallstudien durch die Verwendung von elektronischen Tagebüchern im Rahmen des „Ecological Momentary Assessment (EMA)“<sup>18</sup> mit

---

<sup>18</sup> Diese Methode wird auch als „Experience Sampling Method“ oder „Ambulantes Assessment“ bezeichnet.

Hilfe von Smartphones oder Tablets z. B. während der Wartezeit (4 Wochen, 4 Mal am Tag) durchgeführt werden, um eine spätere Behandlungsempfehlung abzuleiten<sup>19</sup>. Bei der Bildung von idiographischen Symptomzusammenhängen müssten bei einer individuellen Patientin Symptome über einen längeren Zeitraum wiederholt gemessen werden, wobei diese intensiven Längsschnittdaten zur Vorhersage der individuellen Patientin herangezogen werden (Fisher & Boswell, 2016; Fisher, Reeves, Lawyer, Medaglia, & Rubel, 2017; Husen, Rafaeli, Rubel, Bar-Kalifa, & Lutz, 2016; Lutz et al., 2018; Rubel, Fisher, Husen, & Lutz, 2018). Eine offene Studie zu einer personalisierten modularen Therapie bei Depression und Angststörungen spricht für die Wirksamkeit eines idiographischen Vorgehens (Fisher et al., 2019). Da die Forschung bisher nur wenig replizierbare Prädiktoren finden konnte, scheint es wichtig, kritische Prozesse in Patient\*innen mit einem erwünschten Outcome im Rahmen von idiographischen Analysen zu identifizieren (Castonguay & Beutler, 2006). Im Rahmen von der Weiterentwicklung von KVT wird zunehmend eine Prozessorientierung vorgenommen (Hofmann & Hayes, 2019), d.h. dass psychologische Therapie gezielt Prozesse des Lernens, der Emotionsregulation und der Bildung von Gewohnheiten abzielt.

Im Bereich der Psychotherapieverlaufsforschung<sup>20</sup> im Spezifischen und Prozessforschung im Generellen eignet sich zunehmend die Verwendung von EMA. Diese Methode eignet sich neben der Generierung von idiographischen Daten zur personalisierten Behandlungsempfehlung auch im Rahmen der patient\*innenfokussierten Forschung zur Analyse von Mediatoren. Gleichzeitig birgt die große Menge an Daten und wiederholte Messung innerhalb von Personen Chancen aber auch neue Herausforderungen, wodurch die Psychotherapieforschung ausgereifere statistische Analysemethoden benötigt, um diese Komplexitäten anzugehen, wie z. B. Analysen, die auf Bayesschen Statistiken für kausale Inferenzen basieren (Pearl, 2000; Pearl & Mackenzie, 2018). Mit Hilfe von solchen Prozessdaten können durch Mediatoranalysen besser Veränderungsmechanismen untersucht werden, wie z. B. die Erwartungsverletzung im Rahmen von Expositionstherapie bei Angststörungen (Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014). Um der Komplexität des therapeutischen Prozesses und Veränderung Rechnung zu tragen und Antworten auf die Frage „What works for

---

<sup>19</sup> Eine technische Lösung zur Durchführung von EMA könnte z. B. mit MovisensXS realisiert werden (zu Publikationen mit MovisensXS, siehe folgende Internetseite: <https://www.movisens.com/de/produkte/movisensxs/>).

<sup>20</sup> Das Ziel hierbei ist die Identifikation von Wirkmechanismen, die dafür verantwortlich sind, dass eine therapeutische Intervention zu bestimmten Effekten führt.

whom, and why?“ zu finden, müssen sowohl Moderator- als auch Mediatoranalysen kombiniert durchgeführt werden (Brakemeier & Herpertz, 2019; Zilcha-Mano, 2019). Dazu können neue statistische Verfahren (Craske, 2019), u.a. die Verwendung von Netzwerkanalysen, hilfreich sein (Lorimer, Delgado, Kellett, & Brown, 2019; Lutz et al., 2018).

Durch die Nutzung von Feedback- und Problemlösetools wie z. B. das computerunterstützte Stuttgart-Heidelberg-Model (Kordy, Hannover, & Richard, 2001), oder der technologisch-versiertere TTN (Lutz, Rubel, et al., 2019) im Rahmen eines generellen Routine-Outcome-Managements (ROM) (Lambert & Harmon, 2018; Lambert et al., 2018) in z. B. Hochschulambulanzen und Tageskliniken (Hiller, Schindler, Andor, & Rist, 2011) kann patient\*innenfokussierte Forschung (Castonguay & Muran, 2016) ermöglicht werden. Mit Hilfe solcher Lösungen können Entscheidungshilfen auch im Therapieverlauf ermöglicht werden, welche zur personalisierten Behandlungsanpassung führen, z. B. Generierung von Modellen zum erwarteten Therapieerfolg (Lutz et al., 2015; Rubel, Ehrlich, & Lutz, 2015). Das Feedback zum Outcomefortschritt hat sich als wirksames Werkzeug zur Unterstützung der Therapie erwiesen, vor allem für Patient\*innen mit einem möglichen Therapiemisserfolg aber auch im Supervisionsprozess im Ausbildungskontext (Rubel, Zimmermann, Deisenhofer, Müller, & Lutz, 2017). Dabei können zur Personalisierung z. B. auch Fallkonzeptualisierungen herangezogen werden (Mumma & Fluck, 2016).

Im Rahmen eines ROM könnten auch Verschlechterungsraten im Sinne vom RCI als Outcome (Scott & Young, 2016) berechnet und zusätzlich auch Nebenwirkungen miterhoben werden, welche vorallem auch als Prozessvariable mit in die Analysen einbezogen werden könnten. Prozess-Outcome-Analysen sind dabei wichtig zur Herstellung des Einflusses der Behandlung sowie der Relevanz von Nebenwirkungen auf die individuelle Patientin. Eine erste Pilot-Studie hat, analog zur Studie 6, die Relevanz von negativen Effekten auf eine stationäre CBASP-Behandlung untersucht (Brakemeier et al., 2018). Solche Analysen könnten zur Erhöhung der klinimetrischen Eigenschaften führen (Bech, 2016; Fava et al., 2012; Feinstein, 1987; Guidi et al., 2018). Die Einführung von ROM kann es ermöglichen, auch die Skalen zu negativen Effekten besser zu validieren, z. B. durch längsschnittliche Erhebungen, die bisher retrospektiv erhoben wurden. Dazu wäre eine Prozessskala zur Erfassung von negativen Effekten von Psychotherapie wünschenswert. Eine zukünftige Forschungsfrage bleibt in diesem Forschungsfeld außerdem die Untersuchung von negativen Effekten in spezifischen

Patient\*innenpopulationen (z. B. Kinder und Jugendliche), in verschiedenen Settings (z. B. stationäres Setting, Gruppensetting) als auch die Validierung von Therapeut\*innenratings (z. B. UEA-ATR). Durch die Einführung von standardisierten Follow-up-Messzeitpunkten im Rahmen des ROM und das Begleiten von Patient\*innen über das Ende der Therapie hinaus, könnten zukünftig Analysen ermöglicht werden, die z. B. Prädiktoren für einen Rückfall vorhersagen und so Patient\*innen mit erhöhtem Risiko für einen Rückfall identifizieren.

In der vorliegenden Studie wurde auf einen routinemäßig erhobenen Datensatz zurückgegriffen, der insbesondere aus Patient\*innenvariablen besteht. Neben der Spannweite an unterschiedlichen soziodemographischen, klinischen und psychometrischen Variablen können dennoch wichtige Prädiktoren in der routinemäßigen Erhebung gefehlt haben, insbesondere solche die in der Literatur als „allgemeine Faktoren“ diskutiert werden. Diesen könnte in zukünftigen Untersuchungen noch vermehrt Aufmerksamkeit geschenkt werden. Den Einfluss von solchen Faktoren auf den Erfolg von Psychotherapie haben die Psychotherapieforscher Wampold & Imel (2015) in ihrem kontextuellen Meta-Modell der Psychotherapie („Contextual Model“, Wampold & Imel, 2015) genauer beschrieben. Demnach gebe es drei zentrale Pfade, über die Psychotherapie eine positive Wirkung entfalten könne: die „echte“ therapeutische Beziehung, Erwartungen und spezifische Elemente. Die Grundvoraussetzung für die Wirkung dieser drei Pfade sei laut Wampold & Imel (2015) eine positive initiale therapeutische Beziehung. Erwartungen der Patient\*innen können laut Wampold & Imel (2015) einen positiven Wirkmechanismus darstellen, wenn es gelingt, den Patient\*innen ein gutes Verständnis davon zu vermitteln, wie die Psychotherapie ihnen bei ihren Problemen helfen kann und wie sie selbst zur Überwindung ihrer Probleme beitragen können. Da Erwartungen einen wesentlichen Prädiktor für den Erfolg von medizinischen Behandlungen darstellen, ist es naheliegend, dass Erwartungen an eine psychotherapeutische Behandlung ähnlich bedeutsam für deren Erfolg sind (Brooks, Bratley, Jones, King, & Lucock, 2020). Tatsächlich konnten mehrere Arbeiten nachweisen, dass die Erwartungen von Patient\*innen an eine Psychotherapie einen starken Einfluss auf den Behandlungserfolg haben (Constantino et al., 2011; Constantino, Visla, et al., 2018; Glass et al., 2001; Greenberg et al., 2006; Price et al., 2008; Price & Anderson, 2012), ebenso wie die Akzeptanz und Wahrnehmung der Glaubwürdigkeit einer Behandlung durch die Patientin (Constantino, Coyne, Boswell, Iles, & Visla, 2018). Demnach stellen Erwartungen der Patient\*innen an eine Psychotherapie einen wichtigen Prädiktor für deren Erfolg dar, sodass angestrebt werden sollte, günstige

Erwartungen an die Psychotherapie zu fördern und Erwartungen, die den Therapieerfolg gefährden könnten, abzubauen (Constantino, Visla, et al., 2018). Patient\*innen kommen nicht mit einer Tabula rasa in Therapie, daher können solche Erwartungen optimiert und im Sinne eines Placebo-Effekts in der Psychotherapie gezielt genutzt werden (Gaab, Kossowsky, Ehlert, & Locher, 2019; Wampold, Minami, Tierney, Baskin, & Bhati, 2005).

Beispielsweise konnte gezeigt werden, dass es sinnvoll ist, Patient\*innenattributionen hinsichtlich der Verbesserungen, die mit einer kombinierten expositionsbasierten und pharmakologischen Behandlungen von Angststörungen erreicht werden, zu erheben (Powers, Smits, Whitley, Bystritsky, & Telch, 2008). Im Bereich des Einflusses von Patient\*innenerwartungen bzgl. der Psychotherapie bedarf es insbesondere noch experimenteller Studien (z. B. Manipulation von Wirkung, bei gleicher Erklärung der psychischen Störung und Präsentation des Behandlungsrationalis). Dazu bedarf es des routinemäßigen Assessments zu Behandlungsbeginn von sowohl Outcome- als auch Behandlungserwartungen. Die Bedeutung von Erwartungen ist im Einklang mit aktuellen interdisziplinären Forschungsbemühungen, u.a. durch neurowissenschaftliche und philosophische Strömungen hervorgerufen. Das Gehirn wird als hypothesen-testender komplexer Mechanismus – im Englischen als „Predictive Mind“ oder „Prediction Machine“ – verstanden, welcher ständig daran beteiligt ist, den Fehler seiner Vorhersagen über den sensorischen Input, den er von der Welt erhält, zu minimieren – ein Prozess, der „prediction error minimization“ genannt wird (Hohwy, 2014). Der Fokus wurde bereits in den letzten Jahren häufig auf die Überprüfung und Modifikation von störungsspezifischen Erwartungen gelegt, z. B. in der Therapie von Angststörungen (Craske et al., 2014). Ein aktueller Ansatz, der wieder zur Annäherung zwischen Neurowissenschaftler\*innen, klinischen Psycholog\*innen und Psychiatrer\*innen geführt hat, ist die Entwicklung von „Predictive Processing“-Perspektiven zur Erklärung der Symptome für verschiedene psychische Störungen (Kube, Berg, Kleim, & Herzog, 2020; Kube, Schwarting, Rozenkrantz, Glombiewski, & Rief, 2019; Sterzer et al., 2018; Van den Bergh, Witthöft, Petersen, & Brown, 2017).

Darüber hinaus könnte die Messung biologischer Marker, die z. B. aus genetischen und Neuroimaging-Daten gewonnen werden, in zukünftigen Studien entlang mehrerer klinischer Marker nützlich sein und wahrscheinlich noch erfolgreicher sein, um eine "Biosignatur" depressiver Störungen und deren Veränderung durch psychologische Behandlungen zu

entwickeln (Trivedi, 2013). Es zeigte sich beispielsweise, dass die genetische Variation das Ansprechen auf eine KVT bei pädiatrischer Ängstlichkeit vorhersagen kann (Beevers & McGeary, 2012). In Übereinstimmung mit der methodischen und statistischen Perspektive, die davon ausgeht, dass Modelle mit mehreren Daten besser passen als Modelle mit einem einzigen niedrigdimensionalen Datentyp, ist es sicherlich sinnvoll, bei der Planung neuer informativer Datenerhebungen und prospektiver Studien die Einbeziehung mehrerer Datenquellen, wie z. B. Neuroimaging und genetische Daten, zusätzlich zu den phänomenologischen Daten in Betracht zu ziehen (Lee et al., 2018).

Einen weiteren wichtigen Punkt im Rahmen der therapeutischen Faktoren, die zum Therapieoutcome beigetragen können, betrifft die therapeutische Beziehung (Stamoulos et al., 2016). Die Gestaltung der therapeutischen Beziehung scheint ein wichtiger Faktor zu sein und einen Einfluss auf den Behandlungserfolg zu haben (Norcross & Lambert, 2018; Norcross & Wampold, 2011). Insbesondere die Therapeut\*innenvariabilität scheint im Rahmen der therapeutischen Beziehung das Therapieoutcome vorherzusagen (Baldwin, Wampold, & Imel, 2007). Im Besonderen wurde der Einfluss der therapeutischen Beziehung auf das Therapieoutcome in unterschiedlichen Kontexten meta-analytisch nachgewiesen (Flückiger et al., 2018; Friedlander et al., 2018; Horvath et al., 2011; Karver et al., 2018) und der Zusammenhang zwischen der „realen“ Beziehung und dem Therapieoutcome gezeigt (Gelso, Kivlighan, & Markin, 2018). Zusätzlich wurde auch der Einfluss eines Bruches der therapeutischen Beziehung auf den Therapieerfolg nachgewiesen (Eubanks, Muran, & Safran, 2018).

Weitere Faktoren scheinen Empathie (Elliott et al., 2011, 2018), Zielkonsens/Zusammenarbeit bzw. Kollaboration (Shick Tryon et al., 2018; Tryon & Winograd, 2011), Akzeptanz/Bestätigung/positive Wertschätzung (Farber & Doolin, 2011; Farber, Suzuki, & Lynch, 2018) und Kongruenz/Echtheit (Kolden, Klein, Wang, & Austin, 2011; Kolden, Wang, Austin, Chang, & Klein, 2018) zu sein; ebenso wie die kulturelle Anpassung der evidenzbasierten Behandlung (Benish et al., 2011). Diese Einflussvariablen müssten im Verlauf der Therapie als Prozessvariable erhoben werden. Bei Gruppentherapien spielt die Kohäsion eine große Rolle für den Therapieerfolg (Burlingame, 2018). Das Sammeln und Bereitstellen von Fortschrittsfeedback im Rahmen von ROM (Boswell, Constantino, Kraus, Bugatti, & Oswald, 2016), wie oben erwähnt, scheint auch einen Einfluss auf den Therapieerfolg zu haben

(Lambert et al., 2018), ebenso wie Gegenübertragungsmanagement (Hayes, Gelso, Goldberg, & Kivlighan, 2018) und der emotionale Ausdruck von Therapeut\*in und Patient\*in (Peluso & Freund, 2018). Eine qualitative Meta-Analyse deutet zudem auf die Rolle der therapeutischen Selbstöffnung sowie Unmittelbarkeit im therapeutischen Prozess hin (Hill, Knox, & Pinto-Coelho, 2018).

Darüber hinaus wurden keine Variablen bzgl. der Therapeut\*innen erhoben: Therapeut\*inneneffekte scheinen jedoch eine wichtige Rolle zu spielen (Berglar et al., 2016; Goldberg et al., 2018; Imel et al., 2015) und erklären ca. 5 % in der Variabilität des Therapieoutcomes (Johns et al., 2019), wobei die Effekte meist im naturalistischen Setting größer als in klinischen Studien waren (Laska, Smith, Wislocki, Minami, & Wampold, 2013). Daher ist es in naturalistischen Settings umso wichtiger dafür zu kontrollieren, insbesondere in der stationären Psychotherapie, da Therapeut\*inneneffekte mit schwerer erkrankten Patient\*innen größer zu sein scheinen (Saxon & Barkham, 2012). Dies scheint an der Fähigkeit zu liegen, wie interpersonell herausfordernde Begegnungen mit Patient\*innen gelöst werden (Anderson, Ogles, Patterson, Lambert, & Vermeersch, 2009). In einem aktuellen Review konnte gezeigt werden, dass effektivere Therapeut\*innen durch professionell kultivierte interpersonelle Fertigkeiten gekennzeichnet sind, die wahrscheinlich in ihrem persönlichen Leben und ihrer Bindungsgeschichte verwurzelt sind (Heinonen & Nissen-Lie, 2019).

Außerdem verdeutlicht die Studie 6 die Rolle der Suizidalität im Rahmen der stationären Psychotherapie bei Depression. Eine international renommierte Forscher\*innengruppe weist in einem Statement zu wichtigen zukünftigen Forschungsbereichen in der Psychotherapieforschung daraufhin, dass unser Wissen, trotz Kenntnissen über das Verständnis von Risikofaktoren zur Vorhersage eines wahrscheinlichen Suizidversuchs, über suizidales Verhalten und der Rettung von Leben begrenzt ist, Fortschritte aufgrund vieler offener Fragen dringend notwendig sind und daher zukünftig wieder mehr in den Fokus der Forschung gerückt werden sollte (Holmes et al., 2018).

Alle oben genannten offenen Forschungsfelder beinhalten Strategien, um die Effekte von Psychotherapien in verschiedenen Settings noch weiter zu optimieren und langfristig zu einer Personalisierung von Psychotherapie auf die individuellen Bedürfnisse der Patientin beizutragen. Wenn sie als Teil eines größeren Qualitätsverbesserungssystems im Rahmen praxisorientierter Forschung integriert werden, könnten Prozess-Outcome-Daten

Informationen über den Fortschritt der Behandlungserfolgs in Echtzeit liefern und praktikable alternative Behandlungspläne vorschlagen (Wampold & Imel, 2015). Um die vorangegangenen Fragestellungen jedoch bearbeiten zu können, ist es wichtig, dass auch die notwendige Finanzierung zur Verfügung gestellt wird. Dabei müssen Ausgaben für die Erforschung psychologischer Behandlungen in einem angemessenen Verhältnis zu ihren Auswirkungen stehen. Forschung zu psychologischen Therapien ist aktuell nicht ausreichend unterstützt: nur ca. 2 % aller Forschungsgelder der EU unterstützen Forschungsprojekte, die psychologische Therapien untersuchen, obwohl psychische Störungen ca. 15% der Krankheitsbelastung in entwickelten Ländern und damit mehr als alle Krebsarten zusammen ausmachen (Holmes et al., 2014).



### 5.5 Fazit

Insgesamt unterstreicht die vorliegende Dissertation die Bedeutung von Phase IV-Studien auf der Basis von sehr großen Datensätzen von Patient\*innen mit verschiedenen psychischen Erkrankungen aus der stationären Routineversorgung in Deutschland und zeigt Ansatzpunkte zur Optimierung dieser Effekte auf.

Es kann anhand verschiedener relevanter Outcomemaße gezeigt werden, dass die Implementierung von evidenzbasierten Psychotherapien in der stationären psychosomatischen Routineversorgung insbesondere von der Posttraumatischen Belastungsstörung, Zwangsstörung und depressiven Störungen gelungen ist. Patient\*innen mit Borderline-Persönlichkeitsstörung scheinen weiterhin trotz eines intensiven stationären Psychotherapiekonzepts basierend auf der DBT insgesamt davon nur in geringerem Ausmaß zu profitieren. Alles in allem sprechen die Ergebnisse hingegen für die absolute Wirksamkeit von Psychotherapie. Die berechneten Effektstärken können dabei als interne Benchmark verstanden werden, gegen die zukünftig in vergleichbaren Settings getestet werden kann. Mit Hilfe verschiedener statistischer Verfahren wurden zudem klinische multivariate Prädiktionsmodelle entwickelt, welche Praktiker\*innen helfen sollen, Patient\*innen mit dem Risiko eines geringeren Psychotherapieerfolgs schon während der Behandlungsplanung zu identifizieren und mit Hilfe geeigneter Interventionen zu adressieren, um dieses Risiko zu minimieren und den Behandlungserfolg zu maximieren. Im Sinne des Qualitätsmanagements können die Ergebnisse dieser Studien daher helfen, zukünftig Rückmeldeschleifen an die Praktiker\*innen zu entwickeln, um die Behandlung zu optimieren.

Neben positiven und ausbleibenden Effekten wurde sich in dieser Dissertation außerdem dem Thema der negativen Effekte gewidmet. Dazu wurde im Rahmen eines systematischen Reviews bisher entwickelte und teils validierte Messinstrumente zusammengefasst, um eine Standortbestimmung dieses Themas in der aktuellen Forschung vorzunehmen. Es wurde für eine einheitliche Begriffsverwendung im Rahmen einer konsensualen Definition und Konzeptualisierung von negativen Effekten plädiert sowie Empfehlungen für die Erfassung von negativen Effekten im Rahmen der Psychotherapie abgeleitet, wodurch weitere wissenschaftlichen Bemühungen in diesem noch „jungen“ Forschungsfeld stimuliert werden können. Auf der Basis der Ergebnisse wurde geschlussfolgert, dass das Auftreten von negativen Effekten stets mit dem Psychotherapieerfolg in Zusammenhang zu bringen sei, um die Relevanz

von Nebenwirkungen abzuschätzen. Dies wurde in einer Studie exemplarisch am Beispiel eines stationären CBASP-Programms umgesetzt, in der gezeigt werden konnte, dass negative Effekte zwar sehr oft auftreten, allerdings nur wenige dieser Effekte im Zusammenhang zum Behandlungserfolg stehen.

Die Ergebnisse der sechs Studien im Rahmen der praxisorientierten Forschung liefern praxisbasierte Evidenz und tragen zu einer Förderung des Praxis-Forschungsnetzwerk in Deutschland bei.

## Literaturverzeichnis

- Adams, R. A., Huys, Q. J. M., & Roiser, J. P. (2016). Computational Psychiatry: towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery, and Psychiatry*, *87*(1), 53–63. <https://doi.org/10.1136/jnnp-2015-310737>
- Addis, M. E., & Jacobson, N. S. (1996). Reasons for depression and the process and outcome of cognitive-behavioral psychotherapies. *Journal of Consulting and Clinical Psychology*, *64*(6), 1417–1424. <https://doi.org/10.1037//0022-006x.64.6.1417>
- Ahn, H. N., & Wampold, B. E. (2001). Where oh where are the specific ingredients? A meta-analysis of component studies in counseling and psychotherapy. *Journal of Counseling Psychology*, *48*(3), 251–257. <https://doi.org/10.1037/0022-0167.48.3.251>
- Anderson, T., Ogles, B. M., Patterson, C. L., Lambert, M. J., & Vermeersch, D. A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology*, *65*(7), 755–768. <https://doi.org/10.1002/jclp.20583>
- Andrade, C. (2015). The numbers needed to treat and harm (NNT, NNH) statistics: what they tell us and what they do not. *The Journal of Clinical Psychiatry*, *76*(3), e330-3. <https://doi.org/10.4088/JCP.15f09870>
- APA, P. T. F. on E.-B. P. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*, 271–285.
- Arean, P. A., & Kraemer, H. C. (2015). *High Quality Psychotherapy Research*. New York: Oxford University Press.
- Association American Psychiatric. (2015). Diagnostisches und Statistisches Manual Psychischer Störungen DSM-5®. In P. Falkai & H.-U. Wittchen (Eds.), *Diagnostisches und Statistisches Manual Psychischer Störungen DSM-5®*. <https://doi.org/10.1026/02803-000>
- Baardseth, T. P., Goldberg, S. B., Pace, B. T., Wislocki, A. P., Frost, N. D., Siddiqui, J. R., ... Wampold, B. E. (2013). Cognitive-behavioral therapy versus other therapies: Redux. *Clinical Psychology Review*, Vol. 33, pp. 395–405. <https://doi.org/10.1016/j.cpr.2013.01.004>
- Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the Alliance-Outcome

Correlation: Exploring the Relative Importance of Therapist and Patient Variability in the Alliance. *Journal of Consulting and Clinical Psychology*, 75(6), 842–852. <https://doi.org/10.1037/0022-006X.75.6.842>

Bandelow, B., Reitt, M., Röver, C., Michaelis, S., Görlich, Y., & Wedekind, D. (2015). Efficacy of treatments for anxiety disorders: A meta-analysis. *International Clinical Psychopharmacology*, 30(4), 183–192. <https://doi.org/10.1097/YIC.0000000000000078>

Bar-Kalifa, E., Atzil-Slonim, D., Rafaeli, E., Peri, T., Rubel, J., & Lutz, W. (2016). Therapist-Client Agreement in Assessments of Clients' Functioning. *Journal of Consulting and Clinical Psychology*, 84(12), 1127–1134. <https://doi.org/10.1037/ccp0000157>

Barber, J. P., & Solomonov, N. (2019). Toward a personalized approach to psychotherapy outcome and the study of therapeutic change. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 18(3), 291–292. <https://doi.org/10.1002/wps.20666>

Barkham, M., Hardy, G. E., & Mellor-Clark, J. (2010). *Developing and delivering practice-based evidence: a guide for the psychological therapies*. Chichester, UK: Wiley-Blackwell.

Barkham, M., & Margison, F. (2008). Practice-based Evidence as a Complement to Evidence-based Practice: From Dichotomy to Chiasmus. In *Handbook of Evidence-based Psychotherapies: A Guide for Research and Practice* (pp. 443–476). <https://doi.org/10.1002/9780470713242.ch25>

Barkham, M., & Mellor-Clark, J. (2003). Bridging evidence-based practice and practice-based evidence: developing a rigorous and relevant knowledge for the psychological therapies. *Clinical Psychology & Psychotherapy*, 10(6), 319–327. <https://doi.org/10.1002/cpp.379>

Barkham, M., Stiles, W. B., Connell, J., Twigg, E., Leach, C., Lucock, M., ... Angus, L. (2008). Effects of psychological therapies in randomized trials and practice-based studies. *British Journal of Clinical Psychology*, 47(4), 397–415. <https://doi.org/10.1348/014466508X311713>

Barlow, D. H. (2004). Psychological treatments. *American Psychologist*, 59(9), 869–878. <https://doi.org/10.1037/0003-066X.59.9.869>

Barlow, D. H. (2010). Negative effects from psychological treatments: a perspective. *The American Psychologist*, 65(1), 13–20. <https://doi.org/10.1037/a0015643>

- Barlow, D. H., Ellard, K. K., Fairholme, C. P., Farchione, T. J., Boisseau, C. L., Allen, L. B., & Ehrenreich-May, J. T. (2015). *Unified Protocol for Transdiagnostic Treatment of Emotional Disorders*. Oxford: Oxford University Press. <https://doi.org/10.1093/med:psych/9780199772674.001.0001>
- Barnicot, K., Katsakou, C., Marougka, S., & Priebe, S. (2011). Treatment completion in psychotherapy for borderline personality disorder - a systematic review and meta-analysis. *Acta Psychiatrica Scandinavica*, *123*(5), 327–338. <https://doi.org/10.1111/j.1600-0447.2010.01652.x>
- Barnicot, K., Katsakou, C., Bhatti, N., Savill, M., Fearn, N., & Priebe, S. (2012). Factors predicting the outcome of psychotherapy for borderline personality disorder. *Clinical Psychology Review*, *32*(5), 400–412. <https://doi.org/10.1016/j.cpr.2012.04.004>
- Baron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research. Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., ... Cuijpers, P. (2013). Comparative Efficacy of Seven Psychotherapeutic Interventions for Patients with Depression: A Network Meta-Analysis. *PLoS Medicine*, *10*(5), e1001454. <https://doi.org/10.1371/journal.pmed.1001454>
- Bateman, A., & Fonagy, P. (2009). Randomized Controlled Trial of Outpatient Mentalization-Based Treatment Versus Structured Clinical Management for Borderline Personality Disorder. *American Journal of Psychiatry*, *166*(12), 1355–1364. <https://doi.org/10.1176/appi.ajp.2009.09040539>
- Bateman, A., & Fonagy, P. (2009). *Psychotherapy for borderline personality disorder: mentalization-based treatment*. New York: Oxford University Press.
- Bateman, Anthony, & Fonagy, P. (2010). Mentalization based treatment for borderline personality disorder. *World Psychiatry*, *9*(1), 11–15. <https://doi.org/10.1002/j.2051-5545.2010.tb00255.x>
- Bauer, S., Lambert, M. J., & Nielsen, S. L. (2004). *Clinical Significance Methods: A Comparison*

of Statistical Techniques. *Journal of Personality Assessment*, 82(1), 60–70.  
[https://doi.org/10.1207/s15327752jpa8201\\_11](https://doi.org/10.1207/s15327752jpa8201_11)

Bech, P. (2016). Clinimetric Dilemmas in Outcome Scales for Mental Disorders. *Psychotherapy and Psychosomatics*, 85(6), 323–326. <https://doi.org/10.1159/000448810>

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory-II*. San Antonio, TX: Harcourt Assessment Inc.

Beevers, C. G., & McGeary, J. E. (2012). Therapygenetics: moving towards personalized psychotherapy treatment. *Trends in Cognitive Sciences*, 16(1), 11–12.  
<https://doi.org/10.1016/j.tics.2011.11.004>

Bell, E. C., Marcus, D. K., & Goodlad, J. K. (2013). Are the parts as good as the whole? A meta-analysis of component treatment studies. *Journal of Consulting and Clinical Psychology*, 81(4), 722–736. <https://doi.org/10.1037/a0033004>

Benish, S. G., Quintana, S., & Wampold, B. E. (2011). Culturally adapted psychotherapy and the legitimacy of myth: A direct-comparison meta-analysis. *Journal of Counseling Psychology*, 58(3), 279–289. <https://doi.org/10.1037/a0023626>

Bennett, C. C., & Hauser, K. (2013). Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. *Artificial Intelligence in Medicine*, 57(1), 9–19. <https://doi.org/10.1016/j.artmed.2012.12.003>

Bergfeld, I. O., Mantione, M., Figuee, M., Schuurman, P. R., Lok, A., & Denys, D. (2018). Treatment-resistant depression and suicidality. *Journal of Affective Disorders*, 235, 362–367. <https://doi.org/10.1016/j.jad.2018.04.016>

Bergin, A. E. (1963). The effects of psychotherapy: Negative results revisited. *Journal of Counseling Psychology*, 10(3), 244–250. <https://doi.org/10.1037/h0043353>

Berglar, J., Cramer, A., Von Wyl, A., Koemeda-Lutz, M., Köhler, M., Staczan, P., ... Tschuschke, V. (2016). Therapist Effects on Treatment Outcome in Psychotherapy: A Multilevel Modelling Analysis. *International Journal of Psychotherapy*, 20(2).

Berk, M., & Parker, G. (2009). The elephant on the couch : side - effects of psychotherapy. *Australian and New Zealand Journal of Psychiatry*, 43(August), 787–794.

<https://doi.org/10.1080/00048670903107559>

Bethge, M., Markus, M., Streibelt, M., Gerlich, C., & Schuler, M. (2019). Effects of nationwide implementation of work-related medical rehabilitation in Germany: Propensity score matched analysis. *Occupational and Environmental Medicine*, 76(12), 913–919. <https://doi.org/10.1136/oemed-2019-106065>

Bienenstein, S., & Rother, M. (2007). *Fehler in der Psychotherapie: Theorie Beispiele und Lösungsansätze für die Praxis*. Wien: Springer-Verlag.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Bisson, J. I., Ehlers, A., Matthews, R., Pilling, S., Richards, D., & Turner, S. (2007). Psychological treatments for chronic post-traumatic stress disorder: Systematic review and meta-analysis. *British Journal of Psychiatry*, Vol. 190, pp. 97–104. <https://doi.org/10.1192/bjp.bp.106.021402>

Blackwell, S. E., Woud, M. L., Margraf, J., & Schönbrodt, F. D. (2019). Introducing the Leapfrog Design: A Simple Bayesian Adaptive Rolling Trial Design for Accelerated Treatment Development and Optimization. *Clinical Psychological Science*, 216770261985807. <https://doi.org/10.1177/2167702619858071>

Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and Initial Psychometric Evaluation. *Journal of Traumatic Stress*, 28(6), 489–498. <https://doi.org/10.1002/jts.22059>

Blom, M. B. J., Spinhoven, P., Hoffman, T., Jonker, K., Hoencamp, E., Haffmans, P. M. J., & van Dyck, R. (2007). Severity and duration of depression, not personality factors, predict short term outcome in the treatment of major depression. *Journal of Affective Disorders*, 104(1–3), 119–126. <https://doi.org/10.1016/J.JAD.2007.03.010>

Boettcher, J., Rozental, A., Andersson, G., & Carlbring, P. (2014). Side effects in Internet-based interventions for Social Anxiety Disorder. *Internet Interventions*, 1(1), 3–11. <https://doi.org/10.1016/j.invent.2014.02.002>

Boettcher, J., Schaeuffele, C., & Renneberg, B. (2018). Das Unified Protocol zur Behandlung

emotionaler Störungen. *Psychotherapeut*, 63(3), 182–187.  
<https://doi.org/10.1007/s00278-018-0282-4>

Bohus, M., Herpertz, S. C., & Falkai, P. (2012). Modulare Psychotherapie – Rationale und Grundprinzipien. *Die Psychiatrie*, 9(02), 89–97. <https://doi.org/10.1055/s-0038-1671776>

Bohus, M. (2015). Elfenbeintürme im Treibsand oder: Was macht es so schwierig, Erkenntnisse aus der Forschung in der therapeutischen Praxis umzusetzen? *Verhaltenstherapie*, 25(2), 145–155. <https://doi.org/10.1159/000381154>

Boschloo, L., Bekhuis, E., Weitz, E. S., Reijnders, M., DeRubeis, R. J., Dimidjian, S., ... Cuijpers, P. (2019). The symptom-specific efficacy of antidepressant medication vs. cognitive behavioral therapy in the treatment of depression: results from an individual patient data meta-analysis. *World Psychiatry*, 18(2), 183–191. <https://doi.org/10.1002/wps.20630>

Boswell, J. F., Constantino, M. J., Kraus, D. R., Bugatti, M., & Oswald, J. M. (2016). The Expanding Relevance of Routinely Collected Outcome Data for Mental Health Care Decision Making. *Administration and Policy in Mental Health and Mental Health Services Research*, 43(4), 482–491. <https://doi.org/10.1007/s10488-015-0649-6>

Böttche, M., Ehring, T., Krüger-Gottschalk, A., Rau, H., Schäfer, I., Schellong, J., ... Knaevelsrud, C. (2018). Testing the ICD-11 proposal for complex PTSD in trauma-exposed adults: factor structure and symptom profiles. *European Journal of Psychotraumatology*, 9(1), 1512264. <https://doi.org/10.1080/20008198.2018.1512264>

Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2016). Psychometric properties of the PTSD checklist for diagnostic and statistical manual of mental disorders-fifth edition (PCL-5) in veterans. *Psychological Assessment*, 28(11), 1379–1391. <https://doi.org/10.1037/pas0000254>

Box, G. E. P., & Draper, N. R. (1987). Empirical model-building and response surfaces. Oxford: John Wiley & Sons.

Bradley, R., Greene, J., Russ, E., Dutra, L., & Westen, D. (2005). A Multidimensional Meta-Analysis of Psychotherapy for PTSD. *American Journal of Psychiatry*, 162(2), 214–227. <https://doi.org/10.1176/appi.ajp.162.2.214>



- Brakemeier, E.-L., Herzog, P., Radtke, M., Schneibel, R., Breger, V., Becker, M., ... Normann, C. (2018). CBASP als stationäres Behandlungskonzept der therapieresistenten chronischen Depression: Eine Pilotstudie zum Zusammenhang von Nebenwirkungen und Therapieerfolg. *PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie*, 68(09/10), 399–407. <https://doi.org/10.1055/a-0629-7802>
- Brakemeier, E. L., & Herpertz, S. C. (2019). Innovative psychotherapy research: towards an evidence-based and process-based individualized and modular psychotherapy. *Nervenarzt*, 90(11), 1125–1134. <https://doi.org/10.1007/s00115-019-00808-9>
- Brooks, J., Bratley, R., Jones, L., King, N., & Lucock, M. (2020). Expectations and experiences of psychological therapy from the client perspective: a qualitative study. *British Journal of Guidance & Counselling*. <https://doi.org/10.1080/03069885.2019.1707167>
- Brown, G. S., Lambert, M. J., Jones, E. R., & Minami, T. (2005). Identifying highly effective psychotherapists in a managed care environment. *American Journal of Managed Care*, 11(8), 513–520.
- Brujniks, S. J. E., Lemmens, L. H. J. M., Hollon, S. D., Peeters, F. P. M. L., Cuijpers, P., Arntz, A., ... Huibers, M. J. H. (2020). The effects of once- versus twice-weekly sessions on psychotherapy outcomes in depressed patients. *The British Journal of Psychiatry*, 1–9. <https://doi.org/10.1192/bjp.2019.265>
- Buchkremer, G., & Klingberg, S. (2001). Was ist wissenschaftlich fundierte psychotherapie? Zur diskussion um leitlinien für die psychotherapieforschung. *Nervenarzt*, Vol. 72, pp. 20–30. <https://doi.org/10.1007/s001150050708>
- Buckman, J. E. J., Saunders, R., Cohen, Z. D., Clarke, K., Ambler, G., DeRubeis, R. J., ... Pilling, S. (2019). What factors indicate prognosis for adults with depression in primary care? A protocol for meta-analyses of individual patient data using the Dep-GP database. *Wellcome Open Research*, 4, 69. <https://doi.org/10.12688/wellcomeopenres.15225.1>
- Burlingame, G. M. (2018). Cohesion in Group Therapy: A Meta-Analysis. *Psychotherapy*, 55(4), 384–398. <https://doi.org/10.1037/pst0000173>
- Bzdok, D., Karrer, T. M., Habel, U., & Schneider, F. (2017). Big data approaches in psychiatry: examples in depression research. *Der Nervenarzt*, 89(8), 869–874.

<https://doi.org/10.1007/s00115-017-0456-2>

Cahill, J., Barkham, M., & Stiles, W. B. (2010). Systematic review of practice-based research on psychological therapies in routine clinic settings. *British Journal of Clinical Psychology, 49*(4), 421–453. <https://doi.org/10.1348/014466509X470789>

Camp, R. C. (1989). *Benchmarking : the search for industry best practices that lead to superior performance*. Milwaukee: ASQC Quality Press.

Carter, J. D., Crowe, M. T., Jordan, J., McIntosh, V. V. W., Frampton, C., & Joyce, P. R. (2015). Predictors of response to CBT and IPT for depression; the contribution of therapy process. *Behaviour Research and Therapy, 74*, 72–79. <https://doi.org/10.1016/j.brat.2015.09.003>

Carter, J. D., Luty, S. E., McKenzie, J. M., Mulder, R. T., Frampton, C. M., & Joyce, P. R. (2011). Patient predictors of response to cognitive behaviour therapy and interpersonal psychotherapy in a randomised clinical trial for depression. *Journal of Affective Disorders, 128*(3), 252–261. <https://doi.org/10.1016/j.jad.2010.07.002>

Caspar, F., & Jacobi, F. (2007). Psychotherapieforschung. In W. Hiller, E. Leibing, F. Leichsenring, & S. K. D. Sulz (Eds.), *Lehrbuch der Psychotherapie für die Ausbildung zur/zum Psychologischen PsychotherapeutIn und für ärztliche Weiterbildung, Band 1: Wissenschaftliche Grundlagen der Psychotherapie* (pp. 395–410). München: CIP-Medien.

Castonguay, L. G., & Beutler, L. E. (2006). Principles of therapeutic change: A task force on participants, relationships, and techniques factors. *Journal of Clinical Psychology, 62*(6), 631–638. <https://doi.org/10.1002/jclp.20256>

Castonguay, L. G., & Muran, J. C. (2016). *Practice-oriented research in psychotherapy : building partnerships between clinicians and researchers*. Oxford, New York: Routledge.

Castonguay, L. G., Barkham, M., Lutz, W., & McAleavey, A. A. (2013). Practice-oriented research: Approaches and application. In M. J. Lambert (Ed.), *Bergin & Garfield's handbook of psychotherapy and behavior change* (6th edition). New York: Wiley.

Cepeda, M. S., Reys, J., & Ryan, P. (2018). Finding factors that predict treatment-resistant depression: Results of a cohort study. *Depression and Anxiety, 35*(7), 668–673. <https://doi.org/10.1002/da.22774>

- Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., ... Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry*, 3(3), 243–250. [https://doi.org/10.1016/S2215-0366\(15\)00471-X](https://doi.org/10.1016/S2215-0366(15)00471-X)
- Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, 110, 12–22. <https://doi.org/10.1016/J.JCLINEPI.2019.02.004>
- Churchill, R., Moore, T. H., Furukawa, T. A., Caldwell, D. M., Davies, P., Jones, H., ... Hunot, V. (2013). “Third wave” cognitive and behavioural therapies versus treatment as usual for depression. *Cochrane Database of Systematic Reviews*, (10). <https://doi.org/10.1002/14651858.CD008705.pub2>
- Cipriani, A., & Geddes, J. (2016). Predicting treatment outcome in depression: so far, so good. *The Lancet Psychiatry*, 3(3), 192–194. [https://doi.org/10.1016/S2215-0366\(15\)00542-8](https://doi.org/10.1016/S2215-0366(15)00542-8)
- Citrome, L. (2008). Compelling or irrelevant? Using number needed to treat can help decide. *Acta Psychiatrica Scandinavica*, 117(6), 412–419. <https://doi.org/10.1111/j.1600-0447.2008.01194.x>
- Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry*, 23(4), 318–327. <https://doi.org/10.3109/09540261.2011.606803>
- Clark, D. M. (2018). Realizing the Mass Public Benefit of Evidence-Based Psychological Therapies: The IAPT Program. *Annual Review of Clinical Psychology*, 14(1), 159–183. <https://doi.org/10.1146/annurev-clinpsy-050817-084833>
- Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *The Lancet*, 391(10121), 679–686. [https://doi.org/10.1016/S0140-6736\(17\)32133-5](https://doi.org/10.1016/S0140-6736(17)32133-5)
- Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment Selection in Depression. *Annual Review of Clinical Psychology*, 14, 15.1-15.28. <https://doi.org/10.1146/annurev-clinpsy-050817>

- Cohen, Z. D., Kim, T., Van, H., Dekker, J., & Driessen, E. (2017). *Individual treatment recommendations of cognitive-behavioral or psychodynamic therapy for mild to moderate adult depression*. <https://doi.org/10.17605/OSF.IO/6QXVE> T4 - Improving the Personalized Advantage Index approach M4 - Citavi
- Collins, F. S., & Varmus, H. (2015). A New Initiative on Precision Medicine. *New England Journal of Medicine*, *372*(9), 793–795. <https://doi.org/10.1056/NEJMp1500523>
- Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*, *350*(jan07 4), g7594–g7594. <https://doi.org/10.1136/bmj.g7594>
- Collins, L. M., & Coles, M. E. (2017). Sudden gains in exposure therapy for obsessive-compulsive disorder. *Behaviour Research and Therapy*, *93*, 1–5. <https://doi.org/10.1016/j.brat.2017.03.003>
- Gibbons, M. B. C., Gallop, R., Thompson, D., Luther, D., Crits-Christoph, K., Jacobs, J., ... Crits-Christoph, P. (2016). Comparative Effectiveness of Cognitive Therapy and Dynamic Psychotherapy for Major Depressive Disorder in a Community Mental Health Setting. *JAMA Psychiatry*, *73*(9), 904. <https://doi.org/10.1001/jamapsychiatry.2016.1720>
- Constantino, M. J., Arnkoff, D. B., Glass, C. R., Ametrano, R. M., & Smith, J. A. Z. (2011). Expectations. *Journal of Clinical Psychology*, *67*(2), 184–192. <https://doi.org/10.1002/jclp.20754>
- Constantino, M. J., Coyne, A. E., Boswell, J. F., Iles, B. R., & Visla, A. (2018). A Meta-Analysis of the Association between Patients' Early Perception of Treatment Credibility and Their Posttreatment Outcomes. *Psychotherapy*, *55*(4), 486–495. <https://doi.org/10.1037/pst0000168>
- Constantino, M. J., Manber, R., DeGeorge, J., McBride, C., Ravitz, P., Zuroff, D. C., ... Arnow, B. A. (2008). Interpersonal styles of chronically depressed outpatients: Profiles and therapeutic change. *Psychotherapy: Theory, Research, Practice, Training*, *45*(4), 491–506. <https://doi.org/10.1037/a0014335>
- Constantino, M. J., Visla, A., Coyne, A. E., & Boswell, J. F. (2018). A Meta-Analysis of the Association between Patients' Early Treatment Outcome Expectation and Their

Posttreatment Outcomes. *Psychotherapy*, 55(4), 473–485.  
<https://doi.org/10.1037/pst0000169>

Cook, J. M., Biyanova, T., Elhai, J., Schnurr, P. P., & Coyne, J. C. (2010). What do psychotherapists really do in practice? An internet study of over 2,000 practitioners. *Psychotherapy*, 47(2), 260–267. <https://doi.org/10.1037/a0019788>

Cook, S. C., Schwartz, A. C., & Kaslow, N. J. (2017). Evidence-Based Psychotherapy: Advantages and Challenges. *Neurotherapeutics*, 14(3), 537–545. <https://doi.org/10.1007/s13311-017-0549-4>

Cooper, A. A., & Conklin, L. R. (2015, August 1). Dropout from individual psychotherapy for major depression: A meta-analysis of randomized clinical trials. *Clinical Psychology Review*, Vol. 40, pp. 57–65. <https://doi.org/10.1016/j.cpr.2015.05.001>

Coutanche, M., & Hallion, L. (2019). Machine Learning for Clinical Psychology and Clinical Neuroscience. In A. G. C. Wright & M. N. Hallquist (Eds.), *The Cambridge Handbook of Research Methods in Clinical Psychology*. Cambridge, UK: Cambridge University Press.

Craske, M. G. (2019). Refining our research practices in clinical science: Challenges and steps towards solutions. *Behaviour Research and Therapy*, 116, 90–93. <https://doi.org/10.1016/j.brat.2019.03.006>

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: an inhibitory learning approach. *Behaviour Research and Therapy*, 58, 10–23. <https://doi.org/10.1016/j.brat.2014.04.006>

Crawford, M. J., Thanu, L., Farquharson, L., Palmer, L., Hancock, E., Bassett, P., ... Parry, G. D. (2016). Patient experience of negative effects of psychological treatment: results of a national survey. *The British Journal of Psychiatry*, 208(3), 260–265. <https://doi.org/10.1192/bjp.bp.114.162628>

Creamer, M., Bell, R., & Failla, S. (2003). Psychometric properties of the Impact of Event Scale - Revised. *Behaviour Research and Therapy*, 41(12), 1489–1496. <https://doi.org/10.1016/j.brat.2003.07.010>

Cristea, I. A., Gentili, C., Cotet, C. D., Palomba, D., Barbui, C., & Cuijpers, P. (2017). Efficacy of

Psychotherapies for Borderline Personality Disorder. *JAMA Psychiatry*, 74(4), 319.  
<https://doi.org/10.1001/jamapsychiatry.2016.4287>

Crits-Christoph, P., Hollon, S. D., & Wilson, G. T. (2005). Empirically supported psychotherapies: Comment on Westen, Novotny, and Thompson-Brenner (2004). *Psychological Bulletin*, 131(3), 412–417. <https://doi.org/10.1037/0033-2909.131.3.412>

Cuijpers, P., Karyotaki, E., Reijnders, M., & Ebert, D. D. (2019). Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression. *Epidemiology and Psychiatric Sciences*, 28(1), 21–30. <https://doi.org/10.1017/S2045796018000057>

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychological Medicine*, 40(2), 211–223. <https://doi.org/10.1017/S0033291709006114>

Cuijpers, P. (2019). Targets and outcomes of psychotherapies for mental disorders: an overview. *World Psychiatry*, 18(3), 276–285. <https://doi.org/10.1002/wps.20661>

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A Meta-Analysis of Cognitive-Behavioural Therapy for Adult Depression, Alone and in Comparison with other Treatments. *The Canadian Journal of Psychiatry*, 58(7), 376–385. <https://doi.org/10.1177/070674371305800702>

Cuijpers, P., Karyotaki, E., Reijnders, M., & Ebert, D. D. (2019). Is psychotherapy effective? Pretending everything is fine will not help the field forward. *Epidemiology and Psychiatric Sciences*, 28(03), 356–357. <https://doi.org/10.1017/S204579601800080X>

Cuijpers, P., Li, J., Hofmann, S. G., & Andersson, G. (2010, August). Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: A meta-analysis. *Clinical Psychology Review*, Vol. 30, pp. 768–778. <https://doi.org/10.1016/j.cpr.2010.06.001>

Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C. H., Cipriani, Andrea, & Furukawa, T. A. (2020). A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry*, 19(1), 92–107. <https://doi.org/10.1002/wps.20701>

- Cuijpers, P., van Straten, A., Schuurmans, J., van Oppen, P., Hollon, S. D., & Andersson, G. (2010). Psychotherapy for chronic major depression and dysthymia: A meta-analysis. *Clinical Psychology Review, 30*(1), 51–62. <https://doi.org/10.1016/j.cpr.2009.09.003>
- Cusack, K., Jonas, D. E., Forneris, C. A., Wines, C., Sonis, J., Middleton, J. C., ... Gaynes, B. N. (2016). Psychological treatments for adults with posttraumatic stress disorder: A systematic review and meta-analysis. *Clinical Psychology Review, 43*(290), 128–141. <https://doi.org/10.1016/j.cpr.2015.10.003>
- de Felice, G., Giuliani, A., Halfon, S., Andreassi, S., Paoloni, G., & Orsucci, F. F. (2019). The misleading Dodo Bird verdict. How much of the outcome variance is explained by common and specific factors? *New Ideas in Psychology, 54*, 50–55. <https://doi.org/10.1016/j.newideapsych.2019.01.006>
- de Haan, E., van Oppen, P., van Balkom, A. J., Spinhoven, P., Hoogduin, K. A., & Van Dyck, R. (1997). Prediction of outcome and early vs. late improvement in OCD patients treated with cognitive behaviour therapy and pharmacotherapy. *Acta Psychiatrica Scandinavica, 96*(5), 354–361.
- De Panfilis, C., Marchesi, C., Cabrino, C., Monici, A., Politi, V., Rossi, M., & Maggini, C. (2012). Patient factors predicting early dropout from psychiatric outpatient care for borderline personality disorder. *Psychiatry Research, 200*(2–3), 422–429. <https://doi.org/10.1016/j.psychres.2012.03.016>
- De Smet, M. M., Meganck, R., De Geest, R., Norman, U. A., Truijens, F., & Desmet, M. (2020). What “good outcome” means to patients: Understanding recovery and improvement in psychotherapy for major depression from a mixed-methods perspective. *Journal of Counseling Psychology, 67*(1), 25–39. <https://doi.org/10.1037/cou0000362>
- DeCou, C. R., Comtois, K. A., & Landes, S. J. (2019). Dialectical Behavior Therapy Is Effective for the Treatment of Suicidal Behavior: A Meta-Analysis. *Behavior Therapy, 50*(1), 60–72. <https://doi.org/10.1016/J.BETH.2018.03.009>
- Deisenhofer, A.-K., Delgado, J., Rubel, J. A., Böhnke, J. R., Zimmermann, D., Schwartz, B., & Lutz, W. (2018). Individual treatment selection for patients with posttraumatic stress disorder. *Depression and Anxiety, 35*(6), 541–550. <https://doi.org/10.1002/da.22755>

- Delgadillo, J., Appleby, S., Booth, S., Burnett, G., Carey, A., Edmeade, L., ... Lutz, W. (2020). The Leeds Risk Index: Field-Test of a Stratified Psychological Treatment Selection Algorithm. *Psychotherapy and Psychosomatics*, 1–2. <https://doi.org/10.1159/000505193>
- Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., ... McMillan, D. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: a multisite, open-label, cluster randomised controlled trial. *The Lancet Psychiatry*, 5(7), 564–572. [https://doi.org/10.1016/S2215-0366\(18\)30162-7](https://doi.org/10.1016/S2215-0366(18)30162-7)
- Delgadillo, J., Gellatly, J., & Stephenson-Bellwood, S. (2015). Decision making in stepped care: How do therapists decide whether to prolong treatment or not? *Behavioural and Cognitive Psychotherapy*, 43(3), 328–341. <https://doi.org/10.1017/S135246581300091X>
- Delgadillo, J., & Lutz, W. (2020). A Development Pathway towards Precision Mental Health Care. *JAMA Psychiatry*. <https://doi.org/10.1001/jamapsychiatry.2020.1048>
- Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy*, 79, 15–22. <https://doi.org/10.1016/j.brat.2016.02.003>
- Delgadillo, J., Overend, K., Lucock, M., Groom, M., Kirby, N., McMillan, D., ... de Jong, K. (2017). Improving the efficiency of psychological treatment using outcome feedback technology. *Behaviour Research and Therapy*, 99, 89–97. <https://doi.org/10.1016/j.brat.2017.09.011>
- Delgadillo, J., Rubel, J., & Barkham, M. (2020). Towards Personalized Allocation of Patients to Therapists. *Journal of Consulting and Clinical Psychology*. <https://doi.org/10.1037/ccp0000507>
- DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index. *PLOS ONE*, 9(1), e83875. <https://doi.org/10.1371/journal.pone.0083875>
- Dewan, N. A., Daniels, A., Zieman, G., & Kramer, T. (2000). The national outcomes management project: A benchmarking collaborative. *Journal of Behavioral Health Services and Research*, 27(4), 431–436. <https://doi.org/10.1007/BF02287824>
- DGPPN, BÄK, KBV, & AWMF. (2017). *S3-Leitlinien/Nationale VersorgungsLeitlinie: Unipolare*



*Depression* (2. Auflage). Berlin, Heidelberg: Springer.

DiBernardo, A., Lin, X., Zhang, Q., Xiang, J., Lu, L., Jamieson, C., ... Li, G. (2018). Humanistic outcomes in treatment resistant depression: a secondary analysis of the STAR\*D study. *BMC Psychiatry, 18*(1), 352. <https://doi.org/10.1186/s12888-018-1920-7>

Doering, B. K., Glombiewski, J. A., & Rief, W. (2018). Expectation-Focused Psychotherapy to Improve Clinical Outcomes. *International Review of Neurobiology, 138*, 257–270. <https://doi.org/10.1016/BS.IRN.2018.02.004>

Donenberg, G. R., Lyons, J. S., & Howard, K. I. (1999). Clinical trials versus mental health services research: Contributions and connections. *Journal of Clinical Psychology, 55*(9), 1135–1146. [https://doi.org/10.1002/\(SICI\)1097-4679\(199909\)55:9<1135::AID-JCLP10>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-4679(199909)55:9<1135::AID-JCLP10>3.0.CO;2-O)

Driessen, E., Van, H. L., Don, F. J., Peen, J., Kool, S., Westra, D., ... Dekker, J. J. M. (2013). The Efficacy of Cognitive-Behavioral Therapy and Psychodynamic Therapy in the Outpatient Treatment of Major Depression: A Randomized Clinical Trial. *American Journal of Psychiatry, 170*(9), 1041–1050. <https://doi.org/10.1176/appi.ajp.2013.12070899>

Dunlop, B., Kaye, J., Youngner, C., & Rothbaum, B. (2014). Assessing treatment-resistant Posttraumatic Stress Disorder: The Emory Treatment Resistance Interview for PTSD (E-TRIP). *Behavioral Sciences, 4*(4), 511–527. <https://doi.org/10.3390/bs4040511>

Durham, R. C., Fisher, P. L., Dow, M. G. T., Sharp, D., Power, K. G., Swan, J. S., & Morton, R. V. (2004). Cognitive behaviour therapy for good and poor prognosis generalized anxiety disorder: A clinical effectiveness study. *Clinical Psychology and Psychotherapy, 11*(3), 145–157. <https://doi.org/10.1002/cpp.405>

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology, 14*(1), 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>

Eddy, K. T., Dutra, L., Bradley, R., & Westen, D. (2004). A multidimensional meta-analysis of psychotherapy and pharmacotherapy for obsessive-compulsive disorder. *Clinical Psychology Review, 24*(8), 1011–1030. <https://doi.org/10.1016/j.cpr.2004.08.004>

Eftekhari, A., Ruzek, J. I., Crowley, J. J., Rosen, C. S., Greenbaum, M. A., & Karlin, B. E. (2013).

Effectiveness of National Implementation of Prolonged Exposure Therapy in Veterans Affairs Care. *JAMA Psychiatry*, 70(9), 949. <https://doi.org/10.1001/jamapsychiatry.2013.36>

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(2), 269–276. <https://doi.org/10.1016/j.cpr.2009.12.001>

Ehlers, A., Grey, N., Wild, J., Stott, R., Liness, S., Deale, A., ... Clark, D. M. (2013). Implementation of Cognitive Therapy for PTSD in routine clinical care: Effectiveness and moderators of outcome in a consecutive sample. *Behaviour Research and Therapy*, 51(11), 742–752. <https://doi.org/10.1016/j.brat.2013.08.006>

Eisen, S. V., Ranganathan, G., Seal, P., & Spiro, A. (2007). Measuring clinically meaningful change following mental health treatment. *Journal of Behavioral Health Services and Research*, 34(3), 272–289. <https://doi.org/10.1007/s11414-007-9066-2>

Elkin, I. (1994). The NIMH Treatment of Depression Collaborative Research Program: Where we began and where we are. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change*, 4th ed. (pp. 114–139). Oxford, England: John Wiley & Sons.

Elliott, H., Jones, P. J., & Schmidt, U. (2019). Central Symptoms Predict Posttreatment Outcomes and Clinical Impairment in Anorexia Nervosa: A Network Analysis. *Clinical Psychological Science*, 8, 139–154. <https://doi.org/10.1177/2167702619865958>

Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. *Psychotherapy*, 48(1), 43–49. <https://doi.org/10.1037/a0022187>

Elliott, R., Bohart, A. C., Watson, J. C., & Murphy, D. (2018). Therapist Empathy and Client Outcome: An Updated Meta-Analysis. *Psychotherapy*, 55(4), 399–410. <https://doi.org/10.1037/pst0000175>

Enns, M. W., Cox, B. J., & Pidlubny, S. R. (2002). Group Cognitive Behaviour Therapy for Residual Depression: Effectiveness and Predictors of Response. *Cognitive Behaviour Therapy*, 31(1), 31–40. <https://doi.org/10.1080/16506070252823643>

Erichson, N. B., Zheng, P., Manohar, K., Brunton, S. L., Kutz, J. N., & Aravkin, A. Y. (2018). *Sparse*

*Principal Component Analysis via Variable Projection*. Retrieved from <http://arxiv.org/abs/1804.00341>

Eubanks, C. F., Muran, J. C., & Safran, J. D. (2018). Alliance Rupture Repair: A Meta-Analysis. *Psychotherapy, 55*(4), 508–519. <https://doi.org/10.1037/pst0000185>

Eysenck, H. J. (1952). The effects of psychotherapy: an evaluation. *Journal of Consulting Psychology, 16*(5), 319–324. <https://doi.org/10.1037/h0063633>

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist, Vol. 33*, p. 517. <https://doi.org/10.1037/0003-066X.33.5.517.a>

Falkenstein, M. J., Nota, J. A., Krompinger, J. W., Schreck, M., Garner, L. E., Potluri, S., ... Elias, J. A. (2019). Empirically-derived response trajectories of intensive residential treatment in obsessive-compulsive disorder: A growth mixture modeling approach. *Journal of Affective Disorders, 245*, 827–833. <https://doi.org/10.1016/j.jad.2018.11.075>

Falkenström, F., Josefsson, A., Berggren, T., & Holmqvist, R. (2016). How much therapy is enough? Comparing dose-effect and good-enough models in two different settings. *Psychotherapy, 53*(1), 130–139. <https://doi.org/10.1037/pst0000039>

Farber, B. A., & Doolin, E. M. (2011). Positive Regard. *Psychotherapy, 48*(1), 58–64. <https://doi.org/10.1037/a0022141>

Farber, B. A., Suzuki, J. Y., & Lynch, D. A. (2018). Positive Regard and Psychotherapy Outcome: A Meta-Analytic Review. *Psychotherapy, 55*(4), 411–423. <https://doi.org/10.1037/pst0000171>

Fava, G. A., Rafanelli, C., & Tomba, E. (2012). The Clinical Process in Psychiatry. *The Journal of Clinical Psychiatry, 73*(02), 177–184. <https://doi.org/10.4088/JCP.10r06444>

Fava, M., Evins, A. E., Dorer, D. J., & Schoenfeld, D. A. (2003). The problem of the placebo response in clinical trials for psychiatric disorders: Culprits, possible remedies, and a novel study design approach. *Psychotherapy and Psychosomatics, 72*(3), 115–127. <https://doi.org/10.1159/000069738>

Federici, A., Summerfeldt, L. J., Harrington, J. L., McCabe, R. E., Purdon, C. L., Rowa, K., & Antony Martin M., M. M. (2010). Consistency between self-report and clinician-administered

versions of the Yale-Brown Obsessive-Compulsive Scale. *Journal of Anxiety Disorders*, 24(7), 729–733. <https://doi.org/10.1016/j.janxdis.2010.05.005>

Feinstein, A. R. (1987). *Clinimetrics*. New Haven and London: Yale University Press.

Fennell, M. J. V., & Teasdale, J. D. (1987). Cognitive therapy for depression: Individual differences and the process of change. *Cognitive Therapy and Research*, 11(2), 253–271. <https://doi.org/10.1007/BF01183269>

Fernandez, E., Salem, D., Swift, J. K., & Ramtahal, N. (2015). Meta-analysis of dropout from cognitive behavioral therapy: Magnitude, timing, and moderators. *Journal of Consulting and Clinical Psychology*, 83(6), 1108–1122. <https://doi.org/10.1037/ccp0000044>

Fisher, A. J., Bosley, H. G., Fernandez, K. C., Reeves, J. W., Soyster, P. D., Diamond, A. E., & Barkin, J. (2019). Open trial of a personalized modular treatment for mood and anxiety. *Behaviour Research and Therapy*, 116, 69–79. <https://doi.org/10.1016/j.brat.2019.01.010>

Fisher, A. J., & Boswell, J. F. (2016). Enhancing the Personalization of Psychotherapy With Dynamic Assessment and Modeling. *Assessment*, 23(4), 496–506. <https://doi.org/10.1177/1073191116638735>

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of Abnormal Psychology*, 126(8), 1044–1056. <https://doi.org/10.1037/abn0000311>

Fisher, P. L., & Wells, A. (2005). How effective are cognitive and behavioral treatments for obsessive–compulsive disorder? A clinical significance analysis. *Behaviour Research and Therapy*, 43(12), 1543–1558. <https://doi.org/10.1016/j.brat.2004.11.007>

Flückiger, C., Del, A. C., Wampold, B. E., & Horvath, A. O. (2018). The Alliance in Adult Psychotherapy: A Meta-Analytic Synthesis. *Psychotherapy*, 55(4), 316–340. <https://doi.org/10.1037/pst0000172>

Foa, E. B., McLean, C. P., Zang, Y., Zhong, J., Powers, M. B., Kauffman, B. Y., ... Knowles, K. (2016). Psychometric properties of the posttraumatic diagnostic scale for DSM-5 (PDS-5). *Psychological Assessment*, 28(10), 1166–1171. <https://doi.org/10.1037/pas0000258>

Fonagy, P. (2010). Psychotherapy research: do we know what works for whom? *The British*

*Journal of Psychiatry*, 197(2).

Forman, E. M., Herbert, J. D., Moitra, E., Yeomans, P. D., & Geller, P. A. (2007). A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior Modification*, 31(6), 772–799. <https://doi.org/10.1177/0145445507302202>

Fourie, D. P. (1996). The research/practice gap in psychotherapy: From discovering reality to making sense. *Journal of Contemporary Psychotherapy*, 26(1), 7–22. <https://doi.org/10.1007/BF02307702>

Frank, E., Cassano, G. B., Rucci, P., Thompson, W. K., Kraemer, H. C., Fagiolini, A., ... Forgiione, R. N. (2011). Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. *Psychological Medicine*, 41(1), 151–162. <https://doi.org/10.1017/S0033291710000553>

Frank, F., Wilk, J., Kriston, L., Meister, R., Shimodera, S., Hesse, K., ... Hölzel, L. P. (2015). Effectiveness of a brief psychoeducational group intervention for relatives on the course of disease in patients after inpatient depression treatment compared with treatment as usual – study protocol of a multisite randomised controlled trial. *BMC Psychiatry*, 15(1), 259. <https://doi.org/10.1186/s12888-015-0633-4>

Franklin, M. E., Abramowitz, J. S., Kozak, M. J., Levitt, J. T., & Foa, E. B. (2000). Effectiveness of exposure and ritual prevention for obsessive-compulsive disorder: Randomized compared with nonrandomized samples. *Journal of Consulting and Clinical Psychology*, 68(4), 594–602. <https://doi.org/10.1037/0022-006X.68.4.594>

Freeman, C., & Power, M. J. (2007). *Handbook of evidence-based psychotherapies : a guide for research and practice*. Chichester, UK: Wiley.

Friedlander, M. L., Escudero, V., Welmers-Van De Poll, M. J., & Heatherington, L. (2018). Meta-Analysis of the Alliance-Outcome Relation in Couple and Family Therapy. *Psychotherapy*, 55(4), 356–371. <https://doi.org/10.1037/pst0000161>

Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: The brain as a phantastic organ. *The Lancet Psychiatry*, 1(2), 148–158. [https://doi.org/10.1016/S2215-0366\(14\)70275-5](https://doi.org/10.1016/S2215-0366(14)70275-5)

- Gaab, J., Kossowsky, J., Ehlert, U., & Locher, C. (2019). Effects and Components of Placebos with a Psychological Treatment Rationale – Three Randomized-Controlled Studies. *Scientific Reports*, 9(1), 1–8. <https://doi.org/10.1038/s41598-018-37945-1>
- Gail, M. H., Wieand, S., & Piantadosi, S. (1984). Biased Estimates of Treatment Effect in Randomized Experiments with Nonlinear Regressions and Omitted Covariates. *Biometrika*, 71(3), 431. <https://doi.org/10.2307/2336553>
- Gaston, L., Marmar, C. R., Gallagher, D., & Thompson, L. W. (1989). Impact of confirming patient expectations of change processes in behavioral, cognitive, and brief dynamic psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, 26(3), 296–302. <https://doi.org/10.1037/h0085439>
- Gaynor, S. T., Weersing, V. R., Kolko, D. J., Birmaher, B., Heo, J., & Brent, D. A. (2003). The prevalence and impact of large sudden improvements during adolescent therapy for depression: a comparison across cognitive-behavioral, family, and supportive therapy. *Journal of Consulting and Clinical Psychology*, 71(2), 386–393. <https://doi.org/10.1037/0022-006x.71.2.386>
- Gelso, C. J., Kivlighan, D. M., & Markin, R. D. (2018). The Real Relationship and Its Role in Psychotherapy Outcome: A Meta-Analysis. *Psychotherapy*, 55(4), 434–444. <https://doi.org/10.1037/pst0000183>
- Gersh, E., Hallford, D. J., Rice, S. M., Kazantzis, N., Gersh, H., Gersh, B., & McCarty, C. A. (2017). Systematic review and meta-analysis of dropout rates in individual psychotherapy for generalized anxiety disorder. *Journal of Anxiety Disorders*, Vol. 52, pp. 25–33. <https://doi.org/10.1016/j.janxdis.2017.10.001>
- Geurtzen, N., Keijsers, G. P. J., Karremans, J. C., & Hutschemaekers, G. J. M. (2019). Care dependency may help and hurt psychological treatment: A treatment-analogue study with students in clinical training. *Journal of Psychotherapy Integration*, 29(4), 374–388. <https://doi.org/10.1037/int0000150>
- Giesen-Bloo, J., van Dyck, R., Spinhoven, P., van Tilburg, W., Dirksen, C., van Asselt, T., ... Arntz, A. (2006). Outpatient Psychotherapy for Borderline Personality Disorder. *Archives of General Psychiatry*, 63(6), 649. <https://doi.org/10.1001/archpsyc.63.6.649>

- Gillespie, K., Duffy, M., Hackmann, A., & Clark, D. M. (2002). Community based cognitive therapy in the treatment of posttraumatic stress disorder following the Omagh bomb. *Behaviour Research and Therapy*, 40(4), 345–357. [https://doi.org/10.1016/s0005-7967\(02\)00004-9](https://doi.org/10.1016/s0005-7967(02)00004-9)
- Glass, C. R., Arnkoff, D. B., & Shapiro, S. J. (2001). Expectations and preferences. *Psychotherapy*, 38(4), 455–461. <https://doi.org/10.1037/0033-3204.38.4.455>
- Goldberg, S. B., Hoyt, W. T., Nissen-Lie, H. A., Nielsen, S. L., & Wampold, B. E. (2018). Unpacking the therapist effect: Impact of treatment length differs for high- and low-performing therapists. *Psychotherapy Research*, 28(4), 532–544. <https://doi.org/10.1080/10503307.2016.1216625>
- Goldfried, M. R. (1980). Toward the delineation of therapeutic change principles. *American Psychologist*, 35(11), 991–999. <https://doi.org/10.1037/0003-066X.35.11.991>
- Goldfried, M. R. (2010). The Future of Psychotherapy Integration: Closing the Gap Between Research and Practice. *Journal of Psychotherapy Integration*, 20(4), 386–396. <https://doi.org/10.1037/a0022036>
- Goldfried, M. R. (2019). Obtaining consensus in psychotherapy: What holds us back? *American Psychologist*, 74(4), 484–496. <https://doi.org/10.1037/amp0000365>
- Gönner, S., Limbacher, K., & Ecker, W. (2012). Stationäre kognitive Verhaltenstherapie bei Zwangsstörungen: Effektivität und Erfolgsprädiktoren in der Routineversorgung. *Verhaltenstherapie*, 22(1), 17–26. <https://doi.org/10.1159/000335776>
- Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., ... Charney, D. S. (1989). The Yale-Brown Obsessive Compulsive Scale. *Archives of General Psychiatry*, 46(11), 1006. <https://doi.org/10.1001/archpsyc.1989.01810110048007>
- Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3), 424. <https://doi.org/10.2307/1912791>
- Grant, N., Hotopf, M., Breen, G., Cleare, A., Grey, N., Hepgul, N., ... Tylee, A. (2014). Predicting outcome following psychological therapy in IAPT (PROMPT): a naturalistic project protocol. *BMC Psychiatry*, 14(1), 170. <https://doi.org/10.1186/1471-244X-14-170>

- Grawe, K. (1994). *Psychotherapie im Wandel: Von der Konfession zur Profession*. Göttingen: Hogrefe.
- Grawe, K. (1997). Research-informed psychotherapy. *Psychotherapy Research*, 7(1), 1–19. Retrieved from <https://psycnet.apa.org/record/1997-04327-001>
- Greenberg, R. P., Constantino, M. J., & Bruce, N. (2006). Are patient expectations still relevant for psychotherapy process and outcome? *Clinical Psychology Review*, 26(6), 657–678. <https://doi.org/10.1016/j.cpr.2005.03.002>
- Grøtten, T., Hansen, B., Haseeth, S., Vogel, P. A., Guzey, I. C., & Solem, S. (2018). Three-Week Inpatient Treatment of Obsessive-Compulsive Disorder: A 6-Month Follow-Up Study. *Frontiers in Psychology*, 9, 620. <https://doi.org/10.3389/fpsyg.2018.00620>
- Groves, S. J., Douglas, K. M., & Porter, R. J. (2018). A Systematic Review of Cognitive Predictors of Treatment Outcome in Major Depression. *Frontiers in Psychiatry*, 9. <https://doi.org/10.3389/fpsyg.2018.00382>
- Guidi, J., Brakemeier, E. L., Bockting, C. L. H., Cosci, F., Cuijpers, P., Jarrett, R. B., ... Fava, G. A. (2018). Methodological Recommendations for Trials of Psychological Interventions. *Psychotherapy and Psychosomatics*, 87(5), 276–284. <https://doi.org/10.1159/000490574>
- Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy*, 51(9), 597–606. <https://doi.org/10.1016/j.brat.2013.06.004>
- Haas, E., Hill, R. D., Lambert, M. J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains? *Journal of Clinical Psychology*, 58(9), 1157–1172. <https://doi.org/10.1002/jclp.10044>
- Hahlweg, K., Fiegenbaum, W., Frank, M., Schroeder, B., & von Witzleben, I. (2001). Short- and long-term effectiveness of an empirically supported treatment for agoraphobia. *Journal of Consulting and Clinical Psychology*, 69(3), 375–382. <https://doi.org/10.1037//0022-006x.69.3.375>
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 23(1), 56–62.



- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*(2), 155–163. <https://doi.org/10.1002/jclp.20108>
- Hannöver, W., Richard, M., Hansen, N. B., Martinovich, Z., & Kordy, H. (2002). A classification tree model for decision-making in clinical practice: An application based on the data of the german multicenter study on eating disorders, project tr-eat. *Psychotherapy Research, 12*(4), 445–461. <https://doi.org/10.1080/713664470>
- Hansen, N. B., Lambert, M. J., & Forman, E. M. (2006). The Psychotherapy Dose-Response Effect and Its Implications for Treatment Delivery Services. *Clinical Psychology: Science and Practice, 9*(3), 329–343. <https://doi.org/10.1093/clipsy.9.3.329>
- Harmon, C., Hawkins, E. J., Lambert, M. J., Slade, K., & Whipple, J. S. (2005). Improving outcomes for poorly responding clients: The use of clinical support tools and feedback to clients. *Journal of Clinical Psychology, 61*(2), 175–185. <https://doi.org/10.1002/jclp.20109>
- Harnett, P. H., Whittingham, K., Puhakka, E., Hodges, J., Spry, C., & Dob, R. (2010). The Short-Term Impact of a Brief Group-Based Mindfulness Therapy Program on Depression and Life Satisfaction. *Mindfulness, 1*(3), 183–188. <https://doi.org/10.1007/s12671-010-0024-3>
- Harnett, P., O'Donovan, A., & Lambert, M. J. (2010). The dose response relationship in psychotherapy: Implications for social policy. *Clinical Psychologist, 14*(2), 39–44. <https://doi.org/10.1080/13284207.2010.500309>
- Harrell, F. (2015). *Regression Modeling Strategies - With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Basel, CH: Springer.
- Hatfield, D., McCullough, L., Frantz, S. H. B., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology and Psychotherapy, 17*(1), 25–32. <http://doi.wiley.com/10.1002/cpp.656>
- Hauck, W. W., Anderson, S., & Marcus, S. M. (1998). Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Controlled Clinical Trials, 19*(3), 249–256. [https://doi.org/10.1016/S0197-2456\(97\)00147-5](https://doi.org/10.1016/S0197-2456(97)00147-5)
- Hautzinger, M., Leuzinger-Bohleber, M., Fiedler, G., Bahrke, U., Kallenbach-Kaminski, L.,

Kaufhold, J., ... Beutel, M. (2019). What does the LAC depression study now really reveal? Response to the critical considerations of the methodology. *Psychotherapeut*, *65*(1), 27–31. <https://doi.org/10.1007/s00278-019-00388-8>

Hayes, J. A., Gelso, C. J., Goldberg, S., & Kivlighan, D. M. (2018). Countertransference Management and Effective Psychotherapy: Meta-Analytic Findings. *Psychotherapy*, *55*(4), 496–507. <https://doi.org/10.1037/pst0000189>

Hayes, S. C., & Hofmann, S. G. (2018). *Process-based CBT: the science and core clinical competencies of cognitive behavioral therapy*. Oakland, CA: New Harbinger.

Heinonen, E., & Nissen-Lie, H. A. (2019). The professional and personal characteristics of effective psychotherapists: a systematic review. *Psychotherapy Research*, 1–16. <https://doi.org/10.1080/10503307.2019.1620366>

Hernández, A. V., Eijkemans, M. J. C., & Steyerberg, E. W. (2006). Randomized controlled trials with time-to-event outcomes: How much does prespecified covariate adjustment increase power? *Annals of Epidemiology*, *16*(1), 41–48. <https://doi.org/10.1016/j.annepidem.2005.09.007>

Hernández, A. V., Steyerberg, E. W., & Habbema, J. D. F. (2004). Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *Journal of Clinical Epidemiology*, *57*(5), 454–460. <https://doi.org/10.1016/j.jclinepi.2003.09.014>

Higgins, J. P. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.). The Cochrane Collaboration.

Hill, C. E., Knox, S., & Pinto-Coelho, K. G. (2018). Therapist Self-Disclosure and Immediacy: A Qualitative Meta-Analysis. *Psychotherapy*, *55*(4), 445–460. <https://doi.org/10.1037/pst0000182>

Hill, C. E., Nutt-Williams, E., Heaton, K. J., Thompson, B. J., & Rhodes, R. H. (1996). Therapist Retrospective Recall of Impasses in Long-Term Psychotherapy: A Qualitative Analysis. *Journal of Counseling Psychology*, *43*(2), 207–217. <https://doi.org/10.1037/0022-0167.43.2.207>

- Hill, C. E., Thompson, B. J., Cogar, M. C., & Denman, D. W. (1993). Beneath the Surface of Long-Term Therapy: Therapist and Client Report of Their Own and Each Other's Covert Processes. *Journal of Counseling Psychology, 40*(3), 278–287. <https://doi.org/10.1037/0022-0167.40.3.278>
- Hiller, W., Bleichhardt, G., & Schindler, A. (2009). Evaluation von Psychotherapien aus der Perspektive von Qualitätssicherung und Qualitätsmanagement. *Zeitschrift Fur Psychiatrie, Psychologie Und Psychotherapie, 57*, 7–22. <https://doi.org/10.1024/1661-4747.57.1.7>
- Hiller, W., & Schindler, A. (2011). Response und Remission in der Psychotherapieforschung. *PPmP Psychotherapie Psychosomatik Medizinische Psychologie, 61*(3–4), 170–176. <https://doi.org/10.1055/s-0030-1255040>
- Hiller, W., Schindler, A., Andor, T., & Rist, F. (2011). Vorschläge zur Evaluation regulärer Psychotherapien an Hochschulambulanzen im Sinne der Phase-IV-Therapieforschung. *Zeitschrift Für Klinische Psychologie Und Psychotherapie, 40*(1), 22–32. <https://doi.org/10.1026/1616-3443/a000063>
- Hiller, W., Schindler, A. C., & Lambert, M. J. (2012). Defining response and remission in psychotherapy research: A comparison of the RCI and the method of percent improvement. *Psychotherapy Research, 22*(1), 1–11. <https://doi.org/10.1080/10503307.2011.616237>
- Hofheinz, C., Reder, M., & Michalak, J. (2019). How specific is cognitive change? A randomized controlled trial comparing brief cognitive and mindfulness interventions for depression. *Psychotherapy Research, 30*(5), 675–691. <https://doi.org/10.1080/10503307.2019.1685138>
- Hofmann, S. G., & Hayes, S. C. (2019). The Future of Intervention Science: Process-Based Therapy. *Clinical Psychological Science, 7*(1), 37–50. <https://doi.org/10.1177/2167702618772296>
- Hofmann, S. G., Sawyer, A. T., Witt, A. A., & Oh, D. (2010). The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*(2), 169–183. <https://doi.org/10.1037/a0018555>
- Hohagen, F. (2011). Suizidalität in der Borderline-Behandlung: Non-Suzid-Vertrag oder Non-

Suizid-Entschluss? *Verhaltenstherapie*, 21(3), 200–203.  
<https://doi.org/10.1159/000331455>

Hohwy, J. (2014). *The Predictive Mind*. Oxford: Oxford University Press.

Høifødt, R. S., Mittner, M., Lillevoll, K., Katla, S. K., Kolstrup, N., Eisemann, M., ... Waterloo, K. (2015). Predictors of Response to Web-Based Cognitive Behavioral Therapy With High-Intensity Face-to-Face Therapist Guidance for Depression: A Bayesian Analysis. *Journal of Medical Internet Research*, 17(9), e197. <https://doi.org/10.2196/jmir.4351>

Holmbeck, G. N. (1997, August). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology*, 65, 599–610. <https://doi.org/10.1037/0022-006X.65.4.599>

Holmes, E. A., Craske, M. G., & Graybiel, A. M. (2014). Psychological treatments: A call for mental-health science. *Nature*, 511, 287–289. <https://doi.org/10.1038/511287a>

Holmes, E. A., Ghaderi, A., Harmer, C. J., Ramchandani, P. G., Cuijpers, P., Morrison, A. P., ... Craske, M. G. (2018). The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. *The Lancet Psychiatry*, 5(3), 237–286. [https://doi.org/10.1016/S2215-0366\(17\)30513-8](https://doi.org/10.1016/S2215-0366(17)30513-8)

Holmqvist, R., Philips, B., & Barkham, M. (2015). Developing practice-based evidence: Benefits, challenges, and tensions. *Psychotherapy Research*, 25(1), 20–31. <https://doi.org/10.1080/10503307.2013.861093>

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in Individual Psychotherapy. *Psychotherapy*, 48(1), 9–16. <https://doi.org/10.1037/a0022186>

Houghton, S., Saxon, D., Bradburn, M., Ricketts, T., & Hardy, G. (2010). The effectiveness of routinely delivered cognitive behavioural therapy for obsessive-compulsive disorder: A benchmarking study. *British Journal of Clinical Psychology*, 49(4), 473–489. <https://doi.org/10.1348/014466509X475414>

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy. Efficacy, effectiveness, and patient progress. *The American Psychologist*,

51(10), 1059–1064.

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The Dose-Effect Relationship in Psychotherapy. *American Psychologist*, 41(2), 159–164. <https://doi.org/10.1037/0003-066X.41.2.159>

Huibers, M. J. H., Cohen, Z. D., Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Cuijpers, P., & DeRubeis, R. J. (2015). Predicting Optimal Outcomes in Cognitive Therapy or Interpersonal Psychotherapy for Depressed Individuals Using the Personalized Advantage Index Approach. *PLOS ONE*, 10(11), e0140771. <https://doi.org/10.1371/journal.pone.0140771>

Hundt, N. E., Calleo, J. S., Williams, W., & Cully, J. A. (2016). Does using cognitive-behavioural therapy skills predict improvements in depression? *Psychology and Psychotherapy*, 89(2), 235–238. <https://doi.org/10.1111/papt.12065>

Hung, Y. H., Cheng, C. M., Lin, W. C., Bai, Y. M., Su, T. P., Li, C. T., ... Chen, M. H. (2019). Post-traumatic stress disorder and asthma risk: A nationwide longitudinal study. *Psychiatry Research*, 276, 25–30. <https://doi.org/10.1016/j.psychres.2019.04.014>

Hunot, V., Moore, T. H., Caldwell, D. M., Furukawa, T. A., Davies, P., Jones, H., ... Churchill, R. (2013). “Third wave” cognitive and behavioural therapies versus other psychological therapies for depression. *Cochrane Database of Systematic Reviews*, (10). <https://doi.org/10.1002/14651858.CD008704.pub2>

Hunsley, J., & Lee, C. M. (2007). Research-informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice*, 38, 21–33. <https://doi.org/10.1037/0735-7028.38.1.21>

Husen, K., Rafaeli, E., Rubel, J. A., Bar-Kalifa, E., & Lutz, W. (2016). Daily affect dynamics predict early response in CBT: Feasibility and predictive validity of EMA for outpatient psychotherapy. *Journal of Affective Disorders*, 206, 305–314. <https://doi.org/10.1016/J.JAD.2016.08.025>

Huys, Q. J. M. (2017). Computational Psychiatry. *Zeitschrift Für Psychiatrie, Psychologie Und Psychotherapie*, 65(1), 21–26. <https://doi.org/10.1024/1661-4747/a000297>

- Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, *19*(3), 404–413. <https://doi.org/10.1038/nn.4238>
- Ilardi, S. S., & Craighead, W. E. (1994). The Role of Nonspecific Factors in Cognitive-Behavior Therapy for Depression. *Clinical Psychology: Science and Practice*, *1*(2), 138–155. <https://doi.org/10.1111/j.1468-2850.1994.tb00016.x>
- Ilardi, S. S., & Craighead, W. E. (2006). Rapid Early Response, Cognitive Modification, and Nonspecific Factors in Cognitive Behavior Therapy for Depression: A Reply to Tang and DeRubeis. *Clinical Psychology: Science and Practice*, *6*(3), 295–299. <https://doi.org/10.1093/clipsy.6.3.295>
- Imel, Z. E., Sheng, E., Baldwin, S. A., & Atkins, D. C. (2015). Removing very low-performing therapists: A simulation of performance-based retention in psychotherapy. *Psychotherapy*, *52*(3), 329–336. <https://doi.org/10.1037/pst0000023>
- Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., ... Uher, R. (2016). Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *Journal of Psychiatric Research*, *78*, 94–102. <https://doi.org/10.1016/j.jpsychires.2016.03.016>
- Ioannidis, J. P. A., Evans, S. J. W., Gøtzsche, P. C., O'Neill, R. T., Altman, D. G., Schulz, K., ... CONSORT Group. (2004). Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Annals of Internal Medicine*, *141*(10), 781–788.
- Ionescu, D. F., Niciu, M. J., Henter, I. D., Zarate, C. A., & Jr. (2013). Defining anxious depression: a review of the literature. *CNS Spectrums*, *18*(5), 252–260. <https://doi.org/10.1017/S1092852913000114>
- Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., ... Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*, *64*(2), 295–304. <https://doi.org/10.1037//0022-006x.64.2.295>
- Jacobson, N. S., Follette, W. C., & Revenstorff, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, *15*(4), 336–352. [https://doi.org/10.1016/S0005-7894\(84\)80002-7](https://doi.org/10.1016/S0005-7894(84)80002-7)

- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19. <https://doi.org/10.1037/0022-006X.59.1.12>
- Jarrett, R. B., Eaves, G. G., Grannemann, B. D., & Rush, A. J. (1991). Clinical, cognitive, and demographic predictors of response to cognitive therapy for depression: A preliminary report. *Psychiatry Research, 37*(3), 245–260. [https://doi.org/10.1016/0165-1781\(91\)90061-S](https://doi.org/10.1016/0165-1781(91)90061-S)
- Jensen-Doss, A., & Hawley, K. M. (2010). Understanding barriers to evidence-based assessment: Clinician attitudes toward standardized assessment tools. *Journal of Clinical Child and Adolescent Psychology, 39*(6), 885–896. <https://doi.org/10.1080/15374416.2010.517169>
- Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to Baldwin and Imel's (2013) review. *Clinical Psychology Review, 67*, 78–93. <https://doi.org/10.1016/j.cpr.2018.08.004>
- Johnston, K. M., Powell, L. C., Anderson, I. M., Szabo, S., & Cline, S. (2019). The burden of treatment-resistant depression: A systematic review of the economic and quality of life literature. *Journal of Affective Disorders, 242*, 195–210. <https://doi.org/10.1016/J.JAD.2018.06.045>
- Jonsson, U., Alaie, I., Parling, T., & Arnberg, F. K. (2014). Reporting of harms in randomized controlled trials of psychological interventions for mental and behavioral disorders: A review of current practice. *Contemporary Clinical Trials, 38*(1), 1–8.
- Jonsson, U., Johanson, J., Nilsson, E., & Lindblad, F. (2016). Adverse effects of psychological therapy : An exploratory study of practitioners ' experiences from child and adolescent psychiatry. *Clinical Child Psychology and Psychiatry, 21*(3), 432–446. <https://doi.org/10.1177/1359104515614072>
- Jørgensen, C. R., Freund, C., Bøye, R., Jordet, H., Andersen, D., & Kjølbye, M. (2013). Outcome of mentalization-based and supportive psychotherapy in patients with borderline personality disorder: a randomized trial. *Acta Psychiatrica Scandinavica, 127*(4), 305–317. <https://doi.org/10.1111/j.1600-0447.2012.01923.x>
- Jørgensen, C. R., Bøye, R., Andersen, D., Døssing Blaabjerg, A. H., Freund, C., Jordet, H., &

- Kjølbbye, M. (2014). Eighteen months post-treatment naturalistic follow-up study of mentalization-based therapy and supportive group treatment of borderline personality disorder: Clinical outcomes and functioning. *Nordic Psychology*, *66*(4), 254–273. <https://doi.org/10.1080/19012276.2014.963649>
- Kaiser, T., Kolar, D., Kok, R. N., & Schuster, R. (2019). What does the LAC depression study really show?: Critical considerations of the methodology. *Psychotherapeut*, *65*(1), 22–26. <https://doi.org/10.1007/s00278-019-00392-y>
- Karatzias, T., Murphy, P., Cloitre, M., Bisson, J., Roberts, N., Shevlin, M., ... Hutton, P. (2019). Psychological interventions for ICD-11 complex PTSD symptoms: systematic review and meta-analysis. *Psychological Medicine*, *49*(11), 1761–1775. <https://doi.org/10.1017/S0033291719000436>
- Karlin, B. E., Brown, G. K., Trockel, M., Cunning, D., Zeiss, A. M., & Taylor, C. B. (2012). National dissemination of cognitive behavioral therapy for depression in the department of veterans affairs health care system: Therapist and patient-level outcomes. *Journal of Consulting and Clinical Psychology*, *80*(5), 707–718. <https://doi.org/10.1037/a0029328>
- Karver, M. S., De Nadai, A. S., Monahan, M., & Shirk, S. R. (2018). Meta-Analysis of the Prospective Relation between Alliance and Outcome in Child and Adolescent Psychotherapy. *Psychotherapy*, *55*(4), 341–355. <https://doi.org/10.1037/pst0000176>
- Katz, N., Paillard, F. C., & Van Inwegen, R. (2015). A review of the use of the number needed to treat to evaluate the efficacy of analgesics. *Journal of Pain*, *16*, 116–123. <https://doi.org/10.1016/j.jpain.2014.08.005>
- Kautzky, A., Dold, M., Bartova, L., Spies, M., Kranz, G. S., Souery, D., ... Kasper, S. (2019). Clinical factors predicting treatment resistant depression: affirmative results from the European multicenter study. *Acta Psychiatrica Scandinavica*, *139*(1), 78–88. <https://doi.org/10.1111/acps.12959>
- Kazdin, A. E. (1994). Methodology, design, and evaluation in psychotherapy research. In A. E. Bergin & S. L. Garfield, *Handbook of psychotherapy and behavior change*, 4th ed. (pp. 19–71). Oxford, England: John Wiley & Sons.
- Kazdin, A. E. (2007). Mediators and Mechanisms of Change in Psychotherapy Research. *Annual*



*Review of Clinical Psychology*, 3(1), 1–27.  
<https://doi.org/10.1146/annurev.clinpsy.3.022806.091432>

Kazdin, A. E. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research*, 19(4–5), 418–428.  
<https://doi.org/10.1080/10503300802448899>

Keefe, J. R., Kim, T. T., DeRubeis, R. J., Streiner, D. L., Links, P. S., & McMain, S. F. (2020). Treatment selection in borderline personality disorder between dialectical behavior therapy and psychodynamic psychiatric management. *Psychological Medicine*, 1–9.  
<https://doi.org/10.1017/S0033291720000550>

Keeley, M. L., Storch, E. A., Merlo, L. J., & Geffken, G. R. (2008). Clinical predictors of response to cognitive-behavioral therapy for obsessive-compulsive disorder. *Clinical Psychology Review*, 28(1), 118–130. <https://doi.org/10.1016/j.cpr.2007.04.003>

Kegel, A. F., & Flückiger, C. (2015). Predicting Psychotherapy Dropouts: A Multilevel Approach. *Clinical Psychology & Psychotherapy*, 22(5), 377–386. <https://doi.org/10.1002/cpp.1899>

Kellogg, S. H., & Young, J. E. (2006). Schema therapy for borderline personality disorder. *Journal of Clinical Psychology*, 62, 445–458. <https://doi.org/10.1002/jclp.20240>

Kendall, P. C., Marrs-Garcia, A., Nath, S. R., & Sheldrick, R. C. (1999). Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology*, 67, 285–299. <https://doi.org/10.1037/0022-006X.67.3.285>

Kent, D. M., Steyerberg, E., & Van Klaveren, D. (2018). Personalized evidence based medicine: Predictive approaches to heterogeneous treatment effects. *BMJ (Online)*, 363, p. k4245. <https://doi.org/10.1136/bmj.k4245>

Kertz, S. J., Koran, J., Stevens, K. T., & Björgvinsson, T. (2015). Repetitive negative thinking predicts depression and anxiety symptom improvement during brief cognitive behavioral therapy. *Behaviour Research and Therapy*, 68, 54–63.  
<https://doi.org/10.1016/j.brat.2015.03.006>

Kilbourne, A. M., Williams, M., Bauer, M. S., & Arean, P. (2012). Implementation research: Reducing the research-to-practice gap in depression treatment. *Depression Research and*

*Treatment*, Vol. 2012. 2012: 476027. <https://doi.org/10.1155/2012/476027>

Klatte, R., Strauss, B., Flückiger, C., & Rosendahl, J. (2018). Adverse effects of psychotherapy: protocol for a systematic review and meta-analysis. *Systematic Reviews*, 7(1), 135. <https://doi.org/10.1186/s13643-018-0802-x>

Klein, J. P., Willenborg, B., & Schweiger, U. (2015). Diagnostik affektiver Störungen im DSM-5: Neuerungen und Kontroversen. *Fortschritte Der Neurologie Psychiatrie*, 82(11), 655–664. <https://doi.org/10.1055/s-0034-1385377>

Kliem, S., & Kröger, C. (2013). Prevention of chronic PTSD with early cognitive behavioral therapy. A meta-analysis using mixed-effects modeling. *Behaviour Research and Therapy*, 51(11), 753–761. <https://doi.org/10.1016/j.brat.2013.08.005>

Kliem, S., Kröger, C., & Kosfelder, J. (2010). Dialectical behavior therapy for borderline personality disorder: A meta-analysis using mixed-effects modeling. *Journal of Consulting and Clinical Psychology*, 78(6), 936–951. <https://doi.org/10.1037/a0021015>

Knapstad, M., Nordgreen, T., & Smith, O. R. F. (2018). Prompt mental health care, the Norwegian version of IAPT: clinical outcomes and predictors of change in a multicenter cohort study. *BMC Psychiatry*, 18(1), 260. <https://doi.org/10.1186/s12888-018-1838-0>

Knopp, J., Knowles, S., Bee, P., Lovell, K., & Bower, P. (2013). A systematic review of predictors and moderators of response to psychological therapies in OCD: Do we have enough empirical evidence to target treatment? *Clinical Psychology Review*, 33(8), 1067–1081. <https://doi.org/10.1016/j.cpr.2013.08.008>

Kohls, S., Bents, H., & Pietrowsky, R. (2002). Prädiktoren für Erfolg und Misserfolg in der verhaltenstherapeutischen Behandlung von Zwangspatienten. *Verhaltenstherapie*, 12(2), 98–106. <https://doi.org/10.1159/000064373>

Kolden, G. G., Klein, M. H., Wang, C. C., & Austin, S. B. (2011). Congruence/Genuineness. *Psychotherapy*, 48(1), 65–71. <https://doi.org/10.1037/a0022064>

Kolden, G. G., Wang, C. C., Austin, S. B., Chang, Y., & Klein, M. H. (2018). Congruence/Genuineness: A Meta-Analysis. *Psychotherapy*, 55(4), 424–433. <https://doi.org/10.1037/pst0000162>

- Kordy, H., Hannover, W., & Richard, M. (2001). Computer-assisted feedback-driven quality management for psychotherapy: The Stuttgart-Heidelberg model. *Journal of Consulting and Clinical Psychology, 69*(2), 173–183. <https://doi.org/10.1037/0022-006X.69.2.173>
- Kraemer, H. C., & Kupfer, D. J. (2006). Size of Treatment Effects and Their Importance to Clinical Research and Practice. *Biological Psychiatry, 59*, 990–996. <https://doi.org/10.1016/j.biopsych.2005.09.014>
- Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and Moderators of Treatment Effects in Randomized Clinical Trials. *Archives of General Psychiatry, 59*(10), 877. <https://doi.org/10.1001/archpsyc.59.10.877>
- Krampen, G., Schui, G., & Hütter, J. W. (2008). Evidenzbasierte Psychotherapie und Therapie-ressourcen. Ein Erweitertes 4-phasen-prüfmodell und Seine Anwendung auf die Klinisch-psychologische Fachliteratur aus dem Deutschsprachigen Bereich. *Zeitschrift Fur Klinische Psychologie Und Psychotherapie, 37*(1), 43–51. <https://doi.org/10.1026/1616-3443.37.1.43>
- Kratzer, L., Heinz, P., Schennach, R., Schiepek, G. K., Padberg, F., & Jobst, A. (2018). Inpatient Treatment of Complex PTSD Following Childhood Abuse: Effectiveness and Predictors of Treatment Outcome. *PPmP Psychotherapie Psychosomatik Medizinische Psychologie, 69* (3-04), 114-122. <https://doi.org/10.1055/a-0591-3962>
- Kraus, D. R., Bentley, J. H., Alexander, P. C., Boswell, J. F., Constantino, M. J., Baxter, E. E., & Castonguay, L. G. (2016). Predicting therapist effectiveness from their own practice-based evidence. *Journal of Consulting and Clinical Psychology, 84*(6), 473–483. <https://doi.org/10.1037/ccp0000083>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine, 16*(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Kröger, C., Harbeck, S., Armbrust, M., & Kliem, S. (2013a). Effectiveness, response, and dropout of dialectical behavior therapy for borderline personality disorder in an inpatient setting. *Behaviour Research and Therapy, 51*(8), 411–416. <https://doi.org/10.1016/j.brat.2013.04.008>
- Kröger, C., Schweiger, U., Sipos, V., Arnold, R., Kahl, K. G., Schunert, T., ... Reinecker, H. (2006).

Effectiveness of dialectical behaviour therapy for borderline personality disorder in an inpatient setting. *Behaviour Research and Therapy*, 44(8), 1211–1217. <https://doi.org/10.1016/j.brat.2005.08.012>

Kube, T., Berg, M., Kleim, B., & Herzog, P. (2020). Rethinking post-traumatic stress disorder – A predictive processing perspective. *Neuroscience & Biobehavioral Reviews*, 113(April), 448–460. <https://doi.org/10.1016/j.neubiorev.2020.04.014>

Kube, T., Herzog, P., Michalak, C. M., Glombiewski, J. A., Doering, B. K., & Rief, W. (2019). Further Specifying the Cognitive Model of Depression: Situational Expectations and Global Cognitions as Predictors of Depressive Symptoms. *Clinical Psychology in Europe*, 1(4). <https://doi.org/10.32872/cpe.v1i4.33548>

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2019). Distorted cognitive processes in major depression – A predictive processing perspective. *Biological Psychiatry*, 87(5), 388–398. <https://doi.org/10.1016/j.biopsych.2019.07.017>

Kube, T., Siebers, V. H. A., Herzog, P., Glombiewski, J. A., Doering, B. K., & Rief, W. (2018). Integrating situation-specific dysfunctional expectations and dispositional optimism into the cognitive model of depression – A path-analytic approach. *Journal of Affective Disorders*, 229, 199–205. <https://doi.org/10.1016/j.jad.2017.12.082>

Külz, A. K., Hassenpflug, K., Riemann, D., Linster, H. W., Dornberg, M., & Voderholzer, U. (2010). Ambulante psychotherapeutische Versorgung bei Zwangserkrankungen: Ergebnisse einer anonymen Therapeutenbefragung. *PPmP Psychotherapie Psychosomatik Medizinische Psychologie*, 60(6), 194–201. <https://doi.org/10.1055/s-0029-1202837>

Kyrios, M., Hordern, C., & Fassnacht, D. B. (2015). Predictors of response to cognitive behaviour therapy for obsessive-compulsive disorder. *International Journal of Clinical and Health Psychology*, 15(3), 181–190. <https://doi.org/10.1016/j.ijchp.2015.07.003>

Ladwig, I., Rief, W., & Nestoriuc, Y. (2014). Welche Risiken und Nebenwirkungen hat Psychotherapie? – Entwicklung des Inventars zur Erfassung Negativer Effekte von Psychotherapie (INEP). *Verhaltenstherapie*, 24, 252–263.

Lambert, M. J. (2013). *Bergin and Garfield's handbook of psychotherapy and behavior change*. Hoboken, NJ: John Wiley & Sons, Inc.

- Lambert, M. J. (2017). Maximizing Psychotherapy Outcome beyond Evidence-Based Medicine. *Psychotherapy and Psychosomatics*, *86*(2), 80–89. <https://doi.org/10.1159/000455170>
- Lambert, M. J., & Bailey, R. J. (2012). Measures of clinically significant change. In *APA Handbooks in Psychology®. APA handbook of research methods in psychology, Vol 3: Data analysis and research publication*. (pp. 147–160). <https://doi.org/10.1037/13621-007>
- Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, *69*(2), 159–172. <https://doi.org/10.1037/0022-006X.69.2.159>
- Lambert, M. J., & Harmon, K. L. (2018). The merits of implementing routine outcome monitoring in clinical practice. *Clinical Psychology: Science and Practice*, *25*(4), e12268. <https://doi.org/10.1111/cpsp.12268>
- Lambert, M. J., Hatch, D. R., Kingston, M. D., & Edwards, B. C. (1986). Zung, Beck, and Hamilton Rating Scales as measures of treatment outcome: A meta-analytic comparison. *Journal of Consulting and Clinical Psychology*, *54*(1), 54–59. <https://doi.org/10.1037//0022-006x.54.1.54>
- Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, *48*(1), 72–79. <https://doi.org/10.1037/a0022238>
- Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*, *55*(4), 520–537. <https://doi.org/10.1037/pst0000167>
- Laska, K. M., Smith, T. L., Wislocki, A. P., Minami, T., & Wampold, B. E. (2013). Uniformity of evidence-based treatments in practice? Therapist effects in the delivery of cognitive processing therapy for PTSD. *Journal of Counseling Psychology*, *60*(1), 31–41. <https://doi.org/10.1037/a0031294>
- Lee, Y., Raguett, R. M., Mansur, R. B., Boutilier, J. J., Rosenblat, J. D., Trevizol, A., ... McIntyre, R. S. (2018). Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review. *Journal of Affective Disorders*, *241*, 519–532. <https://doi.org/10.1016/j.jad.2018.08.073>

- Leichsenring, F., Abbass, A., Beutel, M., Gündel, H., Heuft, G., Hoffmann, S. O., ... Steinert, C. (2019). Vom Sinn des Verfahrenskonzepts und der Verfahrensvielfalt – und warum das Baukasten-System in der Psychotherapie nicht funktioniert. *Zeitschrift Für Psychosomatische Medizin Und Psychotherapie*, 65(4), 321–340. <https://doi.org/10.13109/zptm.2019.65.4.0a1>
- Leichsenring, F., Abbass, A., Driessen, E., Hilsenroth, M., Luyten, P., Rabung, S., & Steinert, C. (2018). Equivalence and non-inferiority testing in psychotherapy research. *Psychological Medicine*, 48(11), 1917-1919. <https://doi.org/10.1017/S0033291718001289>
- Leichsenring, F., Abbass, A., Hilsenroth, M. J., Luyten, P., Munder, T., Rabung, S., & Steinert, C. (2018). “Gold standards,” plurality and monocultures: The need for diversity in psychotherapy. *Frontiers in Psychiatry*, 9(APR). <https://doi.org/10.3389/fpsyt.2018.00159>
- Lemmens, L. H. J. M., Arntz, A., Peeters, F., Hollon, S. D., Roefs, A., & Huibers, M. J. H. (2015). Clinical effectiveness of cognitive therapy v. interpersonal psychotherapy for depression: results of a randomized controlled trial. *Psychological Medicine*, 45(10), 2095–2110. <https://doi.org/10.1017/S0033291715000033>
- Lemmens, L. H.J.M., van Bronswijk, S. C., Peeters, F., Arntz, A., Hollon, S. D., & Huibers, M. J. H. (2019). Long-term outcomes of acute treatment with cognitive therapy v. interpersonal psychotherapy for adult depression: follow-up of a randomized controlled trial. *Psychological Medicine*, 49(3), 465–473. <https://doi.org/10.1017/S0033291718001083>
- Leuzinger-Bohleber, M., Hautzinger, M., Fiedler, G., Keller, W., Bahrke, U., Kallenbach, L., ... Beutel, M. (2019). Outcome of Psychoanalytic and Cognitive-Behavioural Long-Term Therapy with Chronically Depressed Patients: A Controlled Trial with Preferential and Randomized Allocation. *The Canadian Journal of Psychiatry*, 64(1), 47–58. <https://doi.org/10.1177/0706743718780340>
- Leuzinger-Bohleber, M., Hautzinger, M., Keller, W., Fiedler, G., Bahrke, U., Kallenbach, L., ... Beutel, M. (2019). Psychoanalytische und kognitiv-behaviorale Langzeitbehandlung chronisch depressiver Patienten bei randomisierter oder präferierter Zuweisung. *PSYCHE*, 73(02), 77–105. <https://doi.org/10.21706/ps-73-2-77>
- Lex, H., Ginsburg, Y., Sitzmann, A. F., Grayhack, C., Maixner, D. F., & Mickey, B. J. (2019). Quality

of life across domains among individuals with treatment-resistant depression. *Journal of Affective Disorders*, 243, 401–407. <https://doi.org/10.1016/j.jad.2018.09.062>

Lieberei, B., & Linden, M. (2008). Unerwünschte effekte, Nebenwirkungen und Behandlungsfehler in der Psychotherapie. *Zeitschrift Fur Evidenz, Fortbildung Und Qualitat Im Gesundheitswesen*, 102(9), 558–562. <https://doi.org/10.1016/j.zefq.2008.09.017>

Lilienfeld, S. O. (2007). Psychological Treatments That Cause Harm. *Perspectives on Psychological Science*, 2(1), 53–70. <https://doi.org/10.1111/j.1745-6916.2007.00029.x>

Lilienfeld, S. O. (2013). Closing the Science-Practice Gap. *APS Observer*, 26(3).

Lincoln, T. M., Rief, W., Hahlweg, K., Frank, M., Von Witzleben, I., Schroeder, B., & Fiegenbaum, W. (2003). Effectiveness of an empirically supported treatment for social phobia in the field. *Behaviour Research and Therapy*, 41(11), 1251–1269. [https://doi.org/10.1016/S0005-7967\(03\)00038-X](https://doi.org/10.1016/S0005-7967(03)00038-X)

Linden, M. (2013). How to define, find and classify side effects in psychotherapy: From unwanted events to adverse treatment reactions. *Clinical Psychology and Psychotherapy*, 20(4), 286–296.

Linden, M. (2017). Definition and Assessment of Disability in Mental Disorders under the Perspective of the International Classification of Functioning Disability and Health (ICF). *Behavioral Sciences and the Law*, 35(2), 124–134. <https://doi.org/10.1002/bsl.2283>

Linden, M., Strauß, B., Scholten, S., Nestoriuc, Y., Brakemeier, E.-L., & Wasilewski, J. (2018). Definition and Decision-Making in the Determination and Detection of Side Effects of Psychotherapy. *PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie*, 68(09/10), 377–382.

Linehan, M. M. (1993). *Cognitive behavioral therapy of borderline personality disorder*. New York: Guilford Press.

Linehan, M. M., Armstrong, H. E., Suarez, A., Allmon, D., & Heard, H. L. (1991). Cognitive-Behavioral Treatment of Chronically Parasuicidal Borderline Patients. *Archives of General Psychiatry*, 48(12), 1060–1064. <https://doi.org/10.1001/archpsyc.1991.01810360024003>

Linehan, M. M., Korslund, K. E., Harned, M. S., Gallop, R. J., Lungu, A., Neacsiu, A. D., ... Murray-

- Gregory, A. M. (2015). Dialectical behavior therapy for high suicide risk in individuals with borderline personality disorder: A randomized clinical trial and component analysis. *JAMA Psychiatry*, *72*(5), 475–482. <https://doi.org/10.1001/jamapsychiatry.2014.3039>
- Lonergan, M. (2014). Cognitive Behavioral Therapy for PTSD: The Role of Complex PTSD on Treatment Outcome. *Journal of Aggression, Maltreatment & Trauma*, *23*(5), 494–512. <https://doi.org/10.1080/10926771.2014.904467>
- Lorenzo-Luaces, L., German, R. E., & DeRubeis, R. J. (2015). It's complicated: The relation between cognitive change procedures, cognitive change, and symptom change in cognitive therapy for depression. *Clinical Psychology Review*, *41*, 3–15. <https://doi.org/10.1016/j.cpr.2014.12.003>
- Lorenzo-Luaces, L., Keefe, J. R., & DeRubeis, R. J. (2016). Cognitive-Behavioral Therapy: Nature and Relation to Non-Cognitive Behavioral Therapy. *Behavior Therapy*, *47*, 785–803. <https://doi.org/10.1016/j.beth.2016.02.012>
- Lorimer, B., Delgado, J., Kellett, S., & Brown, G. (2019). Exploring relapse through a network analysis of residual depression and anxiety symptoms after cognitive behavioural therapy: A proof-of-concept study. *Psychotherapy Research*. <https://doi.org/10.1080/10503307.2019.1650980>
- Lueger, R. J. (2002). Practice-informed research and research-informed psychotherapy. *Journal of Clinical Psychology*, *58*(10), 1265–1276. <https://doi.org/10.1002/jclp.10110>
- Luijken, K., Groenwold, R. H. H., Van Calster, B., Steyerberg, E. W., & van Smeden, M. (2019). Impact of predictor measurement heterogeneity across settings on the performance of prediction models: A measurement error perspective. *Statistics in Medicine*, *38*(18), 3444–3459. <https://doi.org/10.1002/sim.8183>
- Lutz, W. (2010). *Lehrbuch Psychotherapie*. Bern, CH: Hans Huber.
- Lutz, W., De Jong, K., & Rubel, J. (2015). Patient-focused and feedback research in psychotherapy: Where are we and where do we want to go? *Psychotherapy Research*, *25*(6), 625–632. <https://doi.org/10.1080/10503307.2015.1079661>
- Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M.-A., Jorasz, C., ... Tschitsaz-Stucki, A.



(2013). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research*, 23(1), 14–24. <https://doi.org/10.1080/10503307.2012.693837>

Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., ... Iveson, S. (2005). Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. *Journal of Consulting and Clinical Psychology*, 73(5), 904–913. <https://doi.org/10.1037/0022-006X.73.5.904>

Lutz, W., Martinovich, Z., Howard, K. I., & Leon, S. C. (2002). Outcomes management, expected treatment response, and severity-adjusted provider profiling in outpatient psychotherapy. *Journal of Clinical Psychology*, 58(10), 1291–1304. <https://doi.org/10.1002/jclp.10070>

Lutz, W., Neu, R., & Rubel, J. (2019). *Evaluation und Effekterfassung in der Psychotherapie. Standards der Psychotherapie*. Göttingen: Hogrefe.

Lutz, W., Rubel, J. A., Schwartz, B., Schilling, V., & Deisenhofer, A. K. (2019). Towards integrating personalized feedback research into clinical practice: Development of the Trier Treatment Navigator (TTN). *Behaviour Research and Therapy*, 120, 103438. <https://doi.org/10.1016/j.brat.2019.103438>

Lutz, W., Schiefele, A. K., Wucherpfennig, F., Rubel, J., & Stulz, N. (2016). Clinical effectiveness of cognitive behavioral therapy for depression in routine care: A propensity score based comparison between randomized controlled trials and clinical practice. *Journal of Affective Disorders*, 189, 150–158. <https://doi.org/10.1016/j.jad.2015.08.072>

Lutz, W., Schwartz, B., Hofmann, S. G., Fisher, A. J., Husen, K., & Rubel, J. A. (2018). Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-25953-0>

Maercker, A., & Schützwohl, M. (1998). Erfassung von psychischen Belastungsfolgen: Die Impact of Event Skala-revidierte Version. *Diagnostica*, 44(3), 130–141. <https://doi.org/10.1037/t01210-000>

Maercker, A., Brewin, C. R., Bryant, R. A., Cloitre, M., van Ommeren, M., Jones, L. M., ... Reed, G. M. (2013). Diagnosis and classification of disorders specifically associated with stress:

proposals for ICD-11. *World Psychiatry*, 12(3), 198–206.  
<https://doi.org/10.1002/wps.20057>

Maher, M. J., Wang, Y., Zuckoff, A., Wall, M. M., Franklin, M., Foa, E. B., & Simpson, H. B. (2012). Predictors of patient adherence to cognitive-behavioral therapy for obsessive-compulsive disorder. *Psychotherapy and Psychosomatics*, 81(2), 124–126.  
<https://doi.org/10.1159/000330214>

Marcus, D. K., O'Connell, D., Norris, A. L., & Sawaqdeh, A. (2014). Is the Dodo bird endangered in the 21st century? A meta-analysis of treatment comparison studies. *Clinical Psychology Review*, 34, 519–530. <https://doi.org/10.1016/j.cpr.2014.08.001>

May, P. R. (1971). For better or for worse? Psychotherapy and variance change: a critical review of the literature. *The Journal of Nervous and Mental Disease*, 152(3), 184–192.

Mayo-Wilson, E., Dias, S., Mavranouzouli, I., Kew, K., Clark, D. M., Ades, A. E., & Pilling, S. (2014). Psychological and pharmacological interventions for social anxiety disorder in adults: A systematic review and network meta-analysis. *The Lancet Psychiatry*, 1(5), 368–376.  
[https://doi.org/10.1016/S2215-0366\(14\)70329-3](https://doi.org/10.1016/S2215-0366(14)70329-3)

McCauley, E., Berk, M. S., Asarnow, J. R., Adrian, M., Cohen, J., Korslund, K., ... Linehan, M. M. (2018). Efficacy of Dialectical Behavior Therapy for Adolescents at High Risk for Suicide. *JAMA Psychiatry*, 75(8), 777. <https://doi.org/10.1001/jamapsychiatry.2018.1109>

McCullough, J. P. (2003). *Treatment for chronic depression: cognitive behavioral analysis system of psychotherapy (CBASP)*. New York, NY: Guilford Press.

McHugh, R. K., & Barlow, D. H. (2010). The Dissemination and Implementation of Evidence-Based Psychological Treatments: A Review of Current Efforts. *American Psychologist*, 65(2), 73–84. <https://doi.org/10.1037/a0018121>

McKay, D., Sookman, D., Neziroglu, F., Wilhelm, S., Stein, D. J., Kyrios, M., ... Veale, D. (2015). Efficacy of cognitive-behavioral therapy for obsessive-compulsive disorder. *Psychiatry Research*, 227(1), 104–113. <https://doi.org/10.1016/j.psychres.2014.11.058>

McMurrin, M., Huband, N., & Overton, E. (2010, April). Non-completion of personality disorder treatments: A systematic review of correlates, consequences, and interventions. *Clinical*

*Psychology Review*, Vol. 30, pp. 277–287. <https://doi.org/10.1016/j.cpr.2009.12.002>

McMyler, C., & Prymachuk, S. (2008). Do “no-suicide” contracts work? *Journal of Psychiatric and Mental Health Nursing*, 15(6), 512–522. <https://doi.org/10.1111/j.1365-2850.2008.01286.x>

Mehlum, L., Ramberg, M., Tørmoen, A. J., Haga, E., Diep, L. M., Stanley, B. H., ... Grøholt, B. (2016). Dialectical Behavior Therapy Compared With Enhanced Usual Care for Adolescents With Repeated Suicidal and Self-Harming Behavior: Outcomes Over a One-Year Follow-Up. *Journal of the American Academy of Child and Adolescent Psychiatry*, 55(4), 295–300. <https://doi.org/10.1016/j.jaac.2016.01.005>

Meinschmidt, G., Tegethoff, M., Belardi, A., Stalujanis, E., Oh, M., Jung, E. K., ... Lee, J.-H. (2019). Personalized prediction of smartphone-based psychotherapeutic micro-intervention success using machine learning. *Journal of Affective Disorders*, 264, 430–437. <https://doi.org/10.1016/j.jad.2019.11.071>

Meister, R., Lanio, J., Fangmeier, T., Härter, M., Schramm, E., Zobel, I., ... Kriston, L. (2019). Adverse events during a disorder-specific psychotherapy compared to a nonspecific psychotherapy in patients with chronic depression. *Journal of Clinical Psychology*, jclp.22869. <https://doi.org/10.1002/jclp.22869>

Meister, R., von Wolff, A., Mohr, H., Nestoriuc, Y., Härter, M., Hölzel, L., & Kriston, L. (2016). Adverse event methods were heterogeneous and insufficiently reported in randomized trials on persistent depressive disorder. *Journal of Clinical Epidemiology*, 71, 97–108. <https://doi.org/10.1016/j.jclinepi.2015.10.007>

Merrill, K. A., Tolbert, V. E., & Wade, W. A. (2003). Effectiveness of cognitive therapy for depression in a community mental health center: a benchmarking study. *Journal of Consulting and Clinical Psychology*, 71(2), 404–409. <https://doi.org/10.1037/0022-006x.71.2.404>

Metzler, P., & Krause, B. (1997). Methodischer Standard bei Studien zur Therapieevaluation. *Methods of Psychological Research Online*, 2(1), 55–67.

Minami, T., Serlin, R. C., Wampold, B. E., Kircher, J. C., & Brown, G. S. (2008). Using clinical trials to benchmark effects produced in clinical practice. *Quality and Quantity*, 42(4), 513–525.

<https://doi.org/10.1007/s11135-006-9057-z>

Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E. G., Brown, G. S., & Kircher, J. C. (2008). Benchmarking the Effectiveness of Psychotherapy Treatment for Adult Depression in a Managed Care Environment: A Preliminary Study. *Journal of Consulting and Clinical Psychology, 76*(1), 116–124. <https://doi.org/10.1037/0022-006X.76.1.116>

Minami, T., Wampold, B. E., Serlin, R. C., Kircher, J. C., & Brown, G. S. (2007). Benchmarks for Psychotherapy Efficacy in Adult Major Depression. *Journal of Consulting and Clinical Psychology, 75*(2), 232–243. <https://doi.org/10.1037/0022-006X.75.2.232>

Mindlis, I., Morales-Raveendran, E., Goodman, E., Xu, K., Vila-Castelar, C., Keller, K., ... Wisnivesky, J. P. (2017). Post-traumatic stress disorder dimensions and asthma morbidity in World Trade Center rescue and recovery workers. *Journal of Asthma, 54*(7), 723–731. <https://doi.org/10.1080/02770903.2016.1263650>

Mintz, J., Drake, R. E., & Crits-Christoph, P. (1996). Efficacy and effectiveness of psychotherapy: Two paradigms, one science. *American Psychologist, 51*(10), 1084–1085. <https://doi.org/10.1037/0003-066x.51.10.1084>

Mohr, D. C. (1995). Negative Outcome in Psychotherapy: A Critical Review. *Clinical Psychology: Science and Practice, 2*(1), 1–27. <https://doi.org/10.1111/j.1468-2850.1995.tb00022.x>

Moons, K. G. M., Altman, D. G., Reitsma, J. B., Ioannidis, J. P. A., Macaskill, P., Steyerberg, E. W., ... Collins, G. S. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine, 162*(1), W1–W73. <https://doi.org/10.7326/M14-0698>

Moore, A. (2015). Number needed to treat - Just one of the cards in the pack. *Journal of Pain, 16*(2), 124–125. <https://doi.org/10.1016/j.jpain.2014.11.006>

Moritz, S., Fieker, M., Hottenrott, B., Seeralan, T., Cludius, B., Kolbeck, K., ... Nestoriuc, Y. (2015). No pain, no gain? Adverse effects of psychotherapy in obsessive-compulsive disorder and its relationship to treatment gains. *Journal of Obsessive-Compulsive and Related Disorders, 5*, 61–66.

Moritz, S., Külz, A., Voderholzer, U., Hillebrand, T., McKay, D., & Jelinek, L. (2019). “Phobie à

deux” and other reasons why clinicians do not apply exposure with response prevention in patients with obsessive–compulsive disorder. *Cognitive Behaviour Therapy*, 48(2), 162–176. <https://doi.org/10.1080/16506073.2018.1494750>

Moritz, S., Nestoriuc, Y., Rief, W., Klein, J. P., Jelinek, L., & Peth, J. (2018). It can't hurt, right? Adverse effects of psychotherapy in patients with depression. *European Archives of Psychiatry and Clinical Neuroscience*, 269, 577-586. <https://doi.org/10.1007/s00406-018-0931-1>

Moses, E. B., & Barlow, D. H. (2006). A New Unified Treatment Approach for Emotional Disorders Based on Emotion Science. *Current Directions in Psychological Science*, 15(3), 146–150. <https://doi.org/10.1111/j.0963-7214.2006.00425.x>

Moutoussis, M., Shahar, N., Hauser, T. U., & Dolan, R. J. (2018). Computation in Psychotherapy, or How Computational Psychiatry Can Aid Learning-Based Psychological Therapies. *Computational Psychiatry (Cambridge, Mass.)*, 2, 50–73. [https://doi.org/10.1162/CPSY\\_a\\_00014](https://doi.org/10.1162/CPSY_a_00014)

Mulder, R., Murray, G., & Rucklidge, J. (2017). Common versus specific factors in psychotherapy: opening the black box. *The Lancet Psychiatry*, 4, 953–962. [https://doi.org/10.1016/S2215-0366\(17\)30100-1](https://doi.org/10.1016/S2215-0366(17)30100-1)

Mullin, T., Barkham, M., Mothersole, G., Bewick, B. M., & Kinder, A. (2006). Recovery and improvement benchmarks for counselling and the psychological therapies in routine primary care. *Counselling and Psychotherapy Research*, 6(1), 68–80. <https://doi.org/10.1080/14733140600581515>

Mumma, G. H., & Fluck, J. (2016). How valid is your case formulation? Empirically testing your cognitive behavioural case formulation for tailored treatment. *The Cognitive Behaviour Therapist*, 9, e12. <https://doi.org/10.1017/S1754470X16000088>

Munder, T., Flückiger, C., Leichsenring, F., Abbass, A. A., Hilsenroth, M. J., Luyten, P., ... Wampold, B. E. (2019a). Is psychotherapy effective? A re-analysis of treatments for depression. *Epidemiology and Psychiatric Sciences*, 28(3), 268–274. <https://doi.org/10.1017/S2045796018000355>

Munder, T., Flückiger, C., Leichsenring, F., Abbass, A. A., Hilsenroth, M. J., Luyten, P., ...

Wampold, B. E. (2019b). Is psychotherapy effective? A re-analysis of treatments for depression. *Epidemiology and Psychiatric Sciences*, 28(3), 268–274. <https://doi.org/10.1017/S2045796018000355>

Munder, T., Flückiger, C., Leichsenring, F., Abbass, A. A., Hilsenroth, M. J., Luyten, P., ... Wampold, B. E. (2019). Let's distinguish relative and absolute efficacy to move psychotherapy research forward. *Zeitschrift Fur Psychosomatische Medizin Und Psychotherapie*, 65(2), 178–182. <https://doi.org/10.13109/zptm.2019.65.2.178>

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press.

Mynors-Wallis, L., & Gath, D. (1997). Predictors of treatment outcome for major depression in primary care. *Psychological Medicine*, 27(3), S0033291796004126. <https://doi.org/10.1017/S0033291796004126>

Najavits, L. M. (2015). The problem of dropout from “gold standard” PTSD therapies. *F1000Prime Reports*, 7(April), 1–8. <https://doi.org/10.12703/P7-43>

Norcross, J. C., & Lambert, M. J. (2018). Psychotherapy Relationships That Work III. *Psychotherapy*, 55(4), 303–315. <https://doi.org/10.1037/pst0000193>

Norcross, J. C., & Wampold, B. E. (2011). Evidence-Based Therapy Relationships: Research Conclusions and Clinical Practices. *Psychotherapy*, 48(1), 98–102. <https://doi.org/10.1037/a0022161>

Normann, N., & Morina, N. (2018). The efficacy of metacognitive therapy: A systematic review and meta-analysis. *Frontiers in Psychology*, 9(NOV), 1–14. <https://doi.org/10.3389/fpsyg.2018.02211>

Ong, C. W., Lee, E. B., & Twohig, M. P. (2018). A meta-analysis of dropout rates in acceptance and commitment therapy. *Behaviour Research and Therapy*, 104, 14–33. <https://doi.org/10.1016/j.brat.2018.02.004>

Öst, L. G., Havnen, A., Hansen, B., & Kvale, G. (2015). Cognitive behavioral treatments of obsessive-compulsive disorder. A systematic review and meta-analysis of studies published 1993-2014. *Clinical Psychology Review*, Vol. 40, pp. 156–169. <https://doi.org/10.1016/j.cpr.2015.06.003>

- Ougrin, D., Tranah, T., Stahl, D., Moran, P., & Asarnow, J. R. (2015). Therapeutic Interventions for Suicide Attempts and Self-Harm in Adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54(2), 97-107.e2. <https://doi.org/10.1016/j.jaac.2014.10.009>
- Palardy, V., El-Baalbaki, G., Fredette, C., Rizkallah, E., & Guay, S. (2018). Social Support and Symptom Severity Among Patients With Obsessive-Compulsive Disorder or Panic Disorder With Agoraphobia: A Systematic Review. *Europe's Journal of Psychology*, 14(1), 254–286. <https://doi.org/10.5964/ejop.v14i1.1252>
- Panos, P. T., Jackson, J. W., Hasan, O., & Panos, A. (2014). Meta-Analysis and Systematic Review Assessing the Efficacy of Dialectical Behavior Therapy (DBT). *Research on Social Work Practice*, 24(2), 213–223. <https://doi.org/10.1177/1049731513503047>
- Parry, G. D., Crawford, M. J., & Duggan, C. (2016). Iatrogenic harm from psychological therapies—time to move on. *British Journal of Psychiatry*, 208(3), 210–212.
- Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.
- Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. New York, NY: Basic Books.
- Peluso, P. R., & Freund, R. R. (2018). Therapist and Client Emotional Expression and Psychotherapy Outcomes: A Meta-Analysis. *Psychotherapy*, 55(4), 461–472. <https://doi.org/10.1037/pst0000165>
- Perlis, R. H. (2013). A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. *Biological Psychiatry*, 74(1), 7–14. <https://doi.org/10.1016/j.biopsych.2012.12.007>
- Perroud, N., Uher, R., Dieben, K., Nicastro, R., & Huguelet, P. (2010). Predictors of response and drop-out during intensive dialectical behavior therapy. *Journal of Personality Disorders*, 24(5), 634–650. <https://doi.org/10.1521/pedi.2010.24.5.634>
- Persons, J. B., Bostrom, A., & Bertagnolli, A. (1999). Results of Randomized Controlled Trials of Cognitive Therapy for Depression Generalize to Private Practice. *Cognitive Therapy and*

*Research*, 23(5), 535–548. <https://doi.org/10.1023/A:1018724505659>

Pittler, M. H., Blümle, A., Meerpohl, J. J., & Antes, G. (2011). CONSORT 2010: Aktualisierte Leitlinie für Berichte randomisierter Studien im Parallelgruppen-Design. *Deutsche Medizinische Wochenschrift*, 136(8), 20–23. <https://doi.org/10.1055/s-0031-1272980>

Pocock, S. J., Assmann, S. E., Enos, L. E., & Kasten, L. E. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: Current practice and problems. *Statistics in Medicine*, 21(19), 2917–2930. <https://doi.org/10.1002/sim.1296>

Posternak, M. A., & Miller, I. (2001). Untreated short-term course of major depression: a meta-analysis of outcomes from studies using wait-list control groups. *Journal of Affective Disorders*, 66(2–3), 139–146. [https://doi.org/10.1016/s0165-0327\(00\)00304-9](https://doi.org/10.1016/s0165-0327(00)00304-9)

Powers, M. B., Smits, J. A. J., Whitley, D., Bystritsky, A., & Telch, M. J. (2008). The Effect of Attributional Processes Concerning Medication Taking on Return of Fear. *Journal of Consulting and Clinical Psychology*, 76(3), 478–490. <https://doi.org/10.1037/0022-006X.76.3.478>

Price, M., Anderson, P., Henrich, C. C., & Rothbaum, B. O. (2008). Greater Expectations: Using Hierarchical Linear Modeling to Examine Expectancy for Treatment Outcome as a Predictor of Treatment Response. *Behavior Therapy*, 39(4), 398–405. <https://doi.org/10.1016/j.beth.2007.12.002>

Price, M., & Anderson, P. L. (2012). Outcome expectancy as a predictor of treatment response in cognitive behavioral therapy for public speaking fears within social anxiety disorder. *Psychotherapy*, 49(2), 173–179. <https://doi.org/10.1037/a0024734>

Priebe, S., Bhatti, N., Barnicot, K., Bremner, S., Gaglia, A., Katsakou, C., ... Zinkler, M. (2012). Effectiveness and cost-effectiveness of dialectical behaviour therapy for self-harming patients with personality disorder: A pragmatic randomised controlled trial. *Psychotherapy and Psychosomatics*, 81(6), 356–365. <https://doi.org/10.1159/000338897>

Purgato, M., & Adams, C. E. (2012). Heterogeneity: The issue of apples, oranges and fruit pie. *Epidemiology and Psychiatric Sciences*, 21(1), 27–29. <https://doi.org/10.1017/S2045796011000643>



- R.-Mercier, A., Masson, M., Bussi eres, E.-L., & Cellard, C. (2018). Common transdiagnostic cognitive deficits among people with psychiatric disorders exposed to childhood maltreatment: A meta-analysis. *Cognitive Neuropsychiatry*, 23(3), 180–197. <https://doi.org/10.1080/13546805.2018.1461617>
- Radhakrishnan, M., Hammond, G., Jones, P. B., Watson, A., McMillan-Shields, F., & Lafortune, L. (2013). Cost of Improving Access to Psychological Therapies (IAPT) programme: An analysis of cost of session, treatment and recovery in selected Primary Care Trusts in the East of England region. *Behaviour Research and Therapy*, 51(1), 37–45. <https://doi.org/10.1016/j.brat.2012.10.001>
- Raffin, A. L., Guimar es Fachel, J. M., Ferr o, Y. A., Pasquoto de Souza, F., & Cordioli, A. V. (2009). Predictors of response to group cognitive-behavioral therapy in the treatment of obsessive-compulsive disorder. *European Psychiatry: The Journal of the Association of European Psychiatrists*, 24(5), 297–306. <https://doi.org/10.1016/j.eurpsy.2008.12.001>
- Raj, S., Sachdeva, S. A., Jha, R., Sharad, S., Singh, T., Arya, Y. K., & Verma, S. K. (2019). Effectiveness of mindfulness based cognitive behavior therapy on life satisfaction, and life orientation of adolescents with depression and suicidal ideation. *Asian Journal of Psychiatry*, 39, 58–62. <https://doi.org/10.1016/j.ajp.2018.12.001>
- Ravitz, P., McBride, C., & Maunder, R. (2011). Failures in interpersonal psychotherapy (IPT): factors related to treatment resistances. *Journal of Clinical Psychology*, 67(11), 1129–1139. <https://doi.org/10.1002/jclp.20850>
- Regan, A. M., & Hill, C. E. (1992). Investigation of What Clients and Counselors Do Not Say in Brief Therapy. *Journal of Counseling Psychology*, 39(2), 168–174. <https://doi.org/10.1037/0022-0167.39.2.168>
- Reid, A. M., Garner, L. E., Van Kirk, N., Gironde, C., Kropfing, J. W., Brennan, B. P., ... Elias, J. A. (2017). How willing are you? Willingness as a predictor of change during treatment of adults with obsessive-compulsive disorder. *Depression and Anxiety*, (May), 1–8. <https://doi.org/10.1002/da.22672>
- Reimer, C. (2007). *Psychotherapie : ein Lehrbuch f ur  rzte und Psychologen: mit 35 Tabellen*. Berlin, Heidelberg: Springer.

- Reutfors, J., Andersson, T. M.-L., Brenner, P., Brandt, L., DiBernardo, A., Li, G., ... Bodén, R. (2018). Mortality in treatment-resistant unipolar depression: A register-based cohort study in Sweden. *Journal of Affective Disorders, 238*, 674–679. <https://doi.org/10.1016/j.jad.2018.06.030>
- Rheker, J., Beisel, S., Kräling, S., & Rief, W. (2017). Rate and predictors of negative effects of psychotherapy in psychiatric and psychosomatic inpatients. *Psychiatry Research, 254*(April), 143–150.
- Rhodes, R. H., Hill, C. E., Thompson, B. J., & Elliott, R. (1994). Client Retrospective Recall of Resolved and Unresolved Misunderstanding Events. *Journal of Counseling Psychology, 41*(4), 473–483. <https://doi.org/10.1037/0022-0167.41.4.473>
- Richardson, J. D., Contractor, A. A., Armour, C., St. Cyr, K., Elhai, J. D., & Sareen, J. (2014). Predictors of long-term treatment outcome in combat and peacekeeping veterans with military-related PTSD. *Journal of Clinical Psychiatry, 75*(11), e1299–e1305. <https://doi.org/10.4088/JCP.13m08796>
- Riedler-Singer, R. (2005). Behandlungsfehler in der Psychotherapie. *PiD - Psychotherapie Im Dialog, 6*(2), 162–165. <https://doi.org/10.1055/s-2004-834769>
- Rief, W., & Glombiewski, J. A. (2017). The role of expectations in mental disorders and their treatment. *World Psychiatry, 16*(2), 210–211. <https://doi.org/10.1002/wps.20427>
- Rief, W., & Glombiewski, J. A. (2016). Erwartungsfokussierte Psychotherapeutische Interventionen (EFPI). *Verhaltenstherapie, 26*, 47–54. <https://doi.org/10.1159/000442374>
- Rief, W., Glombiewski, J. A., Gollwitzer, M., Schubö, A., Schwarting, R., & Thorwart, A. (2015). Expectancies as core features of mental disorders. *Current Opinion in Psychiatry, 28*(5), 378–385. <https://doi.org/10.1097/YCO.0000000000000184>
- Rief, W., & Hofmann, S. G. (2018). Some problems with non-inferiority tests in psychotherapy research: psychodynamic therapies as an example. *Psychological Medicine, 48*(08), 1392–1394. <https://doi.org/10.1017/S0033291718000247>
- Rief, W., & Hofmann, S. G. (2019). The limitations of equivalence and non-inferiority trials. *Psychological Medicine, 49*(2), 349–350. <https://doi.org/10.1017/S0033291718002891>

- Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E., Moons, K. G. M., & Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction model: Part I - Continuous outcomes. *Statistics in Medicine*, *38*(7), 1262–1275. <https://doi.org/10.1002/sim.7993>
- Riley, R. D., Snell, K. I., Ensor, J., Burke, D. L., Harrell Jr, F. E., Moons, K. G., & Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Statistics in Medicine*, *38*(7), 1276–1296. <https://doi.org/10.1002/sim.7992>
- Rizvi, S., Vogt, D., & Resick, P. (2012). Cognitive and affective predictors of treatment outcome in cognitive processing therapy and prolonged exposure for posttraumatic stress disorder. *Behavioral Research and Therapy*, *47*(9), 737–743. <https://doi.org/10.1016/j.brat.2009.06.003>. Cognitive
- Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2019). The network approach to psychopathology: a review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, 1–14. <https://doi.org/10.1017/S0033291719003404>
- Robinson, L. D., & Jewell, N. P. (1991). Some Surprising Results about Covariate Adjustment in Logistic Regression Models. *International Statistical Review / Revue Internationale de Statistique*, *59*(2), 227. <https://doi.org/10.2307/1403444>
- Robinson, L., Delgadillo, J., & Kellett, S. (2020). The dose-response effect in routinely delivered psychological therapies: A systematic review. *Psychotherapy Research*, *30*(1), 79–96. <https://doi.org/10.1080/10503307.2019.1566676>
- Robinson, L., Kellett, S., & Delgadillo, J. (2020). Dose-response patterns in low and high intensity cognitive behavioral therapy for common mental health problems. *Depression and Anxiety*, da.22999. <https://doi.org/10.1002/da.22999>
- Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin*, *113*(3), 553–565. <https://doi.org/10.1037/0033-2909.113.3.553>
- Rose, T. (2016). *The end of average : how we succeed in a world that values sameness*. San Francisco, CA: HarperOne.

- Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry*, 6(3), 412–415. <https://doi.org/10.1111/j.1939-0025.1936.tb05248.x>
- Roth, A., & Fonagy, P. (2005). *What works for whom?: a critical review of psychotherapy research*. New York, NY: Guilford Press.
- Roth, C., Siegl, J., Aufdermauer, N., & Reinecker, H. (2004). Therapie von Angst- und Zwangspatienten in der verhaltenstherapeutischen Praxis. *Verhaltenstherapie*, 14(1), 16–21. <https://doi.org/10.1159/000078027>
- Rubel, J. A., Fisher, A. J., Husen, K., & Lutz, W. (2018). Translating Person-Specific Network Models into Personalized Treatments: Development and Demonstration of the Dynamic Assessment Treatment Algorithm for Individual Networks (DATA-IN). *Psychotherapy and Psychosomatics*, 87(4), 249–251. <https://doi.org/10.1159/000487769>
- Rubel, J. A., Zilcha-Mano, S., Giesemann, J., Prinz, J., & Lutz, W. (2019). Predicting personalized process-outcome associations in psychotherapy using machine learning approaches-A demonstration. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 1–10. <https://doi.org/10.1080/10503307.2019.1597994>
- Rubel, J. A., Zimmermann, D., Deisenhofer, A. K., Müller, V., & Lutz, W. (2017). Nutzung von psychometrischem Feedback als empirische Unterstützung des Supervisionsprozesses bei Ausbildungstherapien. *Zeitschrift Fur Klinische Psychologie Und Psychotherapie*, 46(2), 83–95. <https://doi.org/10.1026/1616-3443/a000413>
- Rubel, J., Ehrlich, T., & Lutz, W. (2015). Feedback-Ansätze in der Psychotherapie. *PiD - Psychotherapie Im Dialog*, 16(04), 30–34. <https://doi.org/10.1055/s-0041-105247>
- Rubel, J., & Lutz, W. (2016). „Personalisierung statt Verfahrensinflation“ – Oder: Warum die Psychotherapieforschung in Deutschland ausgebaut werden sollte/könnte/müsste? Teil 2. *PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie*, 66(12), 463–464. <https://doi.org/10.1055/s-0042-119573>
- Rudd, B. N., Last, B. S., Gregor, C., Jackson, K., Berkowitz, S., Zinny, A., ... Beidas, R. S. (2019). Benchmarking Treatment Effectiveness of Community-Delivered Trauma-Focused Cognitive Behavioral Therapy. *American Journal of Community Psychology*, 64(3–4), 438–

450. <https://doi.org/10.1002/ajcp.12370>

Rüsch, N., Schiel, S., Corrigan, P. W., Leihener, F., Jacob, G. A., Olschewski, M., ... Bohus, M. (2008). Predictors of dropout from inpatient dialectical behavior therapy among women with borderline personality disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(4), 497–503. <https://doi.org/10.1016/j.jbtep.2007.11.006>

Sackett, D. L., & Rosenberg, W. M. C. (1995a). On the need for evidence-based medicine. *Journal of Public Health*, 88(11), 620–624. <https://doi.org/10.1093/oxfordjournals.pubmed.a043127>

Sackett, D. L., & Rosenberg, W. M. C. (1995b). The need for evidence-based medicine. *Journal of the Royal Society of Medicine*, Vol. 88, pp. 620–624. <https://doi.org/10.1093/oxfordjournals.pubmed.a043127>

Sakiris, N., & Berle, D. (2019). A systematic review and meta-analysis of the Unified Protocol as a transdiagnostic emotion regulation based intervention. *Clinical Psychology Review*, 72, 101751. <https://doi.org/10.1016/j.cpr.2019.101751>

Salkovskis, P. M. (1995). Demonstrating specific effects in cognitive and behavioural therapy. In M. Aveline & D. A. Shapiro (Eds.), *Research foundations for psychotherapy practice*. Chichester, UK: Wiley.

Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: Therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology*, 80(4), 535–546. <https://doi.org/10.1037/a0028898>

Scheeres, K., Wensing, M., Knoop, H., & Bleijenberg, G. (2008). Implementing Cognitive Behavioral Therapy for Chronic Fatigue Syndrome in a Mental Health Center: A Benchmarking Evaluation. *Journal of Consulting and Clinical Psychology*, 76(1), 163–171. <https://doi.org/10.1037/0022-006X.76.1.163>

Schefft, C., Guhn, A., Brakemeier, E.-L., Sterzer, P., & Köhler, S. (2019). Efficacy of inpatient psychotherapy for major depressive disorder: a meta-analysis of controlled trials. *Acta Psychiatrica Scandinavica*, 139(4), 322-335. <https://doi.org/10.1111/acps.12995>

Schermuly-Haupt, M.-L., & Linden, M. (2020). Erwartungen und Kenntnisse von

Ausbildungskandidaten in Verhaltenstherapie zu Psychotherapieebenenwirkungen. *Verhaltenstherapie*, 1–7. <https://doi.org/10.1159/000505594>

Schmoll, D. (2012). Nebenwirkungen bei psychoanalytisch orientierten Therapien. *Psychotherapeut*, 57(5), 395–401. <https://doi.org/10.1007/s00278-012-0933-9>

Schneibel, R., Wilbertz, G., Scholz, C., Becker, M., Brakemeier, E. L., Bschor, T., ... Schmoll, D. (2017). Adverse events of group psychotherapy in the in-patient setting - results of a naturalistic trial. *Acta Psychiatrica Scandinavica*, 1–12.

Schottenbauer, M. A., Glass, C. R., Arnkoff, D. B., Tendick, V., & Gray, S. H. (2008). Nonresponse and Dropout Rates in Outcome Studies on PTSD: Review and Methodological Considerations. *Psychiatry: Interpersonal and Biological Processes*, 71(2), 134–168. <https://doi.org/10.1521/psyc.2008.71.2.134>

Schramm, E., Kriston, L., Elsaesser, M., Fangmeier, T., Meister, R., Bausch, P., ... Härter, M. (2019). Two-Year Follow-Up after Treatment with the Cognitive Behavioral Analysis System of Psychotherapy versus Supportive Psychotherapy for Early-Onset Chronic Depression. *Psychotherapy and Psychosomatics*, 88(3), 154–164. <https://doi.org/10.1159/000500189>

Schramm, E., Kriston, L., Zobel, I., Bailer, J., Wambach, K., Backenstrass, M., ... Härter, M. (2017). Effect of Disorder-Specific vs Nonspecific Psychotherapy for Chronic Depression. *JAMA Psychiatry*, 74(3), 233.

Schruers, K., Koning, K., Luermans, J., Haack, M. J., & Griez, E. (2005). Obsessive-compulsive disorder: A critical review of therapeutic perspectives. *Acta Psychiatrica Scandinavica*, 111(4), 261–271. <https://doi.org/10.1111/j.1600-0447.2004.00502.x>

Schulte, D. (1993). Wie soll Therapieerfolg gemessen werden? *Zeitschrift Für Klinische Psychologie*, 22(4), 374–393.

Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M., & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multilevel autoregressive model. *Psychological Methods*, 21(2), 206–221. <https://doi.org/10.1037/met0000062>

Schwarzmeier, H., Lehr, E. J., Böhnlein, J., Seeger, F. R., Roesmann, K., Gathmann, B., ...

- Dannlowski, U. (2019). Theranostic markers for personalized therapy of spider phobia: Methods of a bicentric external cross-validation machine learning approach. *International Journal of Methods in Psychiatric Research*, e1812. <https://doi.org/10.1002/mpr.1812>
- Scott, J. (2017). The science and practice of psychotherapy: the critical need for quality assurance and vigilance to monitor the ratio of benefits to risks of therapies. *Acta Psychiatrica Scandinavica*, 136(3), 233–235. <https://doi.org/10.1111/acps.12780>
- Scott, J., & Young, A. H. (2016). Psychotherapies should be assessed for both benefit and harm. *British Journal of Psychiatry*, 208(3), 208–209. <https://doi.org/10.1192/bjp.bp.115.169060>
- Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist*, 50(12), 965–974. <https://doi.org/10.1037//0003-066x.50.12.965>
- Sempértegui, G. A., Karreman, A., Arntz, A., & Bekker, M. H. J. (2013). Schema therapy for borderline personality disorder: A comprehensive review of its empirical foundations, effectiveness and implementation possibilities. *Clinical Psychology Review*, 33(3), 426–447. <https://doi.org/10.1016/j.cpr.2012.11.006>
- Serbanescu, I., Walter, H., Schnell, K., Kessler, H., Weber, B., Drost, S., ... Schoepf, D. (2020). Combining baseline characteristics to disentangle response differences to disorder-specific versus supportive psychotherapy in patients with persistent depressive disorder. *Behaviour Research and Therapy*, 124, 103512. <https://doi.org/10.1016/j.brat.2019.103512>
- Serlin, R. C., & Lapsley, D. K. (1985). Rationality in psychological research: The good-enough principle. *American Psychologist*, 40(1), 73–83. <https://doi.org/10.1037/0003-066X.40.1.73>
- Serlin, R. C., & Lapsley, D. K. (1993). Rational appraisal of psychological research and the good-enough principle. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues*. (pp. 199–228). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Shadish, W R, Matt, G. E., Navarro, A. M., Siegle, G., Crits-Christoph, P., Hazelrigg, M. D., ... Weiss, B. (1997). Evidence that therapy works in clinically representative conditions. *Journal of Consulting and Clinical Psychology*, 65(3), 355–365.

- Shadish, William R., Navarro, A. M., Matt, G. E., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin*, *126*(4), 512–529. <https://doi.org/10.1037/0033-2909.126.4.512>
- Shakow, D. (1976). What is clinical psychology? *The American Psychologist*, *31*(8), 553–560. <https://doi.org/10.1037/0003-066X.31.8.553>
- Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review*, *17*(8), 881–901. [https://doi.org/10.1016/S0272-7358\(97\)00056-1](https://doi.org/10.1016/S0272-7358(97)00056-1)
- Shearer-Underhill, C., & Marker, C. (2010). The Use of the Number Needed to Treat (NNT) in Randomized Clinical Trials in Psychological Treatment. *Clinical Psychology: Science and Practice*, *17*(1), 41–47. <https://doi.org/10.1111/j.1468-2850.2009.01191.x>
- Shick Tryon, G., Birch, S. E., & Verkuilen, J. (2018). Meta-Analyses of the Relation of Goal Consensus and Collaboration to Psychotherapy Outcome. *Psychotherapy*, *55*(4), 372–383. <https://doi.org/10.1037/pst0000170>
- Shimazu, K., Shimodera, S., Mino, Y., Nishida, A., Kamimura, N., Sawada, K., ... Inoue, S. (2011). Family psychoeducation for major depression: randomised controlled trial. *British Journal of Psychiatry*, *198*(5), 385–390. <https://doi.org/10.1192/bjp.bp.110.078626>
- Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing Treatment Outcome of Patients at Risk of Treatment Failure: Meta-Analytic and Mega-Analytic Review of a Psychotherapy Quality Assurance System. *Journal of Consulting and Clinical Psychology*, *78*(3), 298–311. <https://doi.org/10.1037/a0019247>
- Shiratori, Y., & Samuelson, K. W. (2012). Relationship between posttraumatic stress disorder and asthma among New York area residents exposed to the World Trade Center disaster. *Journal of Psychosomatic Research*, *73*(2), 122–125. <https://doi.org/10.1016/j.jpsychores.2012.05.003>
- Siddiqui, O., Hung, H. M. J., & O'Neill, R. (2009). MMRM vs. LOCF: A Comprehensive Comparison Based on Simulation Study and 25 NDA Datasets. *Journal of Biopharmaceutical Statistics*, *19*(2), 227–246. <https://doi.org/10.1080/10543400802609797>



- Silvia, P. J. (2007). *How to write a lot: a practical guide to productive academic writing*. American Psychological Association.
- Simon, G. E., & Perlis, R. H. (2010). Personalized medicine for depression: can we match patients with treatments? *The American Journal of Psychiatry*, *167*(12), 1445–1455. <https://doi.org/10.1176/appi.ajp.2010.09111680>
- Simon, W., Lambert, M. J., Harris, M. W., Busath, G., & Vazquez, A. (2012). Providing patient progress information and clinical support tools to therapists: Effects on patients at risk of treatment failure. *Psychotherapy Research*, *22*(6), 638–647. <https://doi.org/10.1080/10503307.2012.698918>
- Simpson, H. B., Maher, M. J., Wang, Y., Bao, Y., Foa, E. B., & Franklin, M. (2011). Patient adherence predicts outcome from cognitive behavioral therapy in obsessive-compulsive disorder. *Journal of Consulting and Clinical Psychology*, *79*(2), 247–252. <https://doi.org/10.1037/a0022659>
- Skapinakis, P., Caldwell, D. M., Hollingworth, W., Welton, N. J., Fineberg, N., Salkovskis, P., ... Lewis, G. (2016). Network meta-analyses and treatment recommendations for obsessive-compulsive disorder - Authors' reply. *The Lancet Psychiatry*, *3*(10), 921–922. [https://doi.org/10.1016/S2215-0366\(16\)30282-6](https://doi.org/10.1016/S2215-0366(16)30282-6)
- Slade, E. P., Gottlieb, J. D., Lu, W., Yanos, P. T., Rosenberg, S., Silverstein, S. M., ... Mueser, K. T. (2017). Cost-Effectiveness of a PTSD Intervention Tailored for Individuals With Severe Mental Illness. *Psychiatric Services (Washington, D.C.)*, *68*(12), 1225–1231. <https://doi.org/10.1176/appi.ps.201600474>
- Sobell, L. C. (1996). Bridging the gap between scientists and practitioners: The challenge before us. *Behavior Therapy*, *27*(3), 297–320. [https://doi.org/10.1016/S0005-7894\(96\)80019-0](https://doi.org/10.1016/S0005-7894(96)80019-0)
- Sollberger, D., & Walter, M. (2010). Psychotherapie der Borderline-Persönlichkeitsstörung: Gemeinsamkeiten und Differenzen evidenzbasierter störungsspezifischer Behandlungen. *Fortschritte Der Neurologie Psychiatrie*, *78*(12), 698–708. <https://doi.org/10.1055/s-0029-1245626>
- Sonne, C., Carlsson, J., Bech, P., Vindbjerg, E., Mortensen, E. L., & Elklit, A. (2016). Psychosocial predictors of treatment outcome for trauma-affected refugees. *European Journal Of*

*Psychotraumatology*, 7, 30907. <https://doi.org/10.3402/ejpt.v7.30907>

Stamoulos, C., Trepanier, L., Bourkas, S., Bradley, S., Stelmaszczyk, K., Schwartzman, D., & Drapeau, M. (2016). Psychologists' perceptions of the importance of common factors in psychotherapy for successful treatment outcomes. *Journal of Psychotherapy Integration*, 26(3), 300–317. <https://doi.org/10.1037/a0040426>

Steinert, C., Hofmann, M., Kruse, J., & Leichsenring, F. (2014). Relapse rates after psychotherapy for depression - stable long-term effects? A meta-analysis. *Journal of Affective Disorders*, 168, 107–118. <https://doi.org/10.1016/j.jad.2014.06.043>

Steinert, C., Munder, T., Rabung, S., Hoyer, J., Leichsenring, F., Rabung, S., & Steinert, C. (2017). Psychodynamic Therapy: As Efficacious as Other Empirically Supported Treatments? A Meta-Analysis Testing Equivalence of Outcomes. *American Journal of Psychiatry*, 174(10), 943–953. <https://doi.org/10.1176/appi.ajp.2017.17010057>

Steinfeld, B. I., Coffman, S. J., & Keyes, J. A. (2009). Implementation of an evidence-based practice in a clinical setting: What happens when you get there? *Professional Psychology: Research and Practice*, 40(4), 410–416. <https://doi.org/10.1037/a0015677>

Steketee, G., Eisen, J., Dyck, I., Warshaw, M., & Rasmussen, S. (1999). Predictors of course in obsessive compulsive disorder. *Psychiatry Research*, 89(3), 229–238. [https://doi.org/10.1016/S0165-1781\(99\)00104-3](https://doi.org/10.1016/S0165-1781(99)00104-3)

Steketee, G., Siev, J., Fama, J. M., Keshaviah, A., Chosak, A., & Wilhelm, S. (2011). Predictors of treatment outcome in modular cognitive therapy for obsessive-compulsive disorder. *Depression and Anxiety*, 28(4), 333–341. <https://doi.org/10.1002/da.20785>

Stephan, K. E., Siemerikus, J., Bischof, M., & Haker, H. (2017). Hat Computational Psychiatry Relevanz für die klinische Praxis der Psychiatrie? *Zeitschrift Für Psychiatrie, Psychologie Und Psychotherapie*, 65(1), 9–19. <https://doi.org/10.1024/1661-4747/a000296>

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., ... Corlett, P. R. (2018). The Predictive Coding Account of Psychosis. *Biological Psychiatry*, 84(9), 634–643. <https://doi.org/10.1016/J.BIOPSYCH.2018.05.015>

Stewart, R. E., & Chambless, D. L. (2009). Cognitive-Behavioral Therapy for Adult Anxiety

Disorders in Clinical Practice: A Meta-Analysis of Effectiveness Studies. *Journal of Consulting and Clinical Psychology*, 77(4), 595–606. <https://doi.org/10.1037/a0016032>

Steyerberg, E. W. (2019). *Clinical Prediction Models* (2nd edition). Cham: Springer International Publishing.

Stiles-Shields, C., Corden, M. E., Kwasny, M. J., Schueller, S. M., & Mohr, D. C. (2015). Predictors of outcome for telephone and face-to-face administered cognitive behavioral therapy for depression. *Psychological Medicine*, 45(15), 3205–3215. <https://doi.org/10.1017/S0033291715001208>

Stiles, W. B., Barkham, M., Mellor-Clark, J., & Connell, J. (2008). Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies in UK primary-care routine practice: Replication in a larger sample. *Psychological Medicine*, 38(5), 677–688. <https://doi.org/10.1017/S0033291707001511>

Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., ... Hardy, G. E. (2003). Early sudden gains in psychotherapy under routine clinic conditions: Practice-based evidence. *Journal of Consulting and Clinical Psychology*, 71(1), 14–21. <https://doi.org/10.1037/0022-006X.71.1.14>

Storebø, O. J., Stoffers-Winterling, J. M., Völlm, B. A., Kongerslev, M. T., Mattivi, J. T., Jørgensen, M. S., ... Simonsen, E. (2020). Psychological therapies for people with borderline personality disorder. *Cochrane Database of Systematic Reviews*, (5). <https://doi.org/10.1002/14651858.CD012955.pub2>

Strauß, B., & Linden, M. (2013). *Risiken und Nebenwirkungen von Psychotherapie : Erfassung, Bewältigung, Risikovermeidung*. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.

Suh, C. S., Strupp, H. H., & O'Malley, S. S. (1986). The Vanderbilt Process Measures: The Psychotherapy Process Scale (VPPS) and the Negative Indicators Scale (VNIS). In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 285–323). New York, London: Guilford Press.

Sunderji, N., Nicholas Angl, E., Polaha, J., & Gao, C. (2019). Why and how to use patient-oriented research to promote translational research. *Families, Systems, & Health*, 37(1), 1–9.

<https://doi.org/10.1037/fsh0000405>

Swenson, C. R., Sanderson, C., Dulit, R. A., & Linehan, M. M. (2001). The Application of Dialectical Behavior Therapy for Patients with Borderline Personality Disorder on Inpatient Units. *Psychiatric Quarterly*, *72*(4), 307–324. <https://doi.org/10.1023/A:1010337231127>

Swift, J. K., Greenberg, R. P., Tompkins, K. A., & Parkin, S. R. (2017). Treatment refusal and premature termination in psychotherapy, pharmacotherapy, and their combination: A meta-analysis of head-to-head comparisons. *Psychotherapy*, *54*(1), 47–57. <https://doi.org/10.1037/pst0000104>

Sydow, K. von, & Borst, U. (2018). *Systemische Therapie in der Praxis*. Weinheim: Beltz.

Szapocznik, J., & Prado, G. (2007). Negative effects on family functioning from psychosocial treatments: a recommendation for expanded safety monitoring. *Journal of Family Psychology*, *21*(3), 468–478.

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, *67*(6), 894–904. <https://doi.org/10.1037//0022-006x.67.6.894>

Tang, T. Z., & DeRubeis, R. J. (2006). Reconsidering Rapid Early Response in Cognitive Behavioral Therapy for Depression. *Clinical Psychology: Science and Practice*, *6*(3), 283–288. <https://doi.org/10.1093/clipsy.6.3.283>

Teachman, B. A., Drabick, D. A. G., Hershenberg, R., Vivian, D., Wolfe, B. E., & Goldfried, M. R. (2012). Bridging the gap between clinical research and clinical practice: Introduction to the special section. *Psychotherapy*, *49*(2), 97–100. <https://doi.org/10.1037/a0027346>

Teismann, T., Koban, C., Illes, F., & Oermann, A. (2016). *Psychotherapie suizidaler Patienten: Therapeutischer Umgang mit Suizidgedanken und Suizidversuchen*. Göttingen: Hogrefe.

Thase, M. E., Reynolds, C. F., Frank, E., Simons, A. D., McGeary, J., Fasiczka, A. L., ... Kupfer, D. J. (1994). Do depressed men and women respond similarly to cognitive behavior therapy? *American Journal of Psychiatry*, *151*(4), 500–505. <https://doi.org/10.1176/ajp.151.4.500>

Tolin, D. F. (2010). Is cognitive-behavioral therapy more effective than other therapies?. A meta-analytic review. *Clinical Psychology Review*, *30*, 710–720.

<https://doi.org/10.1016/j.cpr.2010.05.003>

Tolin, D. F. (2014). Beating a Dead Dodo Bird: Looking at Signal vs. Noise in Cognitive-Behavioral Therapy for Anxiety Disorders. *Clinical Psychology: Science and Practice*, 21(4), 351–362. <https://doi.org/10.1111/cpsp.12080>

Tolin, D. F. (2015). Corrigendum. *Clinical Psychology: Science and Practice*, 22(3), 315–316. <https://doi.org/10.1111/cpsp.12109>

Trivedi, M. H. (2013). Modeling Predictors, Moderators and Mediators of Treatment Outcome and Resistance in Depression. *Biological Psychiatry*, 74(1), 2–4. <https://doi.org/10.1016/J.BIOPSYCH.2013.05.009>

Tryon, G. S., & Winograd, G. (2011). Goal Consensus and Collaboration. *Psychotherapy*, 48(1), 50–57. <https://doi.org/10.1037/a0022061>

Underner, M., Goutaudier, N., Peiffer, G., Perriot, J., Harika-Germaneau, G., & Jaafari, N. (2019). Influence du trouble stress post-traumatique sur l'asthme. *La Presse Médicale*, 48(5), 488–502. <https://doi.org/10.1016/j.lpm.2019.03.002>

van Bronswijk, S. C., DeRubeis, R. J., Lemmens, L. H. J. M., Peeters, F. P. M. L., Keefe, J. R., Cohen, Z. D., & Huibers, M. J. H. (2019). Precision medicine for long-term depression outcomes using the Personalized Advantage Index approach: cognitive therapy or interpersonal psychotherapy? *Psychological Medicine*, (November), 1–11. <https://doi.org/10.1017/s0033291719003192>

van Bronswijk, S. C., Moopen, N., Beijers, L., Ruhe, H. G., & Peeters, F. (2019). Effectiveness of psychotherapy for treatment-resistant depression: a meta-analysis and meta-regression. *Psychological Medicine*, 49(3), 366–379. <https://doi.org/10.1017/S003329171800199X>

Van den Bergh, O., Witthöft, M., Petersen, S., & Brown, R. J. (2017). Symptoms and the body: Taking the inferential leap. *Neuroscience & Biobehavioral Reviews*, 74, 185–203. <https://doi.org/10.1016/j.neubiorev.2017.01.015>

Van Minnen, A., Arntz, A., & Keijsers, G. P. J. (2002). Prolonged exposure in patients with chronic PTSD: Predictors of treatment outcome and dropout. *Behaviour Research and Therapy*, 40(4), 439–457. [https://doi.org/10.1016/S0005-7967\(01\)00024-9](https://doi.org/10.1016/S0005-7967(01)00024-9)

- Van Noppen, B., & Steketee, G. (2009). Testing a conceptual model of patient and family predictors of obsessive compulsive disorder (OCD) symptoms. *Behaviour Research and Therapy*, 47(1), 18–25. <https://doi.org/10.1016/j.brat.2008.10.005>
- Vaughan, B., Goldstein, M. H., Alikakos, M., Cohen, L. J., & Serby, M. J. (2014). Frequency of reporting of adverse events in randomized controlled trials of psychotherapy vs. psychopharmacotherapy. *Comprehensive Psychiatry*, 55(4), 849–855. <https://doi.org/10.1016/j.comppsy.2014.01.001>
- Veale, D., Naismith, I., Miles, S., Gledhill, L. J., Stewart, G., & Hodsoll, J. (2016). Outcomes for residential or inpatient intensive treatment of obsessive-compulsive disorder: A systematic review and meta-analysis. *Journal of Obsessive-Compulsive and Related Disorders*, 8, 38–49. <https://doi.org/10.1016/j.jocrd.2015.11.005>
- Voderholzer, U., Schlegl, S., Diedrich, A., Külz, A. K., Thiel, N., Hertenstein, E., ... Moritz, S. (2015). Versorgung Zwangserkrankter mit kognitiver Verhaltenstherapie als Behandlungsmethode erster Wahl. *Verhaltenstherapie*, 25(3), 183–190. <https://doi.org/10.1159/000438717>
- Voderholzer, U., Schwartz, C., Freyer, T., Zurovski, B., Thiel, N., Herbst, N., ... Kuelz, A. K. (2013). Cognitive functioning in medication-free obsessive-compulsive patients treated with cognitive-behavioural therapy. *Journal of Obsessive-Compulsive and Related Disorders*, 2(3), 241–248. <https://doi.org/10.1016/j.jocrd.2013.03.003>
- Vogel, P. A., Hansen, B., Stiles, T. C., & Götestam, K. G. (2006). Treatment motivation, treatment expectancy, and helping alliance as predictors of outcome in cognitive behavioral treatment of OCD. *Journal of Behavior Therapy and Experimental Psychiatry*, 37(3), 247–255. <https://doi.org/10.1016/j.jbtep.2005.12.001>
- von Brachel, R., Hirschfeld, G., Berner, A., Willutzki, U., Teismann, T., Cwik, J. C., ... Margraf, J. (2019). Long-Term Effectiveness of Cognitive Behavioral Therapy in Routine Outpatient Care: A 5- to 20-Year Follow-Up Study. *Psychotherapy and Psychosomatics*, 88(4), 225–235. <https://doi.org/10.1159/000500188>
- Vos, T., Corry, J., Haby, M. M., Carter, R., & Andrews, G. (2005). Cost-Effectiveness of Cognitive–Behavioural Therapy and Drug Interventions for Major Depression. *Australian & New*

*Zealand Journal of Psychiatry*, 39(8), 683–692. <https://doi.org/10.1080/j.1440-1614.2005.01652.x>

Wade, W. A., Treat, T. A., & Stuart, G. L. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology*, 66(2), 231–239. <https://doi.org/10.1037//0022-006x.66.2.231>

Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An Investigation of Self-Assessment Bias in Mental Health Providers. *Psychological Reports*, 110(2), 639–644. <https://doi.org/10.2466/02.07.17.PR0.110.2.639-644>

Waller, G., & Turner, H. (2016). Therapist drift redux: Why well-meaning clinicians fail to deliver evidence-based therapy, and how to get back on track. *Behaviour Research and Therapy*, 77, 129–137. <https://doi.org/10.1016/j.brat.2015.12.005>

Wampold, B. E., Budge, S. L., Laska, K. M., Del Re, A. C., Beardseth, T. P., Flückiger, C., ... Gunn, W. (2011). Evidence-based treatments for depression and anxiety versus treatment-as-usual: A meta-analysis of direct comparisons. *Clinical Psychology Review*, 31(8), 1304–1312. <https://doi.org/10.1016/j.cpr.2011.07.012>

Wampold, B. E., Flückiger, C., Del Re, A. C., Yulish, N. E., Frost, N. D., Pace, B. T., ... Hilsenroth, M. J. (2017). In pursuit of truth: A critical examination of meta-analyses of cognitive behavior therapy. *Psychotherapy Research*, 27(1), 14–32. <https://doi.org/10.1080/10503307.2016.1249433>

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (Second edition). London: Routledge.

Wampold, B. E., Minami, T., Tierney, S. C., Baskin, T. W., & Bhati, K. S. (2005). The placebo is powerful: Estimating placebo effects in medicine and psychotherapy from randomized clinical trials. *Journal of Clinical Psychology*, 61, 835–854. <https://doi.org/10.1002/jclp.20129>

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. N. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, “all must have prizes.” *Psychological Bulletin*, 122(3), 203–215. <https://doi.org/10.1037/0033->

2909.122.3.203

- Wampold, B. E., & Serlin, R. C. (2000). The consequence of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods*, 5(4), 425–433. <https://doi.org/10.1037/1082-989X.5.4.425>
- Watts, B. V., Schnurr, P. P., Mayo, L., Young-Xu, Y., Weeks, W. B., & Friedman, M. J. (2013). Meta-analysis of the efficacy of treatments for posttraumatic stress disorder. *Journal of Clinical Psychiatry*, 74(6). <https://doi.org/10.4088/JCP.12r08225>
- Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., ... Marx, B. P. (2018). The clinician-administered PTSD scale for DSM-5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychological Assessment*, 30(3), 383–395. <https://doi.org/10.1037/pas0000486>
- Webb, C. A., Cohen, Z. D., Beard, C., Forgeard, M., Peckham, A. D., & Björgvinsson, T. (2020). Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: A comparison of machine learning approaches. *Journal of Consulting and Clinical Psychology*, 88(1), 25–38. <https://doi.org/10.1037/ccp0000451>
- Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist Adherence/Competence and Treatment Outcome: A Meta-Analytic Review. *Journal of Consulting and Clinical Psychology*, 78(2), 200–211. <https://doi.org/10.1037/a0018912>
- Webb, C. A., Trivedi, M. H., Cohen, Z. D., Dillon, D. G., Fournier, J. C., Goer, F., ... Pizzagalli, D. A. (2019). Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study. *Psychological Medicine*, 49(07), 1118–1127. <https://doi.org/10.1017/S0033291718001708>
- Webb, D., & McMurrin, M. (2009). A comparison of women who continue and discontinue treatment for borderline personality disorder. *Personality and Mental Health*, 3(2), 142–149. <https://doi.org/10.1002/pmh.69>
- Weersing, V. R., & Weisz, J. R. (2002). Community clinic treatment of depressed youth: Benchmarking usual care against CBT clinical trials. *Journal of Consulting and Clinical Psychology*, 70(2), 299–310. <https://doi.org/10.1037/0022-006X.70.2.299>



- Weisz, B., Caron, A., Ball, S., Tapp, J., Johnson, M., & Weisz, J. R. (2005). Iatrogenic effects of group treatment for antisocial youth. *Journal of Consulting and Clinical Psychology, 73*, 1036–1044. <https://doi.org/10.1037/0022-006X.73.6.1036>
- Weisz, J. R., Doss, A. J., & Hawley, K. M. (2005). Youth Psychotherapy Outcome Research: A Review and Critique of the Evidence Base. *Annual Review of Psychology, 56*(1), 337–363. <https://doi.org/10.1146/annurev.psych.55.090902.141449>
- Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2006, October). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist, Vol. 61*, pp. 671–689. <https://doi.org/10.1037/0003-066X.61.7.671>
- Weisz, J. R., Weersing, V. R., & Henggeler, S. W. (2005). Jousting with straw men: Comment on Westen, Novotny, and Thompson-Brenner (2004). *Psychological Bulletin, 131*(3), 418–426. <https://doi.org/10.1037/0033-2909.131.3.418>
- Wells, A., & Fisher, P. (2016). *Treating depression: MCT, CBT and Third Wave Therapies*. Hoboken, NJ: Wiley-Blackwell.
- Westbrook, D., & Kirk, J. (2005). The clinical effectiveness of cognitive behaviour therapy: Outcome for a large sample of adults treated in routine practice. *Behaviour Research and Therapy, 43*(10), 1243–1261. <https://doi.org/10.1016/j.brat.2004.09.006>
- Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology, 69*(6), 875–899. <https://doi.org/10.1037/0022-006X.69.6.875>
- Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin, 130*, 631–663. <https://doi.org/10.1037/0033-2909.130.4.631>
- Whiston, A., Bockting, C. L. H., & Semkovska, M. (2019). Towards personalising treatment: A systematic review and meta-analysis of face-to-face efficacy moderators of cognitive-behavioral therapy and interpersonal psychotherapy for major depressive disorder. *Psychological Medicine, 49*(16), 2657–2668.

<https://doi.org/10.1017/S0033291719002812>

Wittmund, B. (2007). Stationäre Psychotherapie – über unerwünschte Wirkungen und andere Katastrophen. *Psychotherapie Im Dialog*, 8, 67–70. <https://doi.org/10.1055/s-2006-951990>

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67–82. <https://doi.org/10.1109/4235.585893>

Wortmann, J. H., Jordan, A. H., Weathers, F. W., Resick, P. A., Dondanville, K. A., Hall-Clark, B., ... Litz, B. T. (2016). Psychometric analysis of the PTSD checklist-5 (PCL-5) among treatment-seeking military service members. *Psychological Assessment*, 28(11), 1392–1403. <https://doi.org/10.1037/pas0000260>

Young, J. E., Klosko, J. S., & Weishaar, M. E. (2003). *Schema therapy : a practitioner's guide*. New York, NY: Guilford Press.

Zala, D., Brabban, A., Stirzaker, A., Kartha, M. R., & McCrone, P. (2019). The Cost-Effectiveness of the Improving Access to Psychological Therapies (IAPT) Programme in Severe Mental Illness: A Decision Analytical Model Using Routine Data. *Community Mental Health Journal*, 55(5), 873–883. <https://doi.org/10.1007/s10597-019-00390-z>

Zilcha-Mano, S. (2019). Major developments in methods addressing for whom psychotherapy may work and why. *Psychotherapy Research*, 29(6), 693–708. <https://doi.org/10.1080/10503307.2018.1429691>

Zimmerman, M., Walsh, E., Friedman, M., Boerescu, D. A., & Attiullah, N. (2018). Are self-report scales as effective as clinician rating scales in measuring treatment response in routine clinical practice? *Journal of Affective Disorders*, 225, 449–452. <https://doi.org/10.1016/j.jad.2017.08.024>

## Anhang A: Studie 1

Running head: Predictors of treatment outcomes in BPD

1

1

2

Drawing the Borderline:

3

Predicting treatment outcomes in patients with borderline personality disorder

4

5

Philipp Herzog\*,

6

Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,

7

Gutenbergstraße 18, D-35032 Marburg, Germany

8

Matthias Feldmann,

9

Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,

10

Gutenbergstraße 18, D-35032 Marburg, Germany

11

Ulrich Voderholzer,

12

Schön-Klinik Roseneck, Psychosomatic Clinic,

13

Am Roseneck 6, D-83209 Prien am Chiemsee, Germany

14

Thomas Gärtner,

15

Schön-Klinik Bad Arolsen, Psychosomatic Clinic,

16

Hofgarten 10, D-34454 Bad Arolsen, Germany

17

Michael Armbrust,

18

Birkenweg 10, Psychosomatic Clinic,

19

Birkenweg 10, D-24576 Bad Bramstedt, Germany

20

Elisabeth Rauh,

21

Schön-Klinik Bad Staffelstein, Psychosomatic Clinic,

22

Am Kurpark 11, D-96231 Bad Staffelstein, Germany

Predictors of treatment outcomes in BPD

2

23 Robert Doerr,

24 Schön-Klinik Berchtesgadener Land, Psychosomatic Clinic,

25 Malterhöh 1, D-83471 Schönau am Königssee, Germany

26 Winfried Rief,

27 Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,

28 Gutenbergstraße 18, D-35032 Marburg, Germany,

29 and Eva-Lotta Brakemeier,

30 Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,

31 Gutenbergstraße 18, D-35032 Marburg, Germany

32 \*Correspondence can be addressed to: Philipp Herzog, Philipps-University of Marburg,  
33 Department of Clinical Psychology and Psychotherapy, Gutenbergstraße 18, D-35032  
34 Marburg, Germany, Email: philipp.herzog@staff.uni-marburg.de, Telephone: +49 6421  
35 2824061

36 **Short title:**

37 Effectiveness and outcome predictors of inpatient treatment in BPD

#### 38 **Acknowledgments**

39 The authors would like to thank the reviewers for their valuable and helpful feedback on  
40 this article.

#### 41 **Contributions**

42 All authors materially participated in the research and/or article preparation. All authors  
43 have approved the final article.

#### 44 **Funding**

45 The authors received no funding from an external source.

46

**Highlights**

- 47   ▪ The naturalistic effectiveness of a DBT program for severely impaired BPD inpatients is  
48    small to moderate.
- 49   ▪ Higher education and comorbid recurrent depressive disorder appear to be the main  
50    predictors of treatment completion.
- 51   ▪ Lower age, absence of work disability, high perceived emotional and physical role  
52    limitations and low reported bodily pain may positively predict functioning-related change  
53    in addition to baseline functioning.
- 54   ▪ The applied variable selection procedure replicates previous research findings on  
55    predictors, but its validation does not appear to be stable.
- 56   ▪ The identification of stable predictors of treatment outcomes in BPD remains difficult.  
57    Tailoring DBT to individual BPD patients should therefore not only be based on  
58    demographic and clinical baseline data.

59

**Abstract**

60 *Background:* A routinely collected big data set was analyzed to determine the effectiveness of  
61 naturalistic inpatient treatment and to identify predictors of treatment outcome and  
62 discontinuation.

63 *Methods:* The sample included 878 patients with borderline personality disorder who received  
64 non-manualized dialectic behavioral therapy in a psychosomatic clinic. Effect sizes (Hedge's  
65  $g$ ) were calculated to determine effectiveness. A bootstrap-enhanced regularized regression  
66 with 91 potential predictors was used to identify stable predictors of residualized symptom- and  
67 functional change and treatment discontinuation. Results were validated in a holdout sample  
68 and repeated cross validation.

69 *Results:* Effect sizes were small to medium ( $g=.28-.51$ ). Positive symptom-related outcome was  
70 predicted by low affect regulation skills and no previous outpatient psychotherapy. Lower age,  
71 absence of work disability, high emotional and physical role limitations and low bodily pain  
72 were associated with greater improvement in functional outcome. Higher education and  
73 comorbid recurrent depressive disorder were the main predictors of treatment completion. The  
74 predictive quality of the models varied, with the best being found for symptom-related outcome  
75 ( $R^2=18\%$ ).

76 *Conclusion:* While the exploratory process of variable selection replicates previous findings,  
77 the validation results suggest that tailoring treatment to the individual patient might not be based  
78 solely on sociodemographic, clinical and psychological baseline data.

79

80 **Keywords:** Borderline personality disorder; effectiveness; predictors; treatment outcome;

81 dropout; premature treatment discontinuation

82 **Introduction**

83 BPD<sup>1</sup> is a serious mental disorder, affecting around 0.7% of the general population  
84 (Coid, Yang, Tyrer, Roberts, & Ullrich, 2006; Torgersen, Kringlen, & Cramer, 2001). In  
85 psychiatric inpatient settings, about 15 to 25% of all patients are diagnosed with BPD  
86 (Leichsenring, Leibing, Kruse, New, & Leweke, 2011). BPD is characterized by a  
87 malfunction of affect regulation, a high risk of suicide, a considerable burden of psychosocial  
88 distress and a high number of psychopathological comorbidities, especially mood disorders,  
89 substance abuse, anxiety and other personality disorders (Leichsenring et al., 2011). Although  
90 BPD is a personality disorder with significant heritability (New, Goodman, Triebwasser, &  
91 Siever, 2008), studies have shown that the disorder is remitted in a considerable number of  
92 cases after more than 10 years (Zanarini, Frankenburg, Hennen, Reich, & Silk, 2006). BPD  
93 has a negative impact on patients' lives, society and the health care system (Meuldijk,  
94 McCarthy, Bourke, & Grenyer, 2017; Soeteman, Hakkaart-van Roijen, Verheul, &  
95 Busschbach, 2008).

96 Psychotherapeutic approaches have proven to be effective in the treatment of BDP  
97 patients, in particular DBT<sup>2</sup> (Cristea et al., 2017; Kliem, Kröger, & Kosfelder, 2010; Linehan,  
98 Armstrong, Suarez, Allmon, & Heard, 1991; Linehan et al., 2015; McCauley et al., 2018;  
99 Mehlum et al., 2016; Ougrin, Tranah, Stahl, Moran, & Asarnow, 2015; Panos, Jackson,  
100 Hasan, & Panos, 2014; Priebe et al., 2012). A recent meta-analysis concluded that DBT  
101 reduces non-suicidal self-injury (NSSI) and suicidal behavior, with effect sizes mostly in the  
102 medium range (DeCou, Comtois, & Landes, 2019). This conclusion must be treated with  
103 caution, however, as most studies do not include an active borderline-specific control group  
104 (Kliem et al., 2010).

---

<sup>1</sup> Borderline personality disorder

<sup>2</sup> Dialectic behavior therapy

## Predictors of treatment outcomes in BPD

6

105 DBT was initially designed as an outpatient treatment and is intended to reduce the  
106 high rates of treatment dropout and suicidality in BPD patients (Linehan, 1993; Swenson,  
107 Sanderson, Dulit, & Linehan, 2001). Later, DBT was adapted as a multidisciplinary concept  
108 for inpatient settings (Bohus et al., 2004). Within the framework of DBT, patients learn  
109 practical emotion regulation skills, ways of dealing with interpersonal problems and coping  
110 with distress and, in the case of inpatients, skills to help them avoid or minimize  
111 hospitalization (Swenson et al., 2001). To date, there are only a few small studies that have  
112 investigated the extent to which DBT provides specific results in routine care under normal  
113 everyday conditions (Higgins & Green, 2011). By determining pre-post effect sizes of  
114 different outcome measures in a large naturalistic data set, we want to address this research  
115 gap.

116 In order to optimize therapy and better adapt it to the needs and problems of each  
117 patient, predictor analyses appear helpful. Therefore, in this large data set, we try to identify  
118 variables that predict the outcome of treatment. Because treatment dropout rates for BPD  
119 patients are high – only 75% (95% CI: 68–82%) completion for interventions of <12 months  
120 duration (Barnicot, Katsakou, Marougka, & Priebe, 2011) –, some authors have argued that  
121 DBT studies should use both dropout rates and quality of life indicators as outcome variables  
122 (Panos et al., 2014). However, most published studies only report, beside the primary  
123 outcome of non-suicidal self-injury (NSSI), measurements of quality of life (Bos, van Wel,  
124 Appelo, & Verbraak, 2010; Carter, Willcox, Lewin, Conrad, & Bendit, 2010).

125 Dropout is associated with negative effects on the patient, such as worse treatment  
126 outcomes, as well as negative effects on the morale of clinical staff and other patients  
127 (McMurrin, Huband, & Overton, 2010; Priebe et al., 2012). Kröger, Roepke, and Kliem  
128 (2014) argued that dropout might also reduce motivation to start a new treatment afterwards.  
129 Identifying factors associated with treatment discontinuation might lead to the development of



## Predictors of treatment outcomes in BPD

7

130 treatments that maximize sustained treatment outcome (Barnicot et al., 2011; McMurrin et  
131 al., 2010). Based on these findings, three different outcomes are investigated in this study:  
132 symptomatic and functional changes as well as early termination of therapy.

133 *Predictors of treatment non-completion*<sup>3</sup>

134 Previous studies have reported that experiential avoidance, poor problem solving  
135 skills, impulsiveness and lack of commitment to change were associated with a higher risk of  
136 dropout (De Panfilis et al., 2012; McMurrin et al., 2010; Rüscher et al., 2008). One meta-  
137 analysis concluded that sociodemographic variables are not associated with risk of dropout  
138 (Barnicot et al., 2011), whereas another concluded that age, educational attainment and  
139 occupational status were negative predictors of dropout (McMurrin et al., 2010). These  
140 studies share the problem of a small sample size, whilst Kröger, Harbeck, Armbrust, and  
141 Kliem (2013) examined treatment discontinuation in a large inpatient DBT sample and found  
142 that dropout was predicted by younger age and substance use disorder.

143 It should be noted that there is variation in the definitions of dropout used in these  
144 studies (McMurrin et al., 2010). Kröger et al. (2014) investigated the reasons for treatment  
145 discontinuation in DBT and found different predictors for exclusion and patient initiated  
146 dropout, whilst Wierzbicki and Pekarik (1993) argued that whether treatment discontinuation  
147 is recorded as patient- or institution-initiated may depend on the practitioner, and that a more  
148 objective approach would be preferable. All in all, a clear definition and valid predictors of  
149 dropout are still lacking.

150

151

---

<sup>3</sup> McMurrin et al. (2010) used the term non-completion to refer to dropout, discontinuation, premature termination and therapy attrition and notes that there has been inconsistency in the use of the terms 'dropout' and 'premature) treatment discontinuation'. We have followed current practice in BPD research and used the term '(premature) treatment discontinuation' in this article.

152 *Predictors of treatment outcome*

153 A meta-analysis of studies of treatment outcomes found that only a few outcomes  
154 were consistently observed (Barnicot et al., 2012). However, the severity of the symptoms  
155 consistently predicted individual treatment effects. In most studies, pretreatment symptom  
156 severity was a positive predictor of treatment effects. Barnicot et al. (2012) attributed this to  
157 regression to the mean (Bland & Altman, 1994) – a statistical mechanism that produces  
158 greater effects of the baseline symptom severity the further an individual score is from the  
159 baseline population mean. Another possible explanation for the effect of symptom severity is  
160 that floor and ceiling effects of the outcome measures may be responsible for the findings -  
161 similar to the prediction of minimally clinically important improvement measures (Ward,  
162 Guthrie, & Alba, 2014). Since it is unclear to what extent a statistical effect can be interpreted  
163 as an indication of clinically meaningful change, this paper focuses on the additional  
164 influence of predictors other than the baseline symptom severity on treatment outcome.

165 Investigation of predictors of (premature) treatment discontinuation may lead to  
166 information that can be used to develop strategies to keep patients in treatment and thus  
167 reduce the severity and chronicity of BPD. As BPD is considered a heterogeneous disorder  
168 (Leichsenring et al., 2011), the important psychotherapy research question which patient  
169 benefit from which treatment strategies (Kazdin, 2007; Paul, 1967; Roth & Fonagy, 2005) is  
170 important to tailor treatment to the specific needs and problems of the individual BPD patient.  
171 To address this issue, generating complex models with clinical value and some global  
172 predictive value are highly auspicious (Steyerberg, 2019). Yet, instead of building such a  
173 black box machine learning model with a high predictive value which can be, for example,  
174 applied into a specific app, but would not contribute to an understanding of the processes  
175 involved, we expect to develop a clinical model with stable predictors by using novel  
176 statistical approaches (such as machine learning algorithms) to foster transparent research.

## Predictors of treatment outcomes in BPD

9

177 Therefore, in the absence of comprehensive models for predicting (premature) treatment  
178 discontinuation or treatment outcome, we used an exploratory, data-driven approach to  
179 identify possible predictors. Following this line and the argument by Yarkoni and Westfall  
180 (2017), we used a prediction-based rather than an explanation-based approach in order to  
181 improve the interpretability of the results, to prevent over-fitting of the data on the basis of  
182 significance tests in a large dataset and to facilitate the translation of research findings into  
183 practical guidelines. As machine learning algorithms are a particularly good way of selecting  
184 a subset of important predictors in the context of feature selection, the goal of this study was  
185 to identify stable predictors for (premature) treatment discontinuation and treatment outcome  
186 using such an exploratory, stability selection-based machine learning approach.

187 To sum, the first aim of our study was to determine the effectiveness of the inpatient  
188 DBT in routine clinical practice. The secondary aim was to develop a clinical predictive  
189 model by identifying patient characteristics (sociodemographic, clinical, and psychological  
190 variables) that predicted (1) a (premature) treatment discontinuation and (2) symptom-specific  
191 as well as functional response in completers in a large naturalistic sample of BPD patients  
192 treated with non-manualized DBT in a German psychosomatic clinic. Whenever appropriate  
193 in our naturalistic study, we applied the requirements of the established TRIPOD statement  
194 paper (Moons et al., 2015). However, this was not always possible, as the resampling  
195 technique used is not explicitly covered in the TRIPOD paper.

196

**Methods****Sample**

198 For this study, we were able to analyze data routinely collected over 5 years from BPD  
199 patients treated in a large German psychosomatic clinic. The patients completed various self-  
200 report questionnaires at admission and at discharge and were interviewed by raters. All  
201 patients gave informed consent to anonymous evaluations of their routinely collected data.  
202 Patients were included in this analyses if they had been diagnosed with BPD (F60.30 or  
203 F60.31) and scored at least five in the SCID-II borderline scale (First, Gibbon, Spitzer, &  
204 Benjamin, 1997). In addition to the SCID-II cutoff in the BPD section, the reliability of the  
205 diagnosis of BPD was further assured by a trained therapist and reassured by the medical head  
206 of the ward. Patients who discontinued treatment for an unknown reason or for medical  
207 reasons not related to BPD were excluded, as were those aged over 65 years or under 18  
208 years. Out of 1.087 patients diagnosed with BPD, 878 met these inclusion and exclusion  
209 criteria. Thus, data from 878 BPD patients could be used in the analyses of the effectiveness  
210 of DBT and predictors of treatment discontinuation. For the analyses with the continuous  
211 outcome variables (e.g., Borderline Symptom List 95, BSL-95, Bohus et al., (2001); Short  
212 Form-36, SF-36, Ware & Sherbourne (1992)) all patients with completely missing values or  
213 outliers had to be excluded (after preprocessing but before imputation). Thus, data from  $N =$   
214 702 (BSL-95) and  $N = 703$  (SF-36) patients were available to compute change scores. The  
215 patient flow diagram is shown in *Figure 1*.

216

--Insert Figure 1 here.--

217 The mean age of the 878 BPD patients was 31.1 years ( $SD = 10.3$ ). 81.2% of the  
218 sample was female, 47.7% were in a relationship and 16% were married. Patients had an  
219 average of 3.5 psychiatric diagnoses. 86.4% of the patient had recurrent depressive disorder as  
220 a comorbidity, 32.5% had a comorbid eating disorder, 21.8% specific anxiety disorder, 17%

221 post-traumatic stress disorder and 14.8% a substance abuse disorder. The mean baseline BSL-  
222 95 score was relatively high ( $M = 2.1$ ,  $SD = 0.7$ ) and the mean GAF baseline score relatively  
223 low ( $M = 42.8$ ,  $SD = 25.4$ ), indicating severe borderline-specific symptoms and moderate to  
224 high impairment in psychosocial adaptation. Consistent with this picture, 25.3% of the  
225 patients had impaired working ability and 34.9% were unemployed. Although 62% of patients  
226 were experiencing their first inpatient treatment, the mean number of previous inpatient  
227 treatments was 2.75 ( $SD = 3.0$ ).

228 Practitioner reports indicated that 22.9% of the patients left treatment ahead of the  
229 scheduled end of the treatment program. Of these, approximately equal portions were  
230 prematurely discontinuations initiated by the patient, discharges by the team, discharge by  
231 mutual agreement and transfers. If only discontinuations that were not agreed are counted, the  
232 dropout rate was 12.2%. In the group patients discharged as planned the mean length of stay  
233 was 66 days ( $SD = 12$  days), so the cutoff for the premature treatment discontinuation  
234 outcome was set to 54 days (see paragraph below *Premature treatment discontinuation* for  
235 further details), which meant that 17.6 % of patients were classified as having discontinued  
236 treatment prematurely.

237 Table 1 presents the sociodemographic and clinical characteristics of the sample.

238 --Insert Table 1 here.--

### 239 **Routine clinical inpatient treatment**

240 The clinic offers a multimodal, non-manualized therapy for BPD patients based on  
241 Linehan's DBT manual (Linehan, 1993; Swenson et al., 2001). A prototypical treatment plan  
242 of the inpatient DBT program comprises the following treatment elements: DBT group  
243 therapy (skills training for 3,5h/week), individual psychotherapy (1h/week), individual co-  
244 therapy (0,5h/week) provided by qualified nursing staff, psychoeducation (basis group

## Predictors of treatment outcomes in BPD

12

245 therapy 1h/week) in strategies for managing distress, emotion regulation and interpersonal  
246 effectiveness, reference groups (daily accompanied morning rounds, weekly interactional  
247 ward group, weekly exercise group, 4h/week), mindfulness group (1h/week), art therapy  
248 group (1,5h/week), body therapy group (1,5h/week) as well as therapist team consultation  
249 meetings (5h/week) and twice a week a senior visit. All DBT-based groups (e.g., skills  
250 training), individual psychotherapy and team meetings including supervision are based on the  
251 DBT manual according to Linehan (Linehan, 1993). Per year, a minimum of 8 hours of team  
252 training is provided by an officially approved external DBT certified supervisor. In addition,  
253 about 2 hours a day are reserved for the preparation and follow-up of the therapy sessions  
254 (behaviour analyses, homework, independent skill training) for the patients. Therefore,  
255 patients received at least 23 hours of therapy over 8-12 weeks (leading to a total of 180-270  
256 therapy hours). For the patient-staff ratio to be proven for DBT certification, two trained DBT  
257 therapists are available (one medical doctor and one certified clinical psychologist), one DBT-  
258 certified therapist in training, and at least 2 DBT-certified nursing personal, through which the  
259 adherence to the traditional DBT manual can be guaranteed. During the weekends, emergency  
260 care of the clinic with medical and nursing personal is warranted, till 10 pm also with  
261 psychological personal.

262         Psychopharmacological treatment was administered when indicated, mainly to treat  
263 psychiatric comorbidities according to the current national and international guidelines.  
264 Nevertheless, the major treatment focus lied on psychotherapeutic work in single and group  
265 settings; medication was therefore continued or reduced to former psychopharmacological  
266 treatment before intake. E.g., a subanalysis revealed that across all involved clinics in 2017,  
267 73% of all patients with the main diagnosis of BPD received antidepressant medication and  
268 61% neuroleptics at discharge.

269 **Assessments**

270 In the big data set, several instruments assessed treatment outcome and were  
271 administered both at admission and at discharge. The instruments we used for our  
272 effectiveness analyses are described below.

273 *Instruments for assessing the effectiveness*274 Primary treatment outcome

275 The BSL-95 (Bohus et al., 2001) was chosen as the primary outcome (also used by  
276 Steil, Dyer, Priebe, Kleindienst, & Bohus, 2011). The BSL-95 is a BPD-specific 95-item, self-  
277 report instrument based on the DSM-IV criteria and is widely used in research on BPD  
278 because of its good internal consistency (Cronbach's  $\alpha = .79$ ), one week retest-reliability ( $r =$   
279  $.84$ ) and sensitivity to change (Bohus et al., 2007). It is used to measure typical BPD  
280 symptoms in the week prior to completion (Bohus et al., 2007) and consists of seven  
281 subscales: self-perception (19 items), affect regulation (13 items), self-destruction (12 items),  
282 dysphoria (10 items), loneliness (12 items), intrusions (11 items) and hostility (6 items).  
283 Responses are given using a five-point Likert scale to indicate the severity of the symptoms  
284 described in each item.

285 Secondary treatment outcomes

286 As treatments for BPD affect quality of life as well as BPD-specific symptoms (Panos  
287 et al., 2014), and quality of life tends to be very low in BPD patients (IsHak et al., 2013), we  
288 used the Medical Outcomes 36-item Short-Form Health Survey (Ware & Sherbourne, 1992)  
289 as a secondary outcome. The Short Form-36 Health Survey (SF-36) is a self-rated  
290 questionnaire assessing perceived health-related quality of life or health status (Ware, Snow,  
291 Kosinski, & Gandek, 1993) in eight domains: physical functioning (PF), role-physical (RP),  
292 bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role-emotional

## Predictors of treatment outcomes in BPD

14

293 (RE), and mental health (MH). The SF-36 has shown excellent internal consistency ( $\alpha > .80$ ),  
294 adequate test-retest-reliability and several forms of validity across studies (e.g., Jenkinson,  
295 Wright, & Coulter, 1994).

296 In order to determine the effectiveness on other outcomes, we also used data of the the  
297 following instruments: the Global Severity Index (GSI from the Brief Symptom Inventory, cf.  
298 Davidson et al., 2006; Doering et al., 2010) as a commonly used measure of overall  
299 psychopathology, the Beck Depression Inventory-II (BDI-II) measuring depressive  
300 symptoms, a self-report questionnaire also used by Steil et al. (2011). Overall functioning was  
301 measured with the therapist-administered Global Assessment of Functioning (GAF) (cf.  
302 Bateman & Fonagy, 2009; Doering et al., 2010; Farrell, Shaw, & Webber, 2009).

**303 *Instruments used for the predictor analyses***

304 The BSL-95 score (disorder-specific outcome, see above), SF-36 score (functional  
305 outcome, see above) and premature treatment discontinuation, were used as dependent  
306 variables in analyses of potential predictors.

**307 Premature treatment discontinuation**

308 Premature treatment discontinuation data were retrieved from medical records. Reason  
309 for discharge data were available for all subjects. In order to construct an objective outcome  
310 independent of the therapists interpretation (by avoiding too large class imbalance), four  
311 practitioner-defined causes of irregular treatment discontinuation (patient induced,  
312 disciplinary, transfers, and discontinuation by mutual agreement) were combined and  
313 restricted in the maximum length of stay to a value of one standard deviation shorter than the  
314 average regular treatment duration as a cutoff (comp. Table 1).

315

316



317 *Candidate predictors*

318 Candidate predictors were selected from the pool of data routinely gathered about the  
319 patients. Binary or composite score variables were constructed from clinical and demographic  
320 categorical variables in order to ensure that variables were interpretable when they were  
321 selected in the variable selection process as an individual predictor. Variables with less than  
322 20% missing values were included in the analyses. Binary variables with severe class  
323 imbalance ( $\geq 90\%$  in one category) were excluded to enhance stability in the variable  
324 selection procedure. In order to make use of the information provided by ICD-10 codes  
325 concerning both somatic comorbidities as well as psychopathological comorbidities, dummy  
326 variables were created for all diagnostic labels assigned to between 10% and 90% of the  
327 patients. If a diagnosis and a single sub-diagnosis from a given branch of the ICD-10 were  
328 selected, only the sub-diagnosis was used in order to avoid highly correlated variables.

329 In addition to the instruments described above, which we use to determine  
330 effectiveness, a number of other routinely collected instruments were available in the data set.  
331 The following questionnaires had less than 20% missing values and were thus included as  
332 candidate predictors in this analysis: the Questionnaire for Thoughts and Feelings-37 (QTF-  
333 37; Renneberg, Schmidt-Rathjens, Hippin, Backenstrass, & Fydrich, 2005; Renneberg &  
334 Seehausen, 2010), the Impulsivity Scale-27 (IS-27; Kröger, Holdstein, Lombe, Schweiger, &  
335 Kosfelder, 2007; Kröger et al., 2010), the Patient Health Questionnaire (PHQ-D; Gräfe,  
336 Zipfel, Herzog, & Löwe, 2004) and the Brief Symptom Inventory (BSI; Derogatis &  
337 Melisaratos, 1983). The QTF-37 assesses borderline-specific cognitions and assumptions and  
338 includes a subscale that measures contradictory cognitions. The IS-27 measures BPD-related  
339 impulsive behavior over the past month. The PHQ-D is organized into three subscales: the  
340 PHQ-9 measures depression, the GAD-7 measures generalized anxiety disorder and the PHQ-

## Predictors of treatment outcomes in BPD

16

341 15 measures somatic symptoms. The BSI measures psychological distress in terms of nine  
342 subscale scores and two global scores.

343 A total of 43 questionnaires or subscales could be included as predictors in the  
344 subsequent analysis, additionally 11 demographic variables, 8 clinical variables, 10 variables  
345 identifying psychopathological comorbidities and 9 variables identifying somatic  
346 comorbidities.

**347 Statistical analyses**

348 All statistical analyses were computed in the statistical processing language R (R Core  
349 Team, 2018).

**350 *Statistical methods for calculating effectiveness***

351 Corrected Hedge's  $g$  pre-post effect sizes for all outcomes (see above) were calculated  
352 to assess the effectiveness of inpatient DBT. To put an upper limit on the potential impact of  
353 treatment discontinuation on effect sizes we computed last observation carried forward  
354 (LOCF) effect sizes.

**355 *Statistical methods for the identification of predictors***

356 The following analyzing steps were computed both on the completer dataset and in  
357 additional subgroups. The selection of the subgroups was based on the largest diagnostic  
358 groups to retain a high power; namely, BPD inpatients with comorbid recurrent depressive  
359 disorder ( $n = 759$ ), with/without comorbid anorexia nervosa ( $n = 285$  vs.  $593$ ), female vs.  
360 male gender ( $n = 713$  vs.  $165$ ) and first inpatient treatment ( $n = 544$ ) vs. previous inpatient  
361 treatments ( $n = 334$ ).

**362 *Preprocessing***

363 Univariate outlier values were removed by a  $p < .001$  criterion (Tabachnick & Fidell,  
364 2013) adapted to the large sample size on the  $z$  distribution for all continuous variables, post

365 treatment values of the outcomes, and change scores. Time-event data were log-transformed.  
366 Missing data were imputed in the data set of predictors using the random forest procedure,  
367 implemented in the R package missForest (Stekhoven & Bühlmann, 2012), a nonparametric  
368 imputation algorithm based on random forests that results in lower imputation and prediction  
369 errors than other commonly used imputation techniques (Waljee et al., 2013). The data were  
370 split randomly into a training set (80%) and a testing set (20%).

### 371 Variable selection

372 As ordinary multiple regression models are not able to deal well with a high number  
373 of correlated variables, we used regularized regression as a basic algorithm (Helwig, 2017). In  
374 regularized regression, a penalty term is added to each coefficient in order to avoid overfitting  
375 to the data. We used elastic net with the package glmnet.cv (Friedman, Hastie, & Tibshirani,  
376 2010) with  $\alpha = 0.5$  in the variable selection process, as it is able to shrink coefficients  
377 exactly to zero and thus perform implicit variable selection. Because it shrinks the coefficients  
378 of highly correlated variables towards each other, it also performs favorably in the presence of  
379 multicollinearity.

380 To control for regression to the mean, for continuous variables a variant further called  
381 residual regularized regression was used, in which the outcome in test and training set was  
382 first predicted via cv.glmnet using the pretest score of the outcome. A second call of  
383 cv.glmnet used additional variables to predict the outcome in test and training set using the  
384 predictions of the first call as an offset. This approach resembles the computation of  
385 residualized change scores in each subsample.

386 Since the main focus of this work is to identify from the multitude of possible  
387 predictors those that can be reproduced with high probability, variables were selected in the  
388 training set using the scheme of bootstrap-enhanced regularized regression (Abram et al.,  
389 2016; Park & Hastie, 2008): Residual elastic net models with an parameter of  $\alpha = 0.5$

390 were trained via the function `glmnet.cv` (Friedman et al., 2010) on 1.000 nonparametric  
391 bootstrap samples of the training set. The variable inclusion probability for a given variable is  
392 the percentage of samples in which that variable is selected as a predictor in a model  
393 (Meinshausen & Bühlmann, 2010). A consistency score was also computed, this indicates the  
394 percentage of times that the coefficient for a variable was positive when that variable was  
395 selected as a predictor, as the coefficients of false positive predictors vary around zero  
396 (Abram et al., 2016).

397 In an attempt to reduce the negative impact of class imbalance on the dichotomous  
398 outcome treatment discontinuation, weights were passed to `glmnet` such that the accumulated  
399 weights of discontinuing patients and completers were equal in both classes of the outcome  
400 variable.

401 For the symptom-related and functional-related outcome, five models with different  
402 sets of variables were constructed: M0 consisting of only the baseline value, M1 additionally  
403 containing predictors with a variable inclusion probability of at least 95%, M2 using a less  
404 strict cutoff of 85%, M3 including all variables and M4 containing also nonlinear  
405 relationships. For the treatment discontinuation outcome, M0 is not applicable, but all other  
406 four models are (comp. Tab. 3). To test for nonlinear relationships, we used a novel statistical  
407 technique `bartMachine` (Kapelner & Bleich, 2016), which has recently been used for the  
408 similar purpose (Cohen, Kim, Van, Dekker, & Driessen, 2019). Predictors were considered  
409 significant and interpretable only if they were included in M1 or showed consistently signed  
410 effects on a self-report measure and treatment discontinuation (comp. Table 2). Whenever a  
411 variable that was included in a model for the prediction of treatment discontinuation had a  
412 contrary effect on change in a self-report measure, it was only interpreted in relation to  
413 treatment discontinuation, because the effect on treatment outcome might be due to dropout  
414 effects.

415 *Validation*

416           The quality of the selection process was determined using a five-fold cross validation  
417 paradigm (Stone, 1974), applied to the full data set and repeated 100 times in order to assess  
418 the stability of prediction results for all models. Additionally, single prediction models were  
419 trained on the training set and used to predict outcomes in the testing set, which was not used  
420 in the variable selection process, thus giving information about the generalizability of the  
421 results.

422           In the validation step, ridge regression (Hoerl & Kennard, 1988) was used instead of  
423 elastic net, as its properties in the estimation of coefficients seem to be favorable (Abram et  
424 al., 2016). The ability of models to predict continuous outcomes was measured with residual  
425 ridge regression predictions using three indicators of fit: R squared, root mean squared error  
426 (RMSE) and mean average error (MAE). The quality of predictions of the dichotomous  
427 outcome was measured as the area under the receiver operating curve (AUC) with the R  
428 package pROC (Robin et al., 2011). An AUC value of one indicates perfect prediction of both  
429 classes, whilst an AUC value of 0.5 represents performance equivalent to random guessing.  
430 DeLong confidence intervals (DeLong, DeLong, & Clarke-Pearson, 1988) for the AUC were  
431 computed via ci.auc. A threshold established in the context of depression of  $R^2=6.3\%$  for  
432 continuous measures (Uher, Tansey, Malki, & Perlis, 2012) and an AUC of .67 was used to  
433 assess clinical significance, respectively.

434

**Results****435 Effectiveness of the inpatient treatment**

436 Hedge's  $g$  for change in BSL-95 score, the primary outcome, was .35 (95% CI [.24,  
437 .45]) and the LOCF-corrected effective size was .31 (95% CI [.22, .40]). The values of  
438 Hedge's  $g$  for the various BSL-95 subscales were, in ascending order, .04 (95% CI [-.07, .14])  
439 for hostility, .06 (95% CI [-.04, .17]) for intrusions, .23 (95% CI [.12, .33]) for self-  
440 destruction, .25 (95% CI [.14, .35]) for affect regulation, .36 (95% CI [.26, .47]) for self-  
441 perception, .38 (95% CI [.27, .48]) for loneliness and .49 (95% CI [.39, .60]) for dysphoria.

442 Hedge's  $g$  for the secondary outcome change in SF-36 score (measure of functioning)  
443 was .31 (95% CI [.20, .42]) and the LOCF-corrected effect size was .28 (95% CI [.18, .37]).

444 Hedge's  $g$  values for the changes in other treatment outcome variables were as  
445 follows: depressive symptoms,  $g = .51$  (95% CI [.41, .61]), LOCF-corrected effect size = .46  
446 (95% CI [.37, .56]); GSI of the BSI  $g = .39$  (95% CI [.28, .49]), LOCF-corrected effect size =  
447 .34 (95% CI [.25, .44]); GAF  $g = .31$  (95% CI [.21, .40]), LOCF-corrected effect size = .31  
448 (95% CI [.21, .40]).

**449 Predictor analysis****450 Variable selection for predictors**

451 The missForest imputation converged with a normalized RMSE of .38 in the case of  
452 continuous variables and 2% falsely classified entries in the case of categorical variables. The  
453 predictors that were included in a model or had a consistent sign in at least two measures are  
454 given in *Table 2*.

455 In the case of residualized symptom change (BSL-95), two predictors were included in  
456 the smaller model M1 (selected in at least 95% of the models) and eight in the larger model  
457 M2 (selected in at least 85% of the models). In M1, high affect regulation deficiencies at

## Predictors of treatment outcomes in BPD

21

458 intake and the absence of a history of outpatient psychotherapy were positively associated  
459 with treatment gains (see Tab. 2).

460         Regarding functional outcome (SF-36) in the smaller model M1, lower age, the  
461 absence of a work disability, high perceived emotional and physical role limitations and low  
462 reported bodily pain were consistently associated with greater improvement in functional  
463 outcome. Five more predictors were included in the larger model M2 of SF-36 score (see Tab.  
464 2).

465         Concerning premature treatment discontinuation (PTD), a higher level of education  
466 and a comorbid severe, recurrent depressive disorder were found to be stable predictors of  
467 treatment completion in the smaller model M1. Additionally, in the larger model M2 two  
468 more variables were included: the presence of substance abuse predicted negatively and  
469 higher age positively treatment completion.

470         The following variables predicted both outcome and treatment completion: Less  
471 severe depression symptoms, lower obsession-compulsion score, fewer intrusions, less  
472 somatization, less self-reported pain, and more perceived emotional role limitations were  
473 associated with both higher treatment gains and increased probability of treatment  
474 completion. The same was true for being diagnosed with an endocrine, metabolic, or  
475 nutritional disease according to ICD-10 (Section E) criteria. A re-analysis of subgroups  
476 showed that 62% of these patients fulfilled the criteria for an obesity diagnosis (Section E66).  
477 A lower level of pain was the only consistent positive predictor of all three outcomes.

478   --Insert Table 2 here.--

479         The subgroup analyses revealed that the underlying patterns of predictor profiles  
480 showed similarities but also some differences, especially for symptom-specific changes and  
481 PTD: an avoidant personality style was negatively associated with treatment gains in BSL-95  
482 for those with comorbid recurrent depression, while living alone was a positive predictor for

483 changes in BSL-95 for female BPD inpatients. Of note, medical diagnoses played a  
484 particularly important role for treatment completion: The presence of a disorder of the thyroid  
485 gland (ICD-10: E0) was positively associated with treatment completion in those with  
486 comorbid recurrent depression; being diagnosed with an endocrine, metabolic, nutritional  
487 disease (ICD-10: Section E) or a diagnosis of a dorsopathy (ICD-10: M5) was a positive  
488 predictor for treatment completion in those with anorexia nervosa; an asthma diagnosis (ICD-  
489 10: J45) was positively linked to treatment completion for female BPD inpatients. For  
490 functional changes in SF-36, a comorbid PTSD diagnosis was negatively associated with  
491 treatment gains for those with anorexia nervosa, and an antisocial personality style was  
492 negatively related to treatment outcome for male BPD inpatients. The identified predictors for  
493 functional changes (physical role limitations, working disability, bodily pain) were  
494 consistently found across subgroups and therefore seemed to be quite stable. The predictors  
495 for all subgroups that were included in the M1 model are displayed in Table 4 (see  
496 supplemental material 1).

#### 497 *Validation*

498 In Table 3, the results concerning the ability of the five models (M0, M1, M2, M3,  
499 M4) to predict the three different outcomes in the full sample and holdout sample are  
500 depicted.

501 Concerning the baseline model M0, baseline BSL-95 score explained on average  
502 13.6% (SD = 4.3%) of the variance in cross-validation and 9.4% in the holdout sample, whilst  
503 baseline SF-36 score explained on average 6.5% (SD = 3.2%) in cross-validation and 8.5% in  
504 the holdout sample. Greater symptom severity at baseline consistently predicted higher  
505 treatment gains.

506 All three 85% VIP models (M2) and the 95% VIP model (M1) for SF-36 score  
507 achieved better predictive performance than the full model (M3) in the cross-validation on the



## Predictors of treatment outcomes in BPD

23

508 full set, which indicates that our variable selection procedure was successful. The functional  
509 outcome models and the 85% VIP model (M2) for treatment discontinuation performed  
510 similarly to the corresponding full models in the holdout sample, whereas the BLS-95 models  
511 accounted for negligible outcome variance in the holdout sample.

512         If the total combined explained variance is considered (adding M0 with M1 or M2 in  
513 the five-fold cross validation), then the combined explained variance in symptom-specific  
514 outcome in the holdout sample is 10.1% (M0 + M1), in functional outcome 14.9% (M0 +  
515 M1), both reaching the clinical significance threshold of an  $R^2$  of 6.3%.

516         If only the additional predictive power is considered (M1 or M2), then the predictive  
517 power in in the symptom-specific outcome did not reach the clinical significance threshold of  
518 an  $R^2$  of 6.3% (0.6% in M1 and 4.2% in M2 in the cross-validation sample and 0.7% in M1  
519 and 0.6% in M2 in the cross-validation sample), whereas the additive predictive power of the  
520 SF-36 model over the baseline model M0 was close to or above this threshold in the cross-  
521 validation (9% in M1 and 10.6% in M2) and holdout sample (6.4% in M1 and 5.4% in M2).  
522 None of the models for the prediction of premature treatment discontinuation reached the  
523 clinical significance threshold, but the confidence intervals in AUC for the 85% VIP model  
524 (M2) did not include 0.5 in cross validation [.53, .75] or the holdout sample [.53, .73],  
525 indicating a predictive power greater than random guessing. Of note, the results of the M4  
526 model were generally not superior to other models, indicating that nonlinear relationships do  
527 not appear to play an important role in our sample.

528                                 --Insert Table 3 here--

529         The results for all subgroups concerning the ability of the five models (M0, M1, M2,  
530 M3, M4) to predict the three different outcomes are depicted in Table 5 (see supplemental  
531 material 2). Overall, the results indicate that the different models of the subgroups do not  
532 reliably perform better in explaining the outcome variance in symptom-specific changes

## Predictors of treatment outcomes in BPD

24

533 (BSL-95) and also PTD, while the prediction quality for functional changes (SF-36) is  
534 somewhat better. For the BSL-95, there is a trend for the subgroup of those BPD inpatients  
535 without anorexia nervosa that the M1 model with the additional predictor affect regulation  
536 difficulties explains more outcome variance than M0, yet failing to reach the clinical  
537 significance cutoff.

538

**Discussion**

539 An important aim in practice-oriented research is that therapists could base their  
540 decision about which patient receives which treatment on scientifically obtained criteria, which  
541 in turn have been obtained in practice, preferably by big data sets (Castonguay, Barkham, Lutz,  
542 & McAleavey, 2013). We have attempted to address this objective by analyzing a routinely  
543 collected big data set of 878 BPD patients to determine the effectiveness of a non-manualized  
544 DBT inpatient treatment under naturalistic conditions and to identify sociodemographic,  
545 clinical and psychological predictors of symptomatic and functional response as well as  
546 premature discontinuation of therapy. In addition, we tried to assess the practical relevance of  
547 the statistical predictions.

***Discussion of main findings concerning the determination of effectiveness***

549 First, the observed effect sizes (Hedge's  $g$ ) for the primary and secondary outcomes  
550 were mostly small, according to Cohen's criteria. With regard to effect sizes for specific  
551 facets of borderline-specific symptoms, the treatment had negligible effect on hostility and  
552 intrusions and small to medium effects on dysphoria, loneliness and self-perception. Thus, the  
553 treatment appears to have been somewhat less effective than in the study by Kröger et al.  
554 (2013), who observed medium effect sizes on average. This difference in effectiveness is  
555 probably due to the difference between the two samples, the non-manualized performance of  
556 the DBT, and the duration of the treatment: Compared with the patients in the study of Kröger  
557 et al. (2013), our sample was more severely impaired in several respects (e.g., higher mean  
558 number of psychiatric diagnoses, higher number of comorbid recurrent depressive disorders,  
559 lower general functional level at baseline) and the duration of the treatment was about 2.5  
560 weeks shorter in our study. This comparatively short inpatient non-manualized DBT treatment  
561 may not have been sufficient to fully address the complexity and heterogeneity of symptoms  
562 experienced by this disabled sample with high comorbidity rates. In particular, the large

563 patient subgroup who also suffered from recurrent depressive disorders (representing 86% of  
564 BPS patients in this sample) may benefit more from an integration of treatment strategies that  
565 more specifically addresses chronic depression, e.g., an augmentation of DBT with strategies  
566 from CBASP<sup>4</sup> – a psychotherapy developed specifically for chronic depression (McCullough,  
567 2003; McCullough, Schramm, & Penberthy, 2015).

568 Notably, in our study practitioner reports indicated that 22.9% of the patients left  
569 treatment ahead of the scheduled end of the treatment program, which is below the median  
570 dropout rate of 37% reported by McMurran et al. (2010). If only discontinuations that were  
571 not mutual agreed by the patient and the team are counted, the dropout rate was 12.2%, which  
572 is comparable to the 10% rate reported by Kröger et al. (2013).

#### 573 *Discussion of main findings concerning the identification of predictors*

574 To the best of our knowledge, this is the first study to predict both treatment outcome  
575 and (premature) treatment discontinuation in a large, inpatient BPD sample. In the following,  
576 we discuss mainly the variables that were included in the M1 model (with most strict cutoff  
577 criteria for inclusion) or showed consistently signed effects on a self-report measure and  
578 treatment discontinuation, since we consider them significant and interpretable.

579 In the case of residualized symptom change, high affect regulation deficiencies at  
580 intake and the absence of a history of outpatient psychotherapy were positively associated  
581 with treatment gains. Thus, the patients who showed the greatest improvement in BPD-  
582 specific symptoms were those with poor affect regulation skills at baseline and no history of  
583 outpatient psychotherapy. This group without therapeutic outpatient prior experience may  
584 have benefited most from the skills training that - as an effective key part of DBT - directly  
585 address affect regulation skills (Swenson et al., 2001).

---

<sup>4</sup> Cognitive Behavioral Analysis System of Psychotherapy

586           Regarding functional outcome, lower age, the absence of a work disability, high  
587 perceived emotional and physical role limitations and low reported bodily pain were  
588 consistently associated with greater improvement in functional outcome. This finding  
589 suggests that patients may be more motivated to engage in and benefit from treatment if they  
590 feel restricted by their symptoms, while perceived bodily pain might have a negative impact  
591 on motivation and outcome. In accordance with this hypothesis, greater role limitations due to  
592 emotional problems and, interestingly, a higher number of children were consistently positive  
593 predictors of both change in functional outcome and treatment completion. Both factors could  
594 be linked with a higher motivation for treatment. In line, internal motivation was found to be  
595 higher in a small sample of women completing therapy than in those who did not (Webb &  
596 McMurrin, 2009).

597           Concerning premature treatment discontinuation, a higher level of education and a  
598 comorbid severe, recurrent depressive disorder were found to be stable predictors of treatment  
599 completion. In line, another study found also that a lower level of education predicted  
600 premature treatment discontinuation (Perroud, Uher, Dieben, Nicastro, & Huguelet, 2010). In  
601 the study of Kröger et al. (2013), comorbid substance abuse disorder and lower age predicted  
602 (premature) treatment discontinuation. We could not confirm these results with our M1 but  
603 with our M2 model, which indicates a possible clinical relevance.

604           Previous studies found that the presence of comorbid anorexia nervosa increased risk  
605 of treatment discontinuation (Kröger et al., 2013). In our study, being diagnosed with an  
606 endocrine, metabolic, or nutritional disease (which was in 62% an obesity diagnosis) was a  
607 positive predictor of both change in symptom-specific outcome and treatment completion.  
608 Thus, this finding reflects perhaps the operation of the same underlying process, albeit in the  
609 opposite direction.

610           In addition, high reported bodily pain seems to reduce treatment gains across all three  
611 outcomes. A core feature of DBT is training to accept aversive experiences, particularly  
612 through mindfulness strategies (Swenson et al., 2001). This might be particular difficult, or  
613 even frustrating, for patients feeling higher levels of bodily pain, which might lead to  
614 avoidance, and hence to reduced treatment gains or treatment discontinuation. This  
615 argumentation is consistent with a previous study, which found that experiential avoidance  
616 predicted treatment discontinuation in an inpatient DBT sample (Rüsch et al., 2008).  
617 Therefore, it seems helpful for BPD patients with pain to supplement DBT with strategies that  
618 are tailored more specifically to address chronic and resistant pain, such as strategies from the  
619 Acceptance and Commitment Therapy (ACT, Hayes, Luoma, Bond, Masuda, & Lillis, 2006).

620           In a sample of female patients with BPD, dissociation predicted a poor response to  
621 DBT (Kleindienst et al., 2011). The authors linked this effect to the interference of  
622 dissociation with learning processes. In our study, psychoticism (measured by the BSI  
623 subscale) appeared to have a negative effect on the symptom-specific outcome (in the M2  
624 model but not in the M1 model). As dissociation has been found to contribute to the  
625 psychoticism scale of the BSI (Allen, Coyne, & Console, 1996), our results could be  
626 cautiously interpreted as a replication of the effect found by Kleindienst et al. (2011).

627           Interestingly, comorbid severe recurrent depressive disorder seems to reduce the  
628 likelihood of (premature) treatment discontinuation but also of treatment benefit. Thus,  
629 patients with BPD and comorbid severe recurrent depressive disorder may be particularly  
630 needy, seek help through psychotherapy while avoiding conflicts and discontinuations;  
631 however, if the therapy is not tailored to their problems (see above), the probability of  
632 improving seems to be decreasing. Consistent with the first part of this hypothesis, De Panfilis  
633 et al. (2012) found that avoidant personality features protected from early dropout.

634           The heterogeneity of BPD inpatients in our sample is greater than in controlled studies  
635 due to less inclusion and exclusion criteria. This heterogeneity might be better addressed by  
636 stratifying for some important variables (e.g., comorbid depression, comorbid anorexia  
637 nervosa, first inpatient treatment, gender) and building subgroups for additional analyses,  
638 while simultaneously retaining high statistical power due to a large sample size. Therefore,  
639 we further performed exploratory subgroup analyses to check whether the heterogeneity of  
640 the sample might have hidden important predictors in the main analysis. There is a trend that  
641 more variance could be explained by a subgroup of BPD inpatients without anorexia nervosa  
642 across different outcomes. In line with the complex interplay between eating disorders and  
643 BPD, a study showed that a current diagnosis of anorexia nervosa was associated with  
644 increased risk of recurrent non-suicidal self-injury in BPD patients (Chen, Brown, Harned, &  
645 Linehan, 2009). Severe psychiatric comorbidities such as eating disorders might have  
646 allocated resources in therapy session in some individuals but not others, which might have  
647 delayed therapy for the core symptoms of BPD or which might have led to improvements by  
648 different mechanisms of change. It might be that the more homogeneous the sample  
649 (especially without having an anorexia nervosa in our case), the more room for improvement  
650 in affect regulation difficulties in DBT treatment. Perhaps, the skills training is particular  
651 effective in this subgroup, while in connection with eating disorders such as anorexia nervosa  
652 it is probably overshadowed by disorder-specific mechanisms, for which no data were  
653 available in this study (weight, etc.). Further, some interesting single findings leading to a  
654 more nuanced predictor model occurred for some subgroups. However, individual coefficients  
655 have to be carefully interpreted due to the large amount of regressions computed. The smaller  
656 sample sizes might have contributed to the selection of fewer predictors. The discussion of  
657 these findings would go beyond the scope of this section. Hereby, we refer to the  
658 supplemental material 2 for a detailed overview.

659           After the discussion of the variables, which have turned out to be significant predictors  
660 in our analyses, we now discuss the quality and validity of the predictor models. First, our  
661 variable selection procedure seems to have been successful, as it resulted in a set of predictors  
662 that explained at least as much variance as a model including all the observed variables (M3).  
663 In addition, we were able to confirm the findings from previous studies (Barnicot et al., 2012)  
664 that greater symptom severity at baseline consistently predicted higher treatment gains  
665 explaining between 9 and 14% in the symptom-related outcome and between 7 and 9% in the  
666 functional-related outcome. The models appear to be particularly of clinical significance if a  
667 few additional variables (M2) are added to the baseline variables, as the proportion of total  
668 explained variance is then increased (18% in symptom-specific outcome and 17% in  
669 functional outcome) exceeding the threshold of clinical significance ( $R^2=6.3\%$ ). In the  
670 prediction of symptom change in depression, Iniesta et al. (2016) using a similar approach  
671 with regularized regression achieved an  $R^2$  of between 3.9% and 10.2% depending on the  
672 subsample in cross-validation but without a prior variable selection procedure. If only the  
673 additional predictive power is considered without the corresponding baseline variables (only  
674 M1 or M2), only the model for functional outcome achieved clinically significant predictive  
675 performance. However, we used the threshold for continuous measures established in the  
676 context of depression to assess clinical significance (Uher et al., 2012). It is questionable  
677 whether this threshold is well transferable to the population of patients with complex and  
678 severe BPS diagnoses. It is noteworthy that the consideration of nonlinear relationships by  
679 using a state-of-the-art statistical procedure (bartMachine) did not increase the predictive  
680 value of our clinical models, as it was not superior to other computed models. Also, building  
681 subgroups has generally not reliably improved the predictive value of our clinical models,  
682 with the exception that there was a trend towards a higher amount of explained variance in the  
683 BSL-95 for the BPD subgroup without anorexia nervosa.



## Predictors of treatment outcomes in BPD

31

684           The prediction of premature treatment discontinuation by a set of four variables (lower  
685 age, lower education, concurrent substance abuse disorder and less severe recurrent  
686 depressive disorder) was significantly better than random guessing – although not clinically  
687 significant.

**688 Limitations**

689           One limitation of this study is the lack of a control group and the non-random nature  
690 of the sample. Effects or the absence of effects can therefore not be directly related to the  
691 treatment administered. Due to the single-treatment design of our study, we had to rely on  
692 LOCF as a pre-post effect size estimation, knowing that this is a very conservative approach  
693 that is usually not recommended due to its disadvantages (Siddiqui, Hung, & O'Neill, 2009).  
694 In controlled studies with at least two treatment arms (e.g., RCTs), mixed models for ES  
695 estimation offer several advantages over LOCF and in future prospective studies, e.g. the  
696 Mixed-Effect Model Repeated Measure (MMRM) can be used (Siddiqui et al., 2009).

697           Since no strict manual for DBT was used for treatment in this naturalistic study,  
698 treatment adherence cannot be guaranteed, and it is likely that practitioners adapted the  
699 techniques, medication and duration of treatment to the needs of individual patients or other  
700 conditions. In addition, no instruments were collected to measure the adherence and  
701 competence of the therapists. However, there are quality assurances through the DBT  
702 certification of the ward, which includes annual DBT team training by an officially  
703 recognized external DBT-certified supervisor and compliance through regular DBT  
704 supervision by the ward management. A further drawback of the study is the lack of the  
705 diagnostic assurance by a standardized interview with an independent blinded rater. Strongly  
706 related to the naturalistic nature of this study, the data set only contained information on  
707 psychopharmacological treatment at discharge and not at admission, which must be  
708 considered as a limitation.

709           A further limitation relates to the choices of primary and secondary outcome. Since data on  
710 suicidal behavior and NSSI, commonly used outcome indicators in BPD research, were not  
711 available, other outcomes had to be chosen. However, our approach has recently been  
712 recommended by several authors (Panos et al., 2014). Both forms of validation used in this

713 study have their own limitations: While the holdout sample relies on a single measurement,  
714 the dataset used for variable selection overlaps with the dataset used for cross validation,  
715 which might have led to an increased prediction quality. Due to the relatively small amount of  
716 explained variance, the limited predictive power, and the fact that the CI of the AUC touches  
717 almost .5 in the validation phase, one might justifiably argue that the clinical utility of these  
718 models is reduced to some extent, especially for the predictor model of symptom-related  
719 change using the BSL-95. We believe that the low explained variance is more a result of the  
720 naturalistic nature of our study, i.e. the noisy data and the heterogeneous sample, rather than a  
721 problem with the statistical model (Christodoulou et al., 2019). The testing of our clinical  
722 predictive models in cases of (premature) treatment discontinuation and functioning-related  
723 changes should be performed in prospective studies, while the interpretation and  
724 recommendation for the predictor model of the BSL-95 should only be done with caution.  
725 Nevertheless, the results of the analyses could help at least in the theory-based downsizing of  
726 the predictor pool especially for the BSL-95 in future studies.

727 Studies suggest that reducing the pool of potential predictors prior to data analyses might be  
728 useful for several reason (Riley, Snell, Ensor, Burke, Harrell, et al., 2019; Riley, Snell, Ensor,  
729 Burke, Harrell Jr, et al., 2019); namely, an increased number of predictors could lead to more  
730 uncertainty regarding an individual predictor (especially in standard linear regression  
731 models). We attempted to reduce these general negative implications of a high number of  
732 predictors using both regularized regression and resampling strategies. In fact, a key  
733 advantage of our data-driven approach – and of machine learning models in general - is their  
734 robustness to using a large number of “features” or predictors, relative to the number of  
735 observations (Coutanche & Hallion, 2019). A further drawback of this study, despite the large  
736 amount of questionnaire data available, is the relatively restricted range of predictors.  
737 Barnicot et al. (2012) noted that studies should investigate how treatment outcomes are

738 affected by process variables or the relationship between patient and therapist. Other  
739 predictors previously shown to be important, such as number of suicide attempts (Rüsch et al.,  
740 2008), were not available either. Thus, no global model for the prediction of outcomes of  
741 BPD inpatient treatment can be derived from these results. Nevertheless, despite these  
742 limitations, these models represent an attempt to address the multilayered nature of BPD by  
743 using multivariate data on a great variety of candidate predictors.

#### 744 **Future research directions**

745 As we were able to replicate some findings of previous studies concerning the  
746 prediction of symptom-specific and functional outcome and (premature) treatment  
747 discontinuation in an exploratory design, an appropriate next step would be to integrate the  
748 results and carry out confirmatory analysis of the models through more controlled studies.  
749 However, as the clinical significance of predictions based on clinical, demographic, and  
750 symptom-focused self-report data at the beginning of treatment remains questionable, further  
751 research should focus on the influence of motivational factors as well as process variables and  
752 the relationship quality between patient and practitioner. This last is of particular importance,  
753 as disturbed relationships seem to be a BPD phenotype (Gunderson, 2007) regarding BPD as  
754 an interpersonal disorder. As for the prediction of the symptom-specific outcome, the baseline  
755 value of this outcome seems to be more important than any other routinely assessed variable.  
756 More sophisticated statistical tests or simulation studies are needed to determine whether the  
757 effect of the baseline variable stems from statistical artifacts such as regression to the mean or  
758 floor and ceiling effects, or from actual symptom change. Additional modalities of  
759 measurement should routinely be included in the measurement of change (Hasler, Hopwood,  
760 Jacob, Brändle, & Schulte-Vels, 2014). Ultimately, the large prediction variance across  
761 subsamples might be caused partly by the heterogeneity of BPD and the large number of  
762 combinations of comorbidities. Our study attempted to identify variables that would predict

763 outcomes in all subgroups, but analysis of specific subgroups or clusters might uncover  
764 variables with greater predictive power and give additional information on the internal  
765 structure of the BPD construct. Lastly, since this study focused on change scores between  
766 admission and discharge, a rather short-term outcome, large trials including follow-up data  
767 collection might give some more insight into the prediction of long-term outcomes of DBT.

768         Noteworthy, from a methodological point of view, the applied machine learning  
769 approaches to feature selection yield some limitations. As a golden rule, the training and  
770 testing set must always be independent, which is very crucial for feature selection techniques  
771 because a violation can artificially inflate a models performance (Coutanche & Hallion,  
772 2019). Our training and testing set was not completely independent. However, we have tried  
773 to address this validation issue by not only relying on cross-validation on data that has at least  
774 partially been used to perform the initial variable selection, but also, as another internal  
775 validation procedure, by keeping a separate holdout set for testing. Yet again, future external  
776 validation studies are encouraged to replicate our models in completely independent BPD  
777 samples to enhance generalization of our results (Steyerberg et al., 2010).

778         As mentioned above, the risk of uncertainty in relation to a single predictor is higher in  
779 an increased number of potential predictors especially in standard linear regression models  
780 (Riley, Snell, Ensor, Burke, Harrell, et al., 2019; Riley, Snell, Ensor, Burke, Harrell Jr, et al.,  
781 2019). Whenever possible, a prior reduction of the item pool might be useful but it must be  
782 based on a strong theory base to avoid arbitrary decisions. At this stage in the scientific  
783 process on this topic, the data presented can partly serve as a scientific basis for the selection  
784 and more rigorous testing of predictor variables in the future. Future naturalistic studies using  
785 a heterogeneous sample might benefit from stratifying on process variables such as individual  
786 therapy time spent on BPD core symptoms which were not available in this study. Future  
787 more controlled studies with large sample sizes should consider a statistical stratification

788 strategy (e.g., by using mixture modeling) in order to identify empirically derived subgroups  
789 that might yield different predictor sets of BPD patients. BPD patients with an eating disorder  
790 diagnosis might be one of these important subgroups (Zanarini, Reichman, Frankenburg,  
791 Reich, & Fitzmaurice, 2010).

## 792 **Conclusion**

793 Moving towards the vision of evidence-based personalized psychotherapy, where  
794 individual patients can be assigned to the treatment that is most helpful to them, an important  
795 first step is to identify which patients will benefit most or least from existing treatments. Here  
796 we have described the exploratory use of a state-of-the-art stability-based machine learning  
797 algorithm to identify the most important predictors of inpatient treatment outcomes and  
798 (premature) treatment discontinuation in a large, naturalistic sample of BPD patients. Many of  
799 the predictors that emerged had been identified in earlier research, which both validated our  
800 analytical strategy and provided corroboration of the earlier findings. Lower age, a comorbid  
801 substance abuse disorder and a lower level of education seem to be important risk factors for  
802 treatment discontinuation. Poor affect regulation skills and the absence of a history of  
803 outpatient psychotherapy at the beginning of the treatment also predicted better symptom-  
804 specific treatment outcomes, which suggests that the effectiveness of DBT might be based in  
805 particular on the training in affect regulation techniques. High bodily pain seems to diminish  
806 therapy gains and the probability of treatment completion, so specialist techniques should be  
807 developed to better tailor treatment for BPD patients who report severe bodily pain. The  
808 results of this study provide some support for the hypothesis of Kleindienst et al. (2011) that  
809 dissociative symptoms diminish treatment effects in DBT by slowing down emotional  
810 learning processes.

811 However, when excluding the baseline model, most of the models we analyzed  
812 predicted disorder-specific symptoms or treatment discontinuation not at a clinically

## Predictors of treatment outcomes in BPD

37

813 significant level in a cross-validation design. Given these results it does not seem advisable to  
814 base individual treatment decisions only on information about comorbidities, self-reports,  
815 demographic or routinely assessed clinical data at baseline. Several predictors seem to be  
816 linked to motivational processes and the development and change concerning the  
817 therapeutical relationship during the course of treatment, which were not assessed directly in  
818 our data set. In the past, such complex data could not be evaluated on a large scale, but the  
819 further development of machine learning algorithms may make it possible to use more  
820 complex model of human patients for psychiatric research (Wright & Hopwood, 2016). Thus,  
821 in the future it may be possible to provide individualized treatment recommendations and to  
822 tailor treatment much more to the needs and problems of individual patients.

823

Predictors of treatment outcomes in BPD

38

824 **Conflict of interest statement**

825 The authors declare no conflict of interest.

826 **Data Availability Statement**

827 Data sharing is not applicable to this article as only secondary analyses were performed on a  
828 routinely collected data set with permission to use for research purposes but without explicit  
829 permission of data sharing by the patients.

830 **CRedit author statement**

831 **Philipp Herzog:** Conceptualization, Methodology, Validation, Visualization, Writing -  
832 Original Draft and Editing

833 **Matthias Feldmann:** Methodology, Software, Formal analysis, Writing - Original Draft  
834 Preparation

835 **Ulrich Voderholzer:** Investigation, Resources, Data Curation, Writing - Review & Editing,  
836 Project administration for collecting data

837 **Thomas Gärtner:** Resources, Data Curation, Writing – Review & Editing

838 **Michael Armbrust:** Resources, Data Curation, Writing - Review & Editing

839 **Elisabeth Rauh:** Resources, Data Curation, Writing - Review

840 **Robert Doerr:** Resources, Data Curation, Writing - Review

841 **Winfried Rief:** Conceptualization, Writing - Review & Editing, Supervision

842 **Eva-Lotta Brakemeier:** Conceptualization, Methodology, Writing - Review & Editing,  
843 Supervision, Project administration



## References

- Abram, S. V., Helwig, N. E., Moodie, C. A., DeYoung, C. G., MacDonald, A. W., & Waller, N. G. (2016). Bootstrap Enhanced Penalized Regression for Variable Selection with Neuroimaging Data. *Frontiers in Neuroscience, 10*, 344.  
<https://doi.org/10.3389/fnins.2016.00344> U6
- Allen, J. G., Coyne, L., & Console, D. A. (1996). Dissociation contributes to anxiety and psychoticism on the Brief Symptom Inventory. *The Journal of Nervous and Mental Disease, 184*(10), 639–641.
- Barnicot, K., Katsakou, C., Marougka, S., & Priebe, S. (2011). Treatment completion in psychotherapy for borderline personality disorder - a systematic review and meta-analysis. *Acta Psychiatrica Scandinavica, 123*(5), 327–338.  
<https://doi.org/10.1111/j.1600-0447.2010.01652.x>
- Barnicot, Kirsten, Katsakou, C., Bhatti, N., Savill, M., Fearn, N., & Priebe, S. (2012). Factors predicting the outcome of psychotherapy for borderline personality disorder. *Clinical Psychology Review, 32*(5), 400–412. <https://doi.org/10.1016/j.cpr.2012.04.004>
- Bateman, A., & Fonagy, P. (2009). Randomized Controlled Trial of Outpatient Mentalization-Based Treatment Versus Structured Clinical Management for Borderline Personality Disorder. *American Journal of Psychiatry, 166*(12), 1355–1364.  
<https://doi.org/10.1176/appi.ajp.2009.09040539>
- Bland, J. M., & Altman, D. G. (1994). Statistic Notes. *BMJ, 308*(6942), 1499.  
<https://doi.org/10.1136/bmj.308.6942.1499>
- Bohus, M., Limberger, M. F., Frank, U., Sender, I., Gratwohl, T., & Stieglitz, R. D. (2001). Entwicklung der Borderline-Symptom-Liste. *Psychotherapie, Psychosomatik, Medizinische Psychologie, 51*(5), 201–211. <https://doi.org/10.1055/s-2001-13281>
- Bohus, Martin, Haaf, B., Simms, T., Limberger, M. F., Schmahl, C., Unckel, C., ... Linehan,

- M. M. (2004). Effectiveness of inpatient dialectical behavioral therapy for borderline personality disorder: A controlled trial. *Behaviour Research and Therapy*, *42*(5), 487–499. [https://doi.org/10.1016/S0005-7967\(03\)00174-8](https://doi.org/10.1016/S0005-7967(03)00174-8)
- Bohus, Martin, Limberger, M. F., Frank, U., Chapman, A. L., Kühler, T., & Stieglitz, R. D. (2007). Psychometric properties of the Borderline Symptom List (BSL). *Psychopathology*, *40*(2), 126–132. <https://doi.org/10.1159/000098493>
- Bos, E. H., van Wel, E. B., Appelo, M. T., & Verbraak, M. J. P. M. (2010). A Randomized Controlled Trial of a Dutch Version of Systems Training for Emotional Predictability and Problem Solving for Borderline Personality Disorder. *The Journal of Nervous and Mental Disease*, *198*(4), 299–304. <https://doi.org/10.1097/NMD.0b013e3181d619cf>
- Carter, G. L., Willcox, C. H., Lewin, T. J., Conrad, A. M., & Bendit, N. (2010). Hunter DBT Project: Randomized Controlled Trial of Dialectical Behaviour Therapy in Women with Borderline Personality Disorder. *Australian & New Zealand Journal of Psychiatry*, *44*(2), 162–173. <https://doi.org/10.3109/00048670903393621>
- Castonguay, L. G., Barkham, M., Lutz, W., & McAleavey, A. A. (2013). Practice-oriented research: Approaches and application. In M. J. Lambert (Ed.), *Bergin & Garfield's handbook of psychotherapy and behavior change* (6th editio). New York: Wiley.
- Chen, E. Y., Brown, M. Z., Harned, M. S., & Linehan, M. M. (2009). A comparison of borderline personality disorder with and without eating disorders. *Psychiatry Research*, *170*(1), 86–90. <https://doi.org/10.1016/j.psychres.2009.03.006>
- Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, *110*, 12–22. <https://doi.org/10.1016/J.JCLINEPI.2019.02.004>

- Cohen, Z. D., Kim, T. T., Van, H. L., Dekker, J. J. M., & Driessen, E. (2019). A demonstration of a multi-method variable selection approach for treatment selection: Recommending cognitive-behavioral versus psychodynamic therapy for mild to moderate adult depression. *Psychotherapy Research*, 1–14.  
<https://doi.org/10.1080/10503307.2018.1563312>
- Coid, J., Yang, M., Tyrer, P., Roberts, A., & Ullrich, S. (2006). Prevalence and correlates of personality disorder in Great Britain. *British Journal of Psychiatry*, 188(05), 423–431.  
<https://doi.org/10.1192/bjp.188.5.423>
- Coutanche, M., & Hallion, L. (2019). Machine Learning for Clinical Psychology and Clinical Neuroscience. In A. G. C. Wright & M. N. Hallquist (Eds.), *The Cambridge Handbook of Research Methods in Clinical Psychology*. Cambridge, UK: Cambridge University Press.
- Cristea, I. A., Gentili, C., Cotet, C. D., Palomba, D., Barbui, C., & Cuijpers, P. (2017). Efficacy of Psychotherapies for Borderline Personality Disorder. *JAMA Psychiatry*, 74(4), 319. <https://doi.org/10.1001/jamapsychiatry.2016.4287>
- Davidson, K., Norrie, J., Tyrer, P., Gumley, A., Tata, P., Murray, H., & Palmer, S. (2006). The Effectiveness of Cognitive Behavior Therapy for Borderline Personality Disorder: Results from the Borderline Personality Disorder Study of Cognitive Therapy (BOSCOT) Trial. *Journal of Personality Disorders*, 20(5), 450–465.  
<https://doi.org/10.1521/pe.2006.20.5.450>
- De Panfilis, C., Marchesi, C., Cabrino, C., Monici, A., Politi, V., Rossi, M., & Maggini, C. (2012). Patient factors predicting early dropout from psychiatric outpatient care for borderline personality disorder. *Psychiatry Research*, 200(2–3), 422–429.  
<https://doi.org/10.1016/j.psychres.2012.03.016>

- DeCou, C. R., Comtois, K. A., & Landes, S. J. (2019). Dialectical Behavior Therapy Is Effective for the Treatment of Suicidal Behavior: A Meta-Analysis. *Behavior Therapy, 50*(1), 60–72. <https://doi.org/10.1016/J.BETH.2018.03.009>
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves. *Biometrics, 44*(3), 837. <https://doi.org/10.2307/2531595>
- Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13*(03), 595. <https://doi.org/10.1017/S0033291700048017>
- Doering, S., Hörz, S., Rentrop, M., Fischer-Kern, M., Schuster, P., Benecke, C., ... Buchheim, P. (2010). Transference-focused psychotherapy v. treatment by community psychotherapists for borderline personality disorder: randomised controlled trial. *British Journal of Psychiatry, 196*(05), 389–395. <https://doi.org/10.1192/bjp.bp.109.070177>
- Farrell, J. M., Shaw, I. A., & Webber, M. A. (2009). A schema-focused approach to group psychotherapy for outpatients with borderline personality disorder: A randomized controlled trial. *Journal of Behavior Therapy and Experimental Psychiatry, 40*(2), 317–328. <https://doi.org/10.1016/j.jbtep.2009.01.002>
- First, M. B., Gibbon, M., Spitzer, R. L., & Benjamin, L. S. (1997). *User's Guide for the Structured Clinical Interview for DSM-IV Axis II Personality Disorders: SCID-II*. American Psychiatric Press.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software, 33*(1), 1–22. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/20808728>
- Gräfe, K., Zipfel, S., Herzog, W., & Löwe, B. (2004). Screening psychischer störungen mit

- dem “Gesundheitsfragebogen für Patienten (PHQ-D)”. Ergebnisse der Deutschen validierungsstudie. *Diagnostica*, 50(4), 171–181. <https://doi.org/10.1026/0012-1924.50.4.171>
- Gunderson, J. G. (2007). Disturbed Relationships as a Phenotype for Borderline Personality Disorder. *American Journal of Psychiatry*, 164(11), 1637–1640. <https://doi.org/10.1176/appi.ajp.2007.07071125>
- Hasler, G., Hopwood, C. J., Jacob, G. A., Brändle, L. S., & Schulte-Vels, T. (2014). Patient-reported outcomes in borderline personality disorder. *Dialogues in Clinical Neuroscience*, 16(2), 255–266.
- Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and Commitment Therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, 44(1), 1–25. <https://doi.org/10.1016/j.brat.2005.06.006>
- Helwig, N. E. (2017). Adding bias to reduce variance in psychological results: A tutorial on penalized regression. *The Quantitative Methods for Psychology*, 13(1), 1–19. <https://doi.org/10.20982/tqmp.13.1.p001>
- Higgins, J. P. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5). The Cochrane Collaboration.
- Hoerl, A., & Kennard, R. (1988). Ridge regression. In *Encyclopedia of Statistical Sciences* (Vol. 8, pp. 129–136). New York: Wiley.
- Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., ... Uher, R. (2016). Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *Journal of Psychiatric Research*, 78, 94–102. <https://doi.org/10.1016/j.jpsychires.2016.03.016>
- IsHak, W. W., Elbau, I., Ismail, A., Delaloye, S., Ha, K., Bolotaulo, N. I., ... Wang, C.

- (2013). Quality of life in borderline personality disorder. *Harvard Review of Psychiatry*, 21(3), 138–150. <https://doi.org/10.1097/HRP.0b013e3182937116>
- Jenkinson, C., Wright, L., & Coulter, A. (1994). Criterion validity and reliability of the SF-36 in a population sample. *Quality of Life Research*, 3(1), 7–12. <https://doi.org/10.1007/BF00647843>
- Kapelner, A., & Bleich, J. (2016). bartMachine: Machine learning with bayesian additive regression trees. *Journal of Statistical Software*, 70(1), 1–40. <https://doi.org/10.18637/jss.v070.i04>
- Kazdin, A. E. (2007). Mediators and Mechanisms of Change in Psychotherapy Research. *Annual Review of Clinical Psychology*, 3(1), 1–27. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091432>
- Kleindienst, N., Limberger, M. F., Ebner-Priemer, U. W., Keibel-Mauchnik, J., Dyer, A., Berger, M., ... Bohus, M. (2011). Dissociation predicts poor response to Dialectical Behavioral Therapy in female patients with Borderline Personality Disorder. *Journal of Personality Disorders*, 25(4), 432–447. <https://doi.org/10.1521/pedi.2011.25.4.432>
- Kliem, S., Kröger, C., & Kosfelder, J. (2010). Dialectical behavior therapy for borderline personality disorder: A meta-analysis using mixed-effects modeling. *Journal of Consulting and Clinical Psychology*, 78(6), 936–951. <https://doi.org/10.1037/a0021015>
- Kröger, C., Harbeck, S., Armbrust, M., & Kliem, S. (2013). Effectiveness, response, and dropout of dialectical behavior therapy for borderline personality disorder in an inpatient setting. *Behaviour Research and Therapy*, 51(8), 411–416. <https://doi.org/10.1016/j.brat.2013.04.008> U6 - <http://www.sciencedirect.com/science/article/pii/S0005796713000752> M4 - Citavi
- Kröger, C., Holdstein, D., Lombe, A., Schweiger, U., & Kosfelder, J. (2007). Konstruktion

eines störungsspezifischen Instruments zur Erfassung der Impulsivität der Borderline-Persönlichkeitsstörung. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, 36(4), 290–297. <https://doi.org/10.1026/1616-3443.36.4.290>

Kröger, C., Roepke, S., & Kliem, S. (2014). Reasons for premature termination of dialectical behavior therapy for inpatients with borderline personality disorder. *Behaviour Research and Therapy*, 60, 46–52. <https://doi.org/10.1016/j.brat.2014.07.001>

Kröger, C., Theysohn, S., Hartung, D., Vonau, M., Lammers, C.-H., & Kosfelder, J. (2010). Die Skala zur Erfassung der Impulsivität der Borderline-Persönlichkeitsstörung (IS-27). *Diagnostica*.

Leichsenring, F., Leibing, E., Kruse, J., New, A. S., & Leweke, F. (2011). Borderline personality disorder. *The Lancet*, 377(9759), 74–84. [https://doi.org/10.1016/S0140-6736\(10\)61422-5](https://doi.org/10.1016/S0140-6736(10)61422-5)

Linehan, M. M. (1993). *Cognitive behavioral therapy of borderline personality disorder*. New York: Guilford Press.

Linehan, M. M., Armstrong, H. E., Suarez, A., Allmon, D., & Heard, H. L. (1991). Cognitive-Behavioral Treatment of Chronically Parasuicidal Borderline Patients. *Archives of General Psychiatry*, 48(12), 1060–1064. <https://doi.org/10.1001/archpsyc.1991.01810360024003>

Linehan, M. M., Korslund, K. E., Harned, M. S., Gallop, R. J., Lungu, A., Neacsiu, A. D., ... Murray-Gregory, A. M. (2015). Dialectical behavior therapy for high suicide risk in individuals with borderline personality disorder: A randomized clinical trial and component analysis. *JAMA Psychiatry*, 72(5), 475–482. <https://doi.org/10.1001/jamapsychiatry.2014.3039>

McCaughey, E., Berk, M. S., Asarnow, J. R., Adrian, M., Cohen, J., Korslund, K., ... Linehan,

- M. M. (2018). Efficacy of Dialectical Behavior Therapy for Adolescents at High Risk for Suicide. *JAMA Psychiatry*, 75(8), 777. <https://doi.org/10.1001/jamapsychiatry.2018.1109>
- McCullough, J. P. (2003). Treatment for chronic depression: Cognitive behavioral analysis system of psychotherapy (CBASP). *Journal of Psychotherapy Integration*, 13(3–4), 241–263. <https://doi.org/10.1037/1053-0479.13.3-4.241>
- McCullough, J. P., Schramm, E., & Penberthy, J. K. (2015). CBASP as a distinctive treatment for persistent depressive disorder. In *Routledge*. New York: Routledge.
- McMurrin, M., Huband, N., & Overton, E. (2010, April). Non-completion of personality disorder treatments: A systematic review of correlates, consequences, and interventions. *Clinical Psychology Review*, Vol. 30, pp. 277–287. <https://doi.org/10.1016/j.cpr.2009.12.002>
- Mehlum, L., Ramberg, M., Tørmoen, A. J., Haga, E., Diep, L. M., Stanley, B. H., ... Grøholt, B. (2016). Dialectical Behavior Therapy Compared With Enhanced Usual Care for Adolescents With Repeated Suicidal and Self-Harming Behavior: Outcomes Over a One-Year Follow-Up. *Journal of the American Academy of Child and Adolescent Psychiatry*, 55(4), 295–300. <https://doi.org/10.1016/j.jaac.2016.01.005>
- Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4), 417–473.
- Meuldijk, D., McCarthy, A., Bourke, M. E., & Grenyer, B. F. S. (2017). The value of psychological treatment for borderline personality disorder: Systematic review and cost offset analysis of economic evaluations. *PLoS ONE*, 12(3). <https://doi.org/10.1371/journal.pone.0171592>
- Moons, K. G. M., Altman, D. G., Reitsma, J. B., Ioannidis, J. P. A., Macaskill, P., Steyerberg, E. W., ... Collins, G. S. (2015). Transparent Reporting of a multivariable prediction



- model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Annals of Internal Medicine*, *162*(1), W1. <https://doi.org/10.7326/M14-0698>
- New, A. S., Goodman, M., Triebwasser, J., & Siever, L. J. (2008). Recent advances in the biological study of personality disorders. *The Psychiatric Clinics of North America*, *31*(3), 441–461, vii. <https://doi.org/10.1016/j.psc.2008.03.011>
- Ougrin, D., Tranah, T., Stahl, D., Moran, P., & Asarnow, J. R. (2015). Therapeutic Interventions for Suicide Attempts and Self-Harm in Adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, *54*(2), 97-107.e2. <https://doi.org/10.1016/j.jaac.2014.10.009>
- Panos, P. T., Jackson, J. W., Hasan, O., & Panos, A. (2014). Meta-Analysis and Systematic Review Assessing the Efficacy of Dialectical Behavior Therapy (DBT). *Research on Social Work Practice*, *24*(2), 213–223. <https://doi.org/10.1177/1049731513503047>
- Park, M. Y., & Hastie, T. (2008). Penalized logistic regression for detecting gene interactions. *Biostatistics*, *9*(1), 30–50. <https://doi.org/10.1093/biostatistics/kxm010>
- Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, *31*(2), 109–118. <https://doi.org/10.1037/h0024436>
- Perroud, N., Uher, R., Dieben, K., Nícastro, R., & Huguelet, P. (2010). Predictors of response and drop-out during intensive dialectical behavior therapy. *Journal of Personality Disorders*, *24*(5), 634–650. <https://doi.org/10.1521/pedi.2010.24.5.634>
- Priebe, S., Bhatti, N., Barnicot, K., Bremner, S., Gaglia, A., Katsakou, C., ... Zinkler, M. (2012). Effectiveness and cost-effectiveness of dialectical behaviour therapy for self-harming patients with personality disorder: A pragmatic randomised controlled trial. *Psychotherapy and Psychosomatics*, *81*(6), 356–365. <https://doi.org/10.1159/000338897>
- R Core Team. (2018). R: A language and environment for statistical computing. R

*Foundation for Statistical Computing, Vienna(Austria)*. Retrieved from <https://www.r-project.org/>

- Renneberg, B., Schmidt-Rathjens, C., Hippin, R., Backenstrass, M., & Fydrich, T. (2005). Cognitive characteristics of patients with borderline personality disorder: Development and validation of a self-report inventory ARTICLE IN PRESS. *Journal of Behavior Therapy and Experimental Psychiatry*, *36*, 173–182.  
<https://doi.org/10.1016/j.jbtep.2005.05.001>
- Renneberg, B., & Seehausen, A. (2010). Fragebogen zu Gedanken und Gefühlen (FGG). *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *39*(3), 170–178.  
<https://doi.org/10.1026/1616-3443/a000031> U6 - <https://doi.org/10.1026/1616-3443/a000031> M4 - Citavi
- Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E., Moons, K. G. M., & Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction model: Part I - Continuous outcomes. *Statistics in Medicine*, *38*(7), 1262–1275.  
<https://doi.org/10.1002/sim.7993>
- Riley, R. D., Snell, K. I., Ensor, J., Burke, D. L., Harrell Jr, F. E., Moons, K. G., & Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Statistics in Medicine*, *38*(7), 1276–1296.  
<https://doi.org/10.1002/sim.7992>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC. *BMC Bioinformatics*, *12*, 77. <https://doi.org/10.1186/1471-2105-12-77>
- Roth, A., & Fonagy, P. (2005). *What works for whom? : a critical review of psychotherapy research*. Guilford Press.
- Rüsch, N., Schiel, S., Corrigan, P. W., Leihener, F., Jacob, G. A., Olschewski, M., ... Bohus,

- M. (2008). Predictors of dropout from inpatient dialectical behavior therapy among women with borderline personality disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(4), 497–503. <https://doi.org/10.1016/j.jbtep.2007.11.006>
- Siddiqui, O., Hung, H. M. J., & O'Neill, R. (2009). MMRM vs. LOCF: A Comprehensive Comparison Based on Simulation Study and 25 NDA Datasets. *Journal of Biopharmaceutical Statistics*, *19*(2), 227–246. <https://doi.org/10.1080/10543400802609797>
- Soeteman, D. I., Hakkaart-van Roijen, L., Verheul, R., & Busschbach, J. J. V. (2008). The Economic Burden of Personality Disorders in Mental Health Care. *The Journal of Clinical Psychiatry*, *69*(2), 259–265. <https://doi.org/10.4088/JCP.v69n0212>
- Steil, R., Dyer, A., Priebe, K., Kleindienst, N., & Bohus, M. (2011). Dialectical behavior therapy for posttraumatic stress disorder related to childhood sexual abuse: A pilot study of an intensive residential treatment program. *Journal of Traumatic Stress*, *24*(1), 102–106. <https://doi.org/10.1002/jts.20617>
- Stekhoven, D. J., & Bühlmann, P. (2012). Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118. <https://doi.org/10.1093/bioinformatics/btr597>
- Steyerberg, E. W. (2019). *Clinical Prediction Models*. <https://doi.org/10.1007/978-3-030-16399-0>
- Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., ... Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology (Cambridge, Mass.)*, *21*(1), 128–138. <https://doi.org/10.1097/EDE.0b013e3181c30fb2>
- Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal*

*of the Royal Statistical Society. Series B (Methodological)*, 36(2), 111–147.

Swenson, C. R., Sanderson, C., Dulit, R. A., & Linehan, M. M. (2001). The Application of Dialectical Behavior Therapy for Patients with Borderline Personality Disorder on Inpatient Units. *Psychiatric Quarterly*, 72(4), 307–324.

<https://doi.org/10.1023/A:1010337231127>

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Boston: Pearson Education.

Torgersen, S., Kringlen, E., & Cramer, V. (2001). The prevalence of personality disorders in a community sample. *Archives of General Psychiatry*, 58(6), 590–596.

<https://doi.org/10.1001/archpsyc.58.6.590>

Uher, R., Tansey, K. E., Malki, K., & Perlis, R. H. (2012). Biomarkers predicting treatment outcome in depression. *Pharmacogenomics*, 13(2), 233–240.

<https://doi.org/10.2217/pgs.11.161>

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., ... Higgins, P. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open*, 3(8), e002847. <https://doi.org/10.1136/bmjopen-2013-002847>

Ward, M. M., Guthrie, L. C., & Alba, M. (2014). Dependence of the minimal clinically important improvement on the baseline value is a consequence of floor and ceiling effects and not different expectations by patients. *Journal of Clinical Epidemiology*, 67(6), 689–696. <https://doi.org/10.1016/j.jclinepi.2013.10.025>

Ware, J. E., & 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.

Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (1993). SF-36 Health Survey Manual and Interpretation Guide. *Boston New England Medical Centre*.

<https://doi.org/10.1097/00007632-200012150-00008>

Webb, D., & McMurrin, M. (2009). A comparison of women who continue and discontinue treatment for borderline personality disorder. *Personality and Mental Health, 3*(2), 142–149. <https://doi.org/10.1002/pmh.69>

Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout.

*Professional Psychology: Research and Practice, 24*(2), 190.

<https://doi.org/10.1037/0735-7028.24.2.190>

Wright, A. G. C., & Hopwood, C. J. (2016). Advancing the Assessment of Dynamic

Psychological Processes. *Assessment, 23*(4), 399–403.

<https://doi.org/10.1177/1073191116654760>

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology.

*Perspectives on Psychological Science : A Journal of the Association for Psychological Science, 12*(6), 1100–1122.

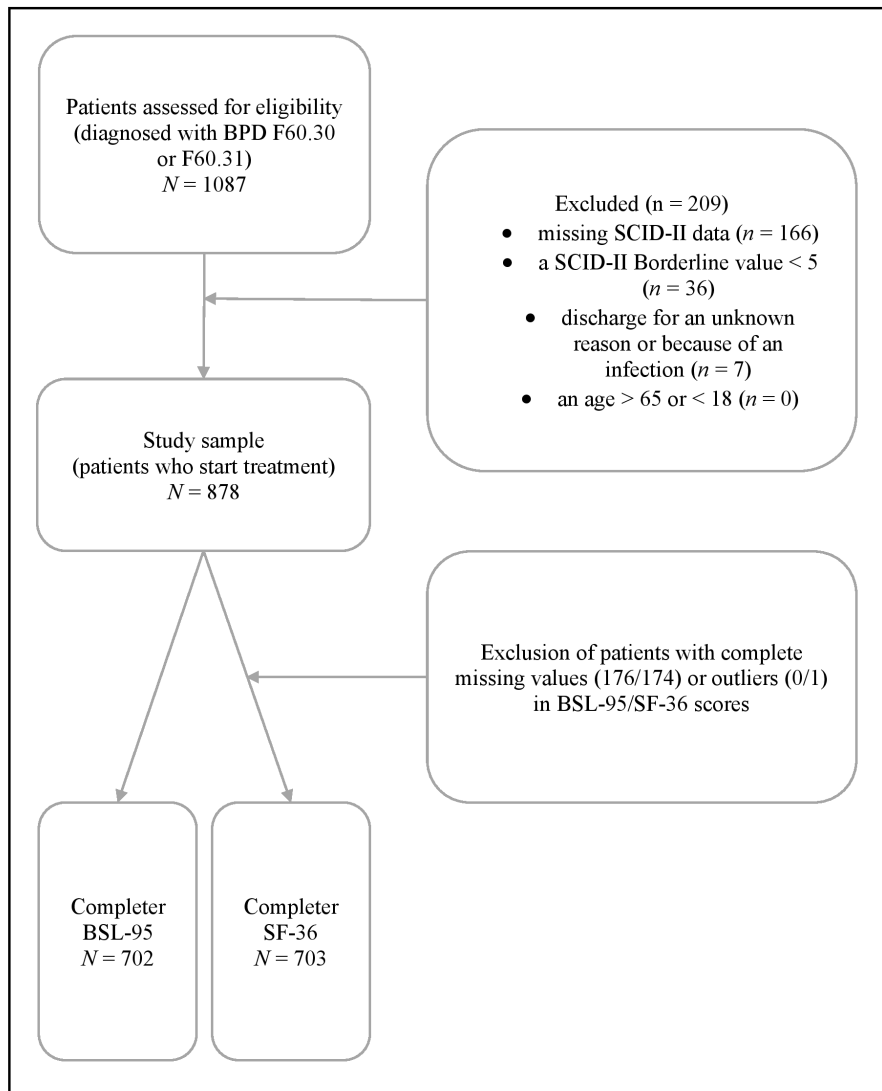
Zanarini, M. C., Frankenburg, F. R., Hennen, J., Reich, D. B., & Silk, K. R. (2006).

Prediction of the 10-Year Course of Borderline Personality Disorder. *American Journal of Psychiatry, 163*(5), 827–832. <https://doi.org/10.1176/ajp.2006.163.5.827>

Zanarini, M. C., Reichman, C. A., Frankenburg, F. R., Reich, D. B., & Fitzmaurice, G.

(2010). The course of eating disorders in patients with borderline personality disorder: A 10-year follow-up study. *International Journal of Eating Disorders, 43*(3), 226–232.

<https://doi.org/10.1002/eat.20689>



**Figure 1.** Patient flow diagram

*Note.* BPD = Borderline Personality Disorder; BSL-95 = Borderline Symptom List-95 (Bohus et al., 2001); SF-36 = Medical Outcomes 36 item Short Form (Ware & Sherbourne, 1992); SCID-II = Structured Clinical Interview for DSM IV II Screening questionnaire (First et al., 1997).

## Predictors of treatment outcomes in BPD

53

Table 1  
*Sociodemographic and clinical characteristics of the sample (N=878 patients and n=702 Completers)*

Characteristics	Patients (N=878)	Completers (n=702) <sup>g</sup>
Age (years) <i>M</i> (SD)	31.1 (10.3)	32.0 (10.4)
Sex <i>n</i> (%)		
male	165 (18.8)	130 (18.5)
female	713 (81.2)	572 (82.5)
Educational score <i>M</i> (SD) <sup>a</sup>	3.0 (0.9)	3.0 (0.9)
Number of mental disorders <i>M</i> (SD) <sup>b</sup>	3.5 (1.2)	3.4 (1.2)
Frequent comorbidities <i>n</i> (%) <sup>b</sup>		
Substance use disorder (F1)	130 (14.8)	85 (12.1)
Depressive episode (F32)	99 (11.3)	70 (10.0)
Recurrent depressive disorder (F33)	759 (86.4)	615 (87.6)
Social phobia (F40.1)	191 (21.8)	142 (20.2)
Post-traumatic stress disorder (F43.1)	149 (17.0)	112 (16.0)
Anorexia nervosa (F50.0)	285 (32.5)	224 (31.9)
ADHD <sup>h</sup> – inattentive type (F90.0)	118 (13.4)	93 (13.2)
Marital status <i>n</i> (%)		
Married	146 (16.6)	122 (17.4)
Divorced or widowed	92 (10.5)	80 (11.4)
Single	640 (72.9)	500 (71.2)
In a relationship <i>n</i> (%)	419 (47.7)	334 (47.6)
Previous inpatient treatments <i>M</i> (SD) <sup>b</sup>	2.75 (3.0)	2.74 (3.0)
First inpatient treatment <i>n</i> (%)	544 (62.0)	432 (61.5)
Occupational status <i>n</i> (%)		
Unemployed	306 (34.9)	229 (32.6)
Retired	147 (17.9)	130 (18.5)
Student, in training, homemaker	133 (15.1)	111 (15.8)
Working full time	159 (18.1)	125 (17.9)
Other	123 (14.0)	106 (16.2)
Working ability <i>n</i> (%)	222 (25.3)	179 (25.5)
Outpatient psychotherapy <i>n</i> (%)	486 (55.4)	393 (56.0)
Outpatient psychiatric treatment <i>n</i> (%)	619 (70.5)	500 (71.2)
Duration of treatment (in days)	59.9 (18.0)	64.4 (13.9)
Discharge type <i>n</i> (%)		
Prematurely discontinued by patient	63 (7.2)	11 (1.6)
Prematurely discharged by the team	44 (5.0)	7 (1.0)
Prematurely by mutual agreement	74 (8.4)	33 (4.7)
Transfer	20 (2.3)	8 (1.1)
Scheduled discharge	677 (77.1)	654 (91.6)
Global BSL-95 score at baseline <i>M</i> (SD) <sup>c</sup>	2.1 (.7)	2.1 (.7)
PHQ-9 score at baseline <i>M</i> (SD) <sup>d</sup>	18.2 (4.9)	18.3 (5.0)
GAF score at baseline <i>M</i> (SD) <sup>e</sup>	42.8 (25.4)	43.3 (28.1)
Global SF-36 score at baseline <i>M</i> (SD)	43.4 (15.4)	43.3 (15.4)

<sup>a</sup> Based on the German school system; scale from 0 (no degree) to 4 (general qualification for university entrance); <sup>b</sup> Diagnosis as given by practitioners according to ICD-10; <sup>c</sup> BSL-95 = Borderline Symptom List, 95 items, scale 0-4 per item; <sup>d</sup> PHQ-9 = Patient Health Questionnaire, 9 items scale 0-3 per item; <sup>e</sup> GAF = Global Assessment of Functioning, scale 0-100; <sup>f</sup> SF-36 = Medical Outcomes 36-item Short-Form Health Survey, 36 items, scale 0-100 per item; <sup>g</sup> BSL-95 pre and post scores exist for these patients; <sup>h</sup> Attention deficit hyperactivity disorder; for each instrument, higher values indicate a more severe manifestation of the symptoms recorded.

Table 2

*Important predictors for residualized symptom change, residualized functional change and premature treatment discontinuation*

	BSL-95 <sup>a</sup>	SF-36 <sup>b</sup>	PTD <sup>g</sup>
<b>Working disability</b>		---	
Substance abuse disorder	++		--
<b>Age</b>		---	++
Anxiety disorder		--	
BDI-II <sup>c</sup>	-		-
BSI <sup>e</sup> anxiety	-	-	
BSI <sup>e</sup> obsession-compulsion		-	-
BSI <sup>e</sup> phobic anxiety	+	+	
BSI <sup>e</sup> positive symptom total	+	+	
BSI <sup>e</sup> psychoticism	--		
<b>BSL-95<sup>a</sup> affect regulation deficiencies</b>	+++	+	
BSL-95 <sup>a</sup> intrusions		-	-
Depressive episode	-	-	
Endocrine, metabolic, nutritional disease	++		+
Duration of work disability	-		-
<b>Educational level</b>		--	+++
Gender female	+	++	
Living alone	++	+	
Number of children	+		+
<b>Outpatient psychotherapy</b>	---		
PHQ-15 <sup>d</sup>		-	-
PHQ-9 <sup>d</sup>	-	-	
<b>Severe recurrent depressive disorder</b>	--	-	+++
<b>SF-36<sup>b</sup> emotional role limitations</b>		+++	-
<b>SF-36<sup>b</sup> bodily pain</b>	--	---	-
<b>SF-36<sup>b</sup> physical role limitations</b>		+++	+
SF-36 <sup>b</sup> social functioning	--		
SCID-II <sup>f</sup> avoidant	-	--	+
SCID-II <sup>f</sup> histrionic	++		
Trained for a job		++	

*Note.* +/- positive/negative effect with consistent sign in 95% of models the predictor was selected in; models were built in 1.000 bootstrap samples; note that for the sake of simplicity an effect in the direction of a favorable outcome is marked with a +, not necessarily a positive coefficient; ++/-- predictor was selected in at least 85% of the samples (in the manuscript referred to as M2); +++/-- selected in at least 95% of the samples (referred to as M1, printed in bold).

<sup>a</sup> Borderline Symptom List-95 (Bohus et al., 2001); <sup>b</sup> Medical Outcomes 36 item Short Form (Ware & Sherbourne, 1992); <sup>c</sup> Beck Depression Inventory-II (Beck, Steer, & Brown, 1996); <sup>d</sup> Patient Health Questionnaire (Gräfe, Zipfel, Herzog, & Löwe, 2004); <sup>e</sup> Brief Symptom Inventory (Derogatis & Melisaratos, 1983); <sup>f</sup> Structured Clinical Interview for DSM IV II Screening questionnaire (First, Gibbon, Spitzer, & Benjamin, 1997); <sup>g</sup> Premature treatment discontinuation



## Predictors of treatment outcomes in BPD

55

Table 3  
*Prediction quality of selected predictors*

	Baseline model	95% VIP model	85% VIP model	Full model	Nonlinear Model
	M0	M1	M2	M3	M4
Five-fold CV					
BSL-95 <sup>a</sup> <i>M (SD) of R<sup>2</sup></i>	.136 (.043)	.006 (.011)	.042 (.029)	.017 (.021)	.123 (.043)
SF-36 <sup>b</sup> <i>M (SD) of R<sup>2</sup></i>	.065 (.032)	.090 (.041)	.106 (.042)	.060 (.034)	.109 (.043)
PTD <sup>c</sup> <i>M (SD) of AUC</i>		.582 (.048)	.642 (.052)	.578 (.05)	.563 (.051)
Holdout sample					
BSL <sup>a</sup> <i>R<sup>2</sup></i>	.094	.007	.006	.025	.090
SF-36 <sup>b</sup> <i>R<sup>2</sup></i>	.085	.064	.054	.058	.206
PTD <sup>c</sup> <i>AUC</i>		.526	.631	.607	.542

*Note.* VIP: variable inclusion probability in elastic net models in 1.000 bootstrap samples in the training set (80%). Five fold CV = Five fold cross validation was repeated 100 times. Holdout sample 20% of complete data set, single prediction trained on rest of the data. M0 = model consisting of only the baseline value, M1 = model additionally containing predictors with a variable inclusion probability of at least 95%, M2 = model using a less strict cutoff of 85%, M3 = model including all variables, M4 = model including also nonlinear relationships using bartMachine (Kapelner & Bleich, 2016); for BSL-95 and SF-36: M1, M2 and M3 improvement over predictive ability of the baseline model M0; AUC: Area under receiver operating curve.

<sup>a</sup> Borderline Symptom List-95 (Bohus et al., 2001); <sup>b</sup> Medical Outcomes 36 item Short Form (Ware & Sherbourne, 1992); <sup>c</sup> Premature treatment discontinuation: Unscheduled treatment discharge before treatment day 55.

## Predictors of treatment outcomes in BPD

56

Table 5  
Most important predictors for subgroups of BPD inpatients

Outcome		Selected predictors in M1	
Subgroup		+	-
<b>BSL-95<sup>a</sup></b>			
Recurrent depressive disorder	BSL-95 affect regulation deficiencies		Outpatient psychotherapy, SCID-II avoidant
Anorexia nervosa	/		/
Without Anorexia nervosa	BSL-95 affect regulation deficiencies		BSI psychoticism, SF-36 bodily pain
First inpatient treatment	/		SF-36 bodily pain
Previous inpatient treatments	/		/
Female	Living alone		/
Male	Disease of the circulatory system		/
<b>SF-36<sup>b</sup></b>			
Recurrent depressive disorder	SF-36 physical role limitations		Working disability, SF-36 bodily pain
Anorexia nervosa	/		Comorbid PTSD (ICD-10: F43.1)
Without Anorexia nervosa	Number of previous inpatient treatments, SF-36 physical role limitations		Age, Working disability, SF-36 social functioning, SF-36 bodily pain
First inpatient treatment	SF-36 physical role limitations, Comorbid F4-diagnosis		Working disability
Previous inpatient treatments	/		Outpatient psychiatric treatment
Female	SF-36 physical role limitations		Working disability, SF-36 bodily pain, Comorbid F4-diagnosis
Male	/		SCID-II antisocial
<b>PTD<sup>c</sup></b>			
Recurrent depressive disorder	Age, Disorders of thyroid gland (ICD-10: E0)		Moderate recurrent depressive disorder
Anorexia nervosa	Endocrine, metabolic, nutritional disease, Other dorsopathies (ICD-10: M5)		/
Without Anorexia nervosa	Age, BSI interpersonal_sensitivity, Severe recurrent depressive disorder		/
First inpatient treatment	/		/
Previous inpatient treatments	/		/
Female	Asthma (ICD-10: J45)		/
Male	/		/

Note. M1 = model additionally containing predictors with a variable inclusion probability of at least 95%. F4 = Anxiety, dissociative, stress-related, somatoform and other nonpsychotic mental disorders according to ICD-10.  
<sup>a</sup> Borderline Symptom List-95 (Bohus et al., 2001); <sup>b</sup> Medical Outcomes 36 item Short Form (Ware & Sherbourne, 1992); <sup>c</sup> Premature treatment discontinuation: Unscheduled treatment discharge before treatment day 55.

## Predictors of treatment outcomes in BPD

57

Table 4  
 Prediction quality for subgroups of BPD inpatients using five-fold CV

	Sample size N (%)	Baseline model M0	95% VIP model M1	85% VIP model M2	Full model M3	Nonlinear model M4
<b>BSL-95<sup>a</sup> M (SD) of R<sup>2</sup></b>						
Recurrent depressive disorder	759 (86.4)	.131 (.048)	.025 (.022)	.047 (.032)	.027 (.030)	.100 (.044)
Anorexia nervosa	285 (32.5)	.133 (.088)	*	-.001 (.009)	-.003 (.012)	.046 (.041)
Without Anorexia nervosa	593 (67.5)	.144 (.057)	.037 (.029)	.070 (.040)	.045 (.033)	.143 (.059)
First inpatient treatment	544 (62.0)	.167 (.064)	.017 (.022)	.029 (.030)	.015 (.023)	.109 (.050)
Previous inpatient treatments	334 (38.0)	.114 (.063)	*	.038 (.043)	-.004 (.015)	.107 (.066)
Female	713 (81.2)	.153 (.049)	.007 (.015)	.043 (.029)	.015 (.017)	.119 (.048)
Male	165 (18.8)	.099 (.090)	.034 (.062)	.076 (.091)	-.005 (.069)	.078 (.087)
<b>SF-36<sup>b</sup> M (SD) of R<sup>2</sup></b>						
Recurrent depressive disorder	759 (86.4)	.066 (.034)	.077 (.040)	.078 (.042)	.052 (.031)	.074 (.041)
Anorexia nervosa	285 (32.5)	.068 (.069)	.030 (.063)	.087 (.063)	.023 (.054)	.105 (.086)
Without Anorexia nervosa	593 (67.5)	.072 (.044)	.117 (.053)	.139 (.055)	.074 (.044)	.124 (.052)
First inpatient treatment	544 (62.0)	.087 (.046)	.056 (.039)	.113 (.059)	.046 (.045)	.079 (.050)
Previous inpatient treatments	334 (38.0)	.057 (.049)	.038 (.052)	.062 (.047)	.024 (.040)	.071 (.058)
Female	713 (81.2)	.067 (.037)	.081 (.043)	.106 (.047)	.045 (.031)	.083 (.041)
Male	165 (18.8)	.075 (.080)	.104 (.099)	.130 (.118)	.012 (.067)	.109 (.098)
<b>PTD<sup>c</sup> M (SD) of AUC</b>						
Recurrent depressive disorder	759 (86.4)		.626 (.066)	.646 (.057)	.597 (.057)	.585 (.053)
Anorexia nervosa	285 (32.5)		.639 (.062)	.709 (.081)	.607 (.084)	.593 (.082)
Without Anorexia nervosa	593 (67.5)		.608 (.060)	.636 (.066)	.547 (.053)	.551 (.049)
First inpatient treatment	544 (62.0)		*	.629 (.067)	.575 (.063)	.561 (.059)
Previous inpatient treatments	334 (38.0)		*	.607 (.080)	.542 (.065)	.555 (.064)
Female	713 (81.2)		.531 (.027)	.629 (.057)	.568 (.053)	.555 (.054)
Male	165 (18.8)		*	.648 (.111)	.543 (.086)	.628 (.101)

Note. VIP: variable inclusion probability in elastic net models in 1,000 bootstrap samples in the training set (80%). Five fold CV = Five fold cross validation was repeated 20 times. M0 = model consisting of only the baseline value, M1 = model additionally containing predictors with a variable inclusion probability of at least 95%, M2 = model using a less strict cutoff of 85%, M3 = model including all variables, M4 = model including also nonlinear relationships using bartMachine (Kaplaner & Bleich, 2016); for BSL-95 and SF-36: M1, M2 and M3 improvement over predictive ability of the baseline model M0; AUC: Area under receiver operating curve.

<sup>a</sup> Borderline Symptom List-95 (Bohus et al., 2001); <sup>b</sup> Medical Outcomes 36 item Short Form (Ware & Sherbourne, 1992); <sup>c</sup> Premature treatment discontinuation: Unscheduled treatment discharge before treatment day 55; \* no predictors were selected.



Predictors of treatment outcome in PTSD

2

- 24 Maria Rolvering-Dijkstra,  
25 Schön-Klinik Roseneck, Psychosomatic Clinic,  
26 Am Roseneck 6, D-83209 Prien am Chiemsee, Germany
- 27 Matthias Feldmann,  
28 Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,  
29 Gutenbergstraße 18, D-35032 Marburg, Germany,
- 30 Winfried Rief,  
31 Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,  
32 Gutenbergstraße 18, D-35032 Marburg, Germany,
- 33 and Eva-Lotta Brakemeier,  
34 Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,  
35 Gutenbergstraße 18, D-35032 Marburg, Germany
- 36 \*Correspondence can be addressed to: Philipp Herzog, Philipps-University of Marburg,  
37 Department of Clinical Psychology and Psychotherapy, Gutenbergstraße 18, D-35032  
38 Marburg, Germany, Email: philipp.herzog@staff.uni-marburg.de, Telephone: +49 6421  
39 2824061
- 40 **Conflict of interest statement**  
41 The authors declare no conflict of interest.
- 42 **Acknowledgments**  
43 The authors would like to thank the reviewers for their helpful and valuable feedback on our  
44 manuscript.
- 45 **Contributions**  
46 All authors materially participated in the research and/or article preparation. All authors have  
47 approved the final article.
- 48 **Funding**  
49 The authors received no funding from an external source.
- 50 **Word length (excluding abstract, tables/figures and references):** 7056

51

**Abstract**

52 *Objective:* The aims of this study were to determine the effectiveness of a routine clinical care  
53 treatment and to identify predictors of treatment outcome in PTSD inpatients.

54 *Methods:* A routinely collected data set of 612 PTSD inpatients ( $M = 42.3$  years [ $SD = 11.6$ ],  
55 75.7% female) having received trauma-focused psychotherapy was analyzed. Primary  
56 outcome was the clinical symptom severity change score, secondary outcomes were assessed  
57 using functional, anxiety, and depression change scores. Hedges  $g$ -corrected pre-post effect  
58 sizes (ES) were computed for all outcomes. Elastic net regulation as a data-driven, stability-  
59 based machine-learning approach was used to build stable clinical prediction models.

60 *Results:* Hedges  $g$  ES indicated medium to large effects on all outcomes. The results of the  
61 predictor analyses suggested that a combined predictor model with sociodemographic,  
62 clinical, and psychometric variables contribute to predicting different treatment outcomes.  
63 Across the clinical and functional outcome, psychoticism, total number of diagnoses, and  
64 bronchial asthma consistently showed a stable negative predictive relationship to treatment  
65 outcome.

66 *Conclusion:* Trauma-focused psychotherapy could effectively be implemented in a routine  
67 inpatient setting. Some important prognostic variables could be identified. If the proposed  
68 models of predictors are replicated, they may help personalize treatment for patients receiving  
69 routine clinical care.

70

71

**Keywords:**

72 posttraumatic stress disorder; PTSD; predictors; prognostic variables; patient characteristics;

73

treatment outcome

74

**75 Clinical Significance of this Article**

76 This study strongly suggests that trauma-focused psychotherapy could effectively be  
77 implemented in a routine inpatient setting and might therefore serve as a benchmark of the  
78 effectiveness of PTSD inpatient treatment.

79 When treating inpatients with PTSD, especially psychoticism, total number of diagnoses,  
80 and bronchial asthma seem to be important prognostic variables of PTSD treatment outcome  
81 and should be addressed at an early treatment stage in order to improve treatment outcome.

82

83

84

85

86

87

**88 Data availability statement**

89 Data sharing is not applicable to this article as only secondary analyses were performed  
90 on a routinely collected data set with permission to use for research purposes but without  
91 explicit permission of data sharing by the patients.

**92 Data transparency statement**

93 The data reported in this manuscript have not been previously published/in press/under  
94 review and were not collected as part of a larger data collection. Findings from the data  
95 collection have not been reported in separate manuscripts.

96

**Introduction**

97 Posttraumatic stress disorder (PTSD) is a severe and often chronic mental disorder in  
98 people who have been exposed to a potentially traumatic, life-threatening event. Prevalence  
99 rates vary from study to study, depending on the sample and measurement used. The median  
100 point prevalence of PTSD in primary care setting across studies worldwide is 12.5% overall;  
101 in the civilian population 11.1%; in a special-risk population in primary care (e.g. refugees  
102 and immigrants) 12.5%; and in veterans about 24.5% (Spottswood, Davydow, & Huang,  
103 2017). Meta-analyses and systematic reviews support the efficacy of psychotherapy for  
104 posttraumatic stress disorder (PTSD), especially trauma-focused cognitive-behavioral  
105 therapy (CBT) in outpatient settings (Bisson et al., 2007; Cusack et al., 2016; Ehlers et al.,  
106 2010). Furthermore, there is first evidence that a short-term, 7-day intensive and standard  
107 weekly cognitive therapy (CT) for PTSD in an outpatient setting is efficacious (Ehlers et al.,  
108 2014), and that CT for PTSD could effectively be implemented in routine clinical outpatient  
109 care (Ehlers et al., 2013). Yet, there is still little evidence of how effective CBT is in routine  
110 clinical care, especially in inpatient settings. Research on the treatment of PTSD indicates  
111 that there is a research-practice-gap (Cook, Schnurr, & Foa, 2004; Karlin & Cross, 2014;  
112 Shiner et al., 2013; Sobell, 1996; Tarrrier, 2010). In addition, nonresponse and dropout rates  
113 of trauma-focused CBT are relatively high, notwithstanding its relatively high overall  
114 efficacy (Bradley, Greene, Russ, Dutra, & Westen, 2005; Schottenbauer, Glass, Arnkoff,  
115 Tendick, & Gray, 2008). In general, nonresponse rates exceed 50% on at least some  
116 measures in many studies, whereas dropout rates can range from 0 to 54%, depending on the  
117 studied population and with no differences across treatments (Bradley, Greene, et al., 2005;  
118 Schottenbauer et al., 2008)). In particular, dropout rates in real-world studies were found to  
119 be higher than those in randomized controlled trials (RCT; Najavits, 2015). These results  
120 suggest there is potential for improving treatment outcome. Determining predictors of



## Predictors of treatment outcome in PTSD

6

121 treatment outcome might be an important initial step, enabling the identification of probable  
122 nonresponders and then tailoring treatments to their specific needs. The question arises, then:  
123 What do we know thus far about the determinants of treatment outcome in PTSD patients?

124 In one of the first attempts to identify demographic or clinical predictors of treatment  
125 outcome in PTSD patients (Van Minnen, Arntz, & Keijsers, 2002), demographic variables;  
126 depression and general anxiety; personality; trauma characteristics; feelings of anger, guilt,  
127 and/or shame; and nonspecific variables concerning treatment were not linked to treatment  
128 efficacy or dropout. Nevertheless, the results of this study indicate a relationship between  
129 higher symptom severity at pretreatment and a higher symptom severity at posttreatment and  
130 followup. The authors concluded therefore that it is difficult to use pre-treatment variables as  
131 a powerful and reliable tool for predicting treatment outcome or dropout. Recent studies  
132 (e.g., Sonne et al., 2016) challenge this result and argue that psychosocial factors in general  
133 predict treatment outcome in PTSD patients and that employment status predicts PTSD-  
134 specific symptom change in particular. Additionally, another study indicates that younger  
135 age, lower intelligence, and less education were related with higher dropout and higher  
136 depression scores and that guilt at pretreatment was linked to better outcome in PTSD  
137 symptoms (Rizvi, Vogt, & Resick, 2012). In line with this finding, comorbid depressive  
138 symptom severity predicted a better treatment outcome measured by decline in PTSD  
139 symptoms (Richardson et al., 2014). There is increasing evidence that patients with complex  
140 PTSD (cPTSD) are less likely to benefit from first-line, evidence-based treatment (Lonergan,  
141 2014). A recent meta-analysis supports that psychological interventions for cPTSD showed  
142 smaller but still moderate to large effects (Karatzias et al., 2019). These partly contradictory  
143 single findings suggest the need for studies identifying patients at risk for not responding and  
144 determining stable predictors of this patient population. A major deficit of prior studies on  
145 predictors lies in their low power due to small sample sizes (e.g., Van Minnen et al., 2002).

## Predictors of treatment outcome in PTSD

7

146 Further limitations concern the use of different outcome measures in the studies, the  
147 inclusion of only a small number of potentially relevant pretreatment variables, and often  
148 insufficient statistical approaches. These limitations complicate the interpretability and  
149 generalizability of the individual findings and partially explain the very small explained  
150 variance in outcome.

151 These concerns suggest the need for a multivariate approach that combines  
152 sociodemographic, clinical, and psychometric data into one model in order to address the  
153 heterogeneity of the PTSD population. Multiple regression is a method that is often chosen  
154 for the task of identifying predictors from a set of candidate predictors in psychological  
155 research. Often, statistical significance is used as a criterium whether a candidate is a true  
156 predictor or not. However, the more candidates are added to a model, the better the model  
157 will be at predicting the data it was trained in, until all variance can be explained, when there  
158 are as many predictors as cases in the dataset – a problem usually called as ‘overfitting’. The  
159 model is then perfect at reproducing the dataset it was trained in (training set) but will  
160 probably fail at predicting for new datasets. In this context, several authors recommend  
161 machine learning as a promising approach (Bzdok, Karrer, Habel, & Schneider, 2017;  
162 Dwyer, Falkai, & Koutsouleris, 2018). Synthesizing multiple variables through machine  
163 learning may improve the prediction of treatment outcomes (Iniesta et al., 2016).  
164 Regularized regression for example with the elastic net (Zou & Hastie, 2005) is a variation  
165 of multiple regression that can deal with a larger set of candidate predictors without severely  
166 overfitting to the training set. The question is not, how well the data in the training set can be  
167 explained, but how well data in an independent testing set can be predicted, which puts the  
168 generalizability and practical use of the model more into focus. However, no significance  
169 intervals can be computed for the regularized regression coefficients. Instead the model is  
170 recomputed on many random subsamples of the training set and the stability of the

## Predictors of treatment outcome in PTSD

8

171 coefficients can be used to decide whether a candidate is a true predictor (Meinshausen &  
172 Bühlmann, 2010). A recent study has demonstrated the usefulness of different machine  
173 learning algorithms for treatment selection for depression and includes also the  
174 aforementioned approach (Z. D. Cohen, Kim, Van, Dekker, & Driessen, 2019). In line with  
175 this, a recent study (Deisenhofer et al., 2018) has shown that machine learning may be useful  
176 for studying and guiding individual treatment selection for PTSD. Addressing these  
177 recommendations and building upon these findings, the aim of the current study was: (1) to  
178 determine the effectiveness of PTSD treatment in routine clinical inpatient care, and (2) to  
179 identify stable predictors of treatment outcome in a large inpatient PTSD sample. In doing  
180 so, this investigation used an exploratory, data-driven approach, based on prediction rather  
181 than explanation, to increase interpretability of the results, prevent overfitting of data into a  
182 large dataset, and ease translation of the findings into practical guidelines (cf. Yarkoni &  
183 Westfall, 2017).

184

185 **Materials and Methods**

186 The study has been conducted in accordance with the ethical standards as laid down  
187 in the 1964 Declaration of Helsinki and its later amendments.

188 **Sample**

189 For this study, we were able to analyze data routinely collected between 2013 and  
190 2017 from PTSD patients treated in five German specialized inpatient clinics. The present  
191 analyses included, consecutively, all patients who were diagnosed and treated for PTSD  
192 (F43.1 according to ICD-10 provided by practitioners), who fulfilled the inclusion criteria  
193 and who gave informed consent to anonymous evaluations of their routinely collected data.  
194 In addition to the IES-R, the reliability of the diagnosis of PTSD was assured by a trained  
195 therapist and reassured by the medical head of the ward. Inclusion criteria were defined as: a  
196 primary diagnosis of PTSD, treatment focus on PTSD, and aged between 18–65 years.  
197 Because the outcome measures were not routinely collected across all clinics and wards,  
198 patients with complete missing values on the primary and secondary outcome were  
199 excluded. Patients who filled in the outcome measures were further excluded if they had  
200 missing data of more than 60% on the outcome measures. To address the heterogeneity of  
201 PTSD patients, no further exclusion criteria were defined a priori. The study was naturalistic  
202 in its nature: Figure 1 shows the patient flow.

203 --Insert Figure 1--

204 In our sample of 612 PTSD inpatients, the mean age at baseline was 42.3 years ( $SD =$   
205 11.6), 75.7% were female, with a mean number of 2.7 comorbid mental disorders.  
206 Approximately two thirds had a comorbid depression (69.1%), on average moderately to  
207 severely pronounced (BDI-II:  $M = 34.9$ ,  $SD = 11.1$ ; in line PHQ-9:  $M = 17.2$ ,  $SD = 5.2$ ),  
208 indicating a highly disabled patient sample. Both instruments were used in order to enhance  
209 comparability with previous studies (for example Deisenhofer et al., 2018). In line, the mean

210 symptom severity score measured by the IES-R at baseline was 77.8 ( $SD = 16.8$ ). The mean  
211 functioning level measured by the GAF was 44.5 at baseline ( $SD = 8.3$ ). Regarding previous  
212 treatment, 68.6% of all patients reported to have received outpatient psychotherapy, 75.3%  
213 had received psychiatric treatment. For 42.8% of the patients, it was the first inpatient  
214 treatment.

215 Table 1 depicts the relevant sociodemographic and clinical characteristics.

216 --Insert Table 1 here--

### 217 **Routine clinical care treatment**

218 The multiple clinics are part of one large clinic group and all offer a multimodal  
219 trauma-focused behavioral treatment program based on national and international guidelines  
220 of psychotherapy for PTSD (e.g. National Institute for Clinical Excellence (NICE), 2005;  
221 Schäfer et al., 2019). In Germany, outpatient treatment of PTSD offers only in  
222 approximately 70% of all patients specific elements of trauma-focused psychoeducation,  
223 skills training, and trauma exposure in sensu as well as in vivo (Equit, Maurer, Michael, &  
224 Köllner, 2018). Trauma-focused inpatient treatment therefore aims to deliver a maximum of  
225 trauma-focused group-therapy containing psychoeducation, coping with trauma-associated  
226 hyperarousal and dissociation, cognitive therapy, training of alternative skills based on  
227 Linehan's DBT manual (Linehan, 1993), and administration of therapist-accompanied  
228 exposure therapy in individual psychotherapy as a crucial change factor in the treatment of  
229 PTSD. Inpatient treatment in specific wards for the treatment of PTSD in these clinics  
230 comprises three stages: 'motivation and psychoeducation', 'exposure-based interventions'  
231 and 'transfer and relapse prevention'.

232 The first stage of 'psychoeducation' includes trigger analyses, the identification of  
233 trauma-related safety behavior and avoidance strategies, the rationale of exposition

## Predictors of treatment outcome in PTSD

11

234 exercises, the development of an individual hierarchy of critical situations and related  
235 negative appraisals (e.g. misinterpretation of trigger situations as threatening, beliefs of  
236 shame, guilt and invalidation) as well as the functionality of dysfunctional behavior. The  
237 second stage ‘exposure-based interventions’ focuses on techniques of Prolonged Exposure  
238 (PE; Foa, Hembree, & Rothbaum, 2015), Imagery Rescription and Reprocessing Therapy  
239 (IRRRT; Smucker, Reschke, & Kögel, 2008) or Eye Movement Desensitization and  
240 Reprocessing Therapy (EMDR; Shapiro, Kierdorf, Höhr, & Mallett, 2013), if patients were  
241 sufficiently stable, comprising accompanied exposition exercises including access to  
242 supportive talks delivered by specifically trained co-therapy (i.e., nursing staff) (1 hour per  
243 week). In the third stage of ‘transfer and relapse prevention’ the focus lies on self-exposure  
244 and coping with triggers under real-life conditions and inclusion of relatives and partners  
245 into individual psychotherapy. PTSD-specific elements are elaborated in PTSD-specific  
246 group therapy (2 hours per week). Inpatient individual psychotherapy (1 hour per week) is  
247 accompanied by continuous training of social skills and assertiveness including functional  
248 assessment of risk situations (2 hours per week), skills training (1 hour per week) and  
249 trauma-focused art therapy (2 hours per week). Additionally, about two hours per day are  
250 reserved for self-controlled behavioral training of functional behavior.

251 Multimodal and transdiagnostic interventions include weekly interactional ward  
252 groups (4 hours per week; e.g. elaboration of alternative social skills and coping with  
253 emotions, identification and treatment of individual functions of safety- and avoidance-  
254 behavior), sports and trauma-focused body therapy groups (3 hours per week), symptom-  
255 specific group therapy for comorbid disorders (2 hours per week; e.g. depression or  
256 somatoform disorders), and, according to individual indication and depending on  
257 comorbidities, mindfulness exercises, progressive muscle relaxation, biofeedback and social  
258 counseling (by trained social workers) (2 hours per week). Additionally, modules of

## Predictors of treatment outcome in PTSD

12

259 Acceptance Commitment Therapy (ACT) with a focus on acceptance-based coping with  
260 trigger situations, intrusions and hyperarousal are included into individual and group  
261 therapies. Therapy is delivered under weekly team supervision and a weekly senior visit.

262 In total, inpatients receive at least 1 hour per week individual psychotherapy, 8 hours  
263 per week of trauma-focused therapy plus access to an average of 11 hours per week of  
264 multimodal and transdiagnostic interventions over 8-10 weeks leading to a total of 152-200  
265 therapy hours. In our sample, the average length of stay was  $M = 54.3$  days ( $SD = 15.5$ ) with  
266 a range 6 – 98 days.

267 For the patient-staff ratio on each PTSD ward, a minimum of two licensed  
268 psychotherapists with specific training in PTSD treatment are available (3-5 years of training  
269 in CBT; one medical doctor and one certified clinical psychologist), 1-2 therapists in  
270 advanced CBT training, and at least 1,5 certified nursing staff, through which the adherence  
271 to trauma-focused psychotherapy including therapist-accompanied exposure could be  
272 guaranteed. During the weekends, the clinic warrants emergency care with medical,  
273 psychological and nursing staff.

274 Psychopharmacological treatment is administered when needed, mainly to treat  
275 psychiatric comorbidities, in compliance with current national and international guidelines  
276 for PTSD treatment (e.g. National Institute for Clinical Excellence (NICE), 2005; Schäfer et  
277 al., 2019) and according to clinical expert supervision. Nevertheless, the major treatment  
278 focus lies on psychotherapeutic work in single and group settings; medication is therefore  
279 continued or reduced to former psychopharmacological treatment before intake.

**280 Measures**

281 In the big data set, there were several self- and observer-rated instruments that  
282 assessed treatment outcome and were administered both at admission ( $T1 =$  pre-

283 measurement on the first day) and discharge (T2 = post-measurement on the last day). The  
284 instruments we used for our analyses are described below.

285 **Clinical outcome: Severity of PTSD symptoms**

286 The Impact-of-Event-Scale-Revised (IES-R; Maercker & Schützwohl, 1998) is a 22-  
287 item self-rating questionnaire measuring three subscales of PTSD symptoms: intrusion,  
288 avoidance, and hyperarousal.

289 **Functional outcome: General Functioning**

290 The Global Assessment of Functioning (GAF; Endicott, Spitzer, Fleiss, & Cohen, 1976)  
291 is an observer-rated index used to determine psychological, social, and occupational  
292 functioning level on a continuum ranging from severe mental illness (0) to mental health  
293 (100). Multimodal inpatient treatment programs such as in our practice-based study aim at  
294 improving the patients functioning level. Within the German mental health care system and  
295 specifically inpatient settings, the GAF is usually used to determine the level of functioning  
296 due to its cost and time efficient administration.

297 **Other secondary outcomes: Depression and Anxiety**

298 The Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996) was used to  
299 measure depressive symptom severity. For this study, in addition to the BDI-II, the PHQ-9  
300 was used to determine the degree of comorbid depressive symptoms in order to achieve a  
301 higher comparability of our sample with other studies (see Deisenhofer et al., 2018).

302 The Generalized Anxiety Disorder-7 (GAD-7; Spitzer, Kroenke, Williams, & Lowe,  
303 2006) is the anxiety module of the Patient Health Questionnaire (PHQ; Gräfe, Zipfel,  
304 Herzog, & Löwe, 2004).



**305 Candidate predictors**

306 For the clinical outcome (IES-R) and functional outcome (GAF), prediction analyses  
307 were performed. In these analyses, additional sociodemographic, clinical und psychometric  
308 variables were included:

309 Patients provided sociodemographic information (e.g., age, sex, inability to work, wish to  
310 retire) at intake in a self-report questionnaire. An initial assessment of a wider set of medical  
311 conditions recorded relevant clinical variables (i.e. mental and medical diagnoses, previous  
312 psychotherapeutic and/or psychiatric treatment, and number of comorbidities). Several other  
313 psychometric variables (Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983),  
314 Satisfaction With Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985), PHQ-D  
315 (Gräfe et al., 2004); see Supplementary Material for detailed descriptions of psychometric  
316 variables) were also administered at intake. For further general details concerning candidate  
317 predictors, see Supplementary Material (Figure 2).

**318 Statistics**

319 All statistical analyses were computed with RStudio (R Core Team, 2018).

**320 Preprocessing**

321 As mentioned above, variables with less than 40% missing values were used. Univariate  
322 outlier values were removed by a  $p < .001$  criterion (Tabachnick & Fidell, 2013) adapted to  
323 the large sample size on the z distribution for all continuous variables. Time-event data were  
324 log-transformed. Interpretable binary or continuous score variables were constructed out of  
325 categorical variables. Binary variables with severe class imbalance ( $\geq 90\%$  in one category)  
326 were excluded in order to enhance stability in the variable selection procedure. In order to  
327 make use of the information included in the ICD-10 medical diagnoses, dummy variables  
328 were created for all diagnosis labels and the class imbalance criterion was applied here as  
329 well. If a diagnosis and a single sub-diagnosis following the tree structure of the ICD-10

330 were selected, only the sub-diagnosis was used. The data then were split randomly into a  
331 training set (65% of the data) and a test set (35%). Missing data imputation was performed  
332 by random forest imputation using the R package missForest (Stekhoven & Bühlmann,  
333 2012), a nonparametric imputation algorithm based on random forests, separately on the  
334 training and testing data sets. Random forest has been shown to be a reliable imputation  
335 technique for psychological research (Golino & Gomes, 2016). Variable selection and  
336 imputation was performed on the training set only to prevent information leakage (Dwyer et  
337 al., 2018; Varma & Simon, 2006).

#### 338 *Statistical methods for calculating effectiveness*

339 Hedges'  $g$ -corrected pre-post effect sizes (ES) were computed for all outcomes (Hedges,  
340 1981). To obtain an upper bound on the influence of ES due to dropout, last observation  
341 carried forward (LOCF) ES were computed as well. Cohen's  $d$  and Hedges  $g$  are interpreted  
342 in a similar way: .20 indicates a small effect, .50 a medium effect and .80 a large effect (J.  
343 Cohen, 1988).

#### 344 *Statistical methods for the identification of predictors*

345 Before data analysis of candidate predictors, preprocessing modules were applied (see  
346 Supplementary Material for a detailed description of the preprocessing).

347 **Variable selection.** To select the most important variables, elastic net regularization  
348 (ENR; Zou & Hastie, 2005) with the R packages glmnet (Friedman, Hastie, & Tibshirani,  
349 2010) and caret (Kuhn, 2008) was performed. ENR is a form of penalized regression that  
350 combines ridge regression (Hoerl & Kennard, 1988) and Lasso regression (Tibshirani, 1996),  
351 and was used before to predict clinical outcomes with demographic and clinical data (Iniesta  
352 et al., 2016). Ridge regression uses the  $L_2$  norm of the coefficients as a penalty to shrink  
353 regression coefficients in order to avoid overfitting to the training data. It also shrinks  
354 coefficients of correlated variables toward each other – also called the grouping effect. Lasso

355 regression uses an  $L_1$  penalty and is thus able to shrink coefficients exactly to zero,  
356 performing implicit variable selection. Unlike ridge regression, it chooses one of a set of  
357 correlated variables. ENR combines the grouping effect with the ability to variable selection,  
358 including or excluding variables as a group, using the scalar alpha to weight  $L_1$  and  $L_2$   
359 penalties. An alpha of 0.5 for elastic net regression was chosen to balance the effects of the  
360 two penalizers. In our setting with many correlated variables, this feature is assumed to  
361 enhance the stability of the results.

362 ENR requires the setting of a shrinkage parameter lambda, which can be empirically  
363 estimated via repeated k-fold cross validation (Friedman et al., 2010; Stone, 1974). For the  
364 purpose of cross validation, the training data were randomly split into k folds. Over various  
365 values of lambda, elastic net models were trained on k-1 folds and tested on the unused fold.  
366 This process was repeated k times, such that each fold was used as a test fold once. The  
367 lambda value *lambda.min*, which produced the lowest root mean squared error cross the k  
368 folds, was selected for each cross validation run. The final lambda was obtained by  
369 averaging *lambda.min* over 10 cross-validation runs. An elastic net model was then fitted  
370 using normalized data *lambda.min* as a parameter.

371 **Statistical Replication.** Based on Park and Hastie (2008) and Cho et al. (2010), the  
372 model-fitting process was repeated on 1000 bootstrap samples in the training set. For each  
373 predictor, the percentage of samples was recorded in which the predictor was included in the  
374 elastic net model. Within the models in which a variable has been included, the percentage  
375 of positive coefficients for that variable was computed. Laurin, Boomsma, and Lubke (2016)  
376 argue that variables with a high frequency and high consistency in their sign are more likely  
377 to be stable predictors.

378 **Model Testing.** To enhance reproducibility, the fitted models were then used to  
379 predict change scores for patients in the training set and test set. Additionally, cross

## Predictors of treatment outcome in PTSD

17

380 predictions were made for the IES-R predictors to predict GAF change scores in the test set  
381 and the other way around. R-squared as well as root mean squared errors (RMSE) and mean  
382 average errors (MAE) were computed for each prediction.

383

**Results****384 Effectiveness of the inpatient treatment**

385 For the symptom-specific outcome, Hedge's  $g$  of the symptom change score in the  
 386 IES-R was .58, 95% CI [.47, .70], LOCF-corrected ES was .59, 95% CI [.48, .70], indicating  
 387 a medium effect. For the functional outcome, Hedge's  $g$  of the functional change outcome on  
 388 the GAF was 1.42, 95% CI [1.29, 1.54], and the LOCF-corrected ES was 1.42, 95% CI  
 389 [1.29, 1.54], indicating a large effect.

390 Moreover, for other secondary outcomes, Hedge's  $g$  of the change score on the GAD-  
 391 7 was .70, 95% CI [.59, .81], the LOCF-corrected ES was .69, 95% CI [.58, .81], Hedge's  $g$   
 392 of the change score on the BDI-II was .97, 95% CI [.86, 1.09], and the LOCF-corrected ES  
 393 was .97, 95% CI [.85, 1.08], indicating medium to large effects.

**394 Prediction of change in symptom-specific and functional outcome**

395 Table 2 displays the results of the variable selection by ENR on the primary  
 396 outcome, i.e., IES-R, and, respectively, on the secondary outcome, i.e., GAF. In a variable  
 397 selection procedure, the ENR identified 15 predictors for the symptom change on the IES-R  
 398 and 12 predictors for the GAF (important variables marked bold in Tab. 2). Regarding the  
 399 symptom-specific change, four sociodemographic factors, five clinical factors, and six  
 400 psychological factors were identified as predictors of treatment outcome. Among the most  
 401 stable sociodemographic factors, a higher *age at baseline* ( $\beta = -.038$ ), the *wish to retire* ( $\beta =$   
 402  $-.078$ ), and *being retired* ( $\beta = -.007$ ) seemed to negatively predict treatment outcome. In  
 403 terms of clinical factors, the *total number of diagnoses* ( $\beta = -.057$ ) seemed to be the most  
 404 stable negative predictor associated with treatment outcome. In terms of stable psychological  
 405 predictors, the *baseline score on the IES-R* ( $\beta = .207$ ), and especially its subscale *avoidance*  
 406 ( $\beta = .088$ ), seemed to positively impact treatment outcome. Further, *higher scores in*  
 407 *psychoticism* ( $\beta = -.110$ ) and *more severe depressive symptoms* ( $\beta = -.068$ ) were associated

## Predictors of treatment outcome in PTSD

19

408 with negative treatment outcome. Slightly less stable negative predictors were *work*  
 409 *disability in the previous year* ( $\beta = -.005$ ) and an *asthma diagnosis* ( $\beta = -.024$ ).  
 410 Concerning the change in general functioning, two sociodemographic factors, five clinical  
 411 factors, and five psychological factors predicted treatment outcome. The most stable  
 412 negative predictors were found to be *psychoticism* ( $\beta = -.047$ ); an *asthma diagnosis* ( $\beta = -$   
 413  $.047$ ); *anxiety symptoms* ( $\beta = -.036$ ); and *diagnoses involving patient nutrition, the*  
 414 *endocrinal, or the metabolic system* ( $\beta = -.045$ ). Slightly less stable negative predictors were  
 415 the *global psychopathological symptom severity* ( $\beta = -.027$ ), *number of children* ( $\beta = -.020$ ),  
 416 *quality of life* ( $\beta = -.014$ ), and a *comorbid personality disorder* ( $\beta = -.016$ ), while *total*  
 417 *number of diagnoses* ( $\beta = .011$ ) were positively associated to functional outcome.  
 418 All other identified predictors seemed to vary in their relation to treatment outcome (comp.  
 419 %pos in Table 2).

420 --Insert Table 2 here--

**Goodness of prediction**

422 The prediction of the extracted model on symptom change on the IES-R on the training  
 423 set was  $R^2 = .17$  and  $R^2 = .16$  on the testing set. This result indicates a stable model that, of  
 424 itself, explained a moderate amount of the variance. Prediction on functional outcome in  
 425 GAF change score was  $R^2 = .11$  on the training set and  $R^2 = .12$  on the testing set. So used,  
 426 this model has good fit but explains less variance than the IES-R model. Cross-predictions  
 427 were not able to explain variance in the testing set, indicating that different factors seem to  
 428 play an important role in clinical (measured by a disorder-specific self-rating instrument) and  
 429 functional treatment outcome (measured by an observer-rated instrument). The prediction  
 430 quality is displayed in Table 3.

431 --Insert Table 3 here--

432

**Discussion**

433 In a naturalistic, multi-site design using routine data from several clinics, this study seeks  
434 to complement current research on PTSD by including machine learning into clinical  
435 psychology research while, as suggested by Deisenhofer et al. (2018), using a disorder-  
436 specific main outcome (IES-R) and a large sample of PTSD patients. The results of this  
437 study suggest that trauma-focused psychotherapy based on CBT can be implemented  
438 effectively in routine clinical inpatient care and, further, that it might be possible to improve  
439 PTSD patients' treatment outcomes by identifying and addressing relevant prognostic  
440 variables at the beginning of treatment.

**441 Main findings and clinical implications**

442 With regard to sociodemographic and clinical variables, the sample was in general  
443 comparable in age and depressive symptom severity to another similar mental health service  
444 sample (Deisenhofer et al., 2018), but with more women in our sample (75.7% vs. 44.6%).  
445 The PTSD-specific symptom severity was only slightly lower than in a sample of inpatients  
446 with more complex forms of PTSD (IES-R; Median = 90.9, Kratzer et al., 2018).

447 The effectiveness of the routine treatment in terms of functional outcome can be  
448 classified as large, while there is space for improvement in the symptom-specific change  
449 where this study showed medium effects. With respect to other secondary outcomes, the  
450 effects for the reduction of depressive symptoms was large and the effects for the reduction  
451 of anxiety symptoms was medium. In general, despite the medium effects in symptom-  
452 specific changes, the naturalistic effectiveness of this inpatient treatment supports the  
453 successful implementation of PTSD treatment in routine clinical care (cf. Ehlers et al.,  
454 2013).

455 The high level of psychopathology, the high number and severity of comorbid disorders,  
456 and the relatively short treatment duration all may have hindered a greater symptom-specific

457 change in this sample. Further, one might cautiously hypothesize that the sample includes a  
458 large number of patients with cPTSD and those may not have been sufficiently captured in  
459 their complexity by the IES-R as a diagnostic tool leading to smaller effects. In line,  
460 psychological interventions for cPTSD showed smaller effects than for PTSD (Karatzias et  
461 al., 2019). The distinction between PTSD and cPTSD as discrete mental disorders has been  
462 validated (Böttche et al., 2018) and proposed as ICD-11 criteria for trauma-related disorders  
463 (Maercker et al., 2013). Yet, this study neither assessed the key feature of chronic exposure  
464 to interpersonal trauma nor included a cPTSD-specific measurement tool such as the ITQ  
465 (Cloitre et al., 2018). Additionally, the average length of stay of 54.3 days is considerably  
466 shorter than in other studies, e.g., 74.8 days on average in another inpatient routine clinical  
467 setting (Kratzer et al., 2018).

468 The results of the predictor analyses indicate that a combined predictor model, using  
469 sociodemographic, clinical, and psychometric scales, may contribute to predicting  
470 differential treatment outcomes for individual patients. Across the clinical and functional  
471 outcome, psychoticism, total number of diagnoses, and bronchial asthma consistently  
472 showed a stable predictive relationship to treatment outcome.

473 In particular, for better symptom-specific treatment outcome in patients with PTSD  
474 symptoms, aside from a higher degree of PTSD symptoms at intake, avoidance seems a  
475 particularly stable predictor. This finding underscores the strong evidence base for trauma-  
476 focused CBT with elements of exposure to the traumatic experience (Bisson et al., 2007;  
477 Cusack et al., 2016; Ehlers et al., 2010). Within this treatment program, the second stage of  
478 exposition-based interventions might have alleviated symptoms through cognitive changes  
479 of dysfunctional negative beliefs according to the cognitive model of PTSD (Ehlers & Clark,  
480 2000). These cognitive changes might be caused by experiences of expectation violation,  
481 which has been shown to be a crucial mechanism of change in exposure therapy (Craske,



482 Treanor, Conway, Zbozinek, & Vervliet, 2014). In this inpatient setting for example,  
483 difficulties in emotion regulation could be directly addressed when occurred to challenge the  
484 belief “I can’t control my emotions.”. Yet, these changes might have been limited due to the  
485 unique setting (i.e. controlled environment) because there are less possibilities for behavioral  
486 experiments (or self-exposure exercises) in daily life to evaluate and modify beliefs about  
487 the trauma and its consequences. In contrast, patients with a higher score on psychoticism,  
488 depression, more comorbidities, older age, being retired or wish to retire, with a medical  
489 diagnosis (bronchial asthma), and higher work disability in the previous year benefited less  
490 from treatment. These characteristics might be regarded as warning signs at the beginning of  
491 treatment. In line with other studies, especially comorbid depressive symptoms were an  
492 important predictor of poorer treatment outcomes (Richardson et al., 2014). Comorbid  
493 depression which is one of the most common comorbidities in PTSD (Bradley et al., 2005)  
494 might have hindered progress in mitigating PTSD symptoms. Therefore, it might be helpful  
495 to address depressive symptoms during treatment planning and treatment, e.g., by integrating  
496 behavioral activation and cognitive as well as interpersonal techniques. Importantly, older  
497 patients and patients already retired or with a desire to retire seem to benefit less from  
498 treatment than other patients. Taken together these predictors, PTSD exacerbation during  
499 older age and delayed onset of PTSD due to exit events and other late life stressors of normal  
500 aging (e.g., deterioration of health and losses of physical and psychological functions, death  
501 of a loved one, retirement) is a phenomenon well-known in research (Bonwick & Morris,  
502 1996; Kaup, Ruskin, & Nyman, 1994). For example, the loss of a spouse in old age is related  
503 to late life bereavement reaction and those are in turn associated with an increase in PTSD  
504 prevalence (O’Connor, 2010). Further, a study with US veterans shows a clinically  
505 significant exacerbation of PTSD symptoms in late life which might be caused by executive  
506 dysfunctions (Mota et al., 2016). In line with those findings, delayed onset PTSD symptoms

507 (Barman, Detweiler, & Kim, 2016) and worsening of existing PTSD symptoms (Mittal,  
508 Torres, Abashidze, & Jimerson, 2001) might be precipitated by cognitive decline. From a  
509 biological view, the results from a pilot study suggest that late-life PTSD may be associated  
510 with increased inflammation (Miller, Driscoll, Smith, & Ramaswamy, 2017). From a  
511 psychological view, it might be that the vulnerability to those stressful live events increase  
512 with aging because of changes in psychosocial functioning, including pretrauma  
513 symptomatology, the availability of adequate social support networks, the use of functional  
514 and dysfunctional coping strategies, and perceptions of the meaning of the traumatic event  
515 (Averill & Beck, 2000). In line, a study indicates that negative health changes, social  
516 support, and death acceptance were significantly related to current and changing PTSD  
517 symptomatology (Port, Engdahl, Frazier, & Eberly, 2002). Thus, PTSD in this age group  
518 seems to present unique features not seen in younger patients due to biological,  
519 psychological, and social changes associated with the aging process (Lapp, Agbokou, &  
520 Ferreri, 2011).

521 While psychotherapy for late life PTSD seem to be efficacious with cognitively intact  
522 older adults (Pless Kaiser, Cook, Glick, & Moye, 2019), literature suggests that older PTSD  
523 patients respond in general rather poorly to trauma-focused CBT such as prolonged exposure  
524 (Hyer & Sohnle, 2001). Until recently, researchers and practitioners called therefore for  
525 adaptations of evidence-based therapies for older patients (Kessler & Tegeler, 2018),  
526 especially for those with PTSD and co-morbidities (e.g., chronic illness, pain, sensory, or  
527 cognitive changes) where the evidence-base is quite limited so far (Pless Kaiser et al., 2019).

528 Further, a higher score in psychoticism, stronger work disability, and more  
529 comorbidities—along with comorbid depressive symptom severity—seemed to diminish  
530 treatment response. This finding is in line with that of Lonergan (2014) and indicates that

531 persons with more severe and complex PTSD need further examination in order to develop  
532 tailored approaches that will optimize treatment.

533       Interestingly, asthma was associated with a higher PTSD morbidity (Shiratori &  
534 Samuelson, 2012), while asthma patients with PTSD also had a higher rate of inpatient  
535 healthcare use (Mindlis et al., 2017). PTSD is a also risk factor for the development of  
536 asthma (Hung et al., 2019) and for the worsening of preexisting asthma (Underner et al.,  
537 2019). Even after careful adjustment for familial and genetic factors and other potential  
538 confounding factors, symptoms of PTSD were associated with an elevated prevalence of  
539 asthma (Goodwin, Fischer, & Goldberg, 2007). Further, PTSD symptoms were found to be  
540 significantly related to asthma severity, quality of life, and health care utilization (Vanderbilt  
541 et al., 2008). In line, people can develop PTSD symptoms and degrees of psychiatric co-  
542 morbid symptoms after suffering asthma attack. The severity of these symptoms relates to  
543 people's perceptions of asthma severity and alexithymia (Chung, Rudd, & Wall, 2012).  
544 Moreover, the physical symptoms of an unexpected asthma attack and PTSD bodily  
545 symptoms associated with intrusions and hyperarousal (e.g. shortness of breath) have a lot in  
546 common: they may bring about feelings of fear and helplessness enhancing loss of control by  
547 reducing self-efficacy in coping mechanism. In line, PTSD and other psychiatric symptoms  
548 in asthma populations are mediated in part by self-efficacy (Wagner, Hoelterhoff, & Chung,  
549 2017). Common physiological symptoms in PTSD patients with asthma make it difficult to  
550 distinguish in highly distressing situations (such as in exposure-based interventions) between  
551 both which might have led to a decreased benefit from treatment. An asthma diagnosis at the  
552 beginning of treatment should therefore be considered a warning sign. For this population,  
553 the standard CBT approach might be adapted to their specific needs.

554       Finally, the explained variance of the best fitted model appears satisfactory, since it  
555 seems to be clinically significant, as defined by Uher, Tansey, Malki, and Perlis (2012).

556 However, some important predictive variables might not have been assessed in routine care  
557 (e.g., expectations: Kube et al., 2018; childhood maltreatment: R.-Mercier, Masson,  
558 Bussi eres, & Cellard, 2018; treatment resistance and chronicity: Dunlop, Kaye, Youngner, &  
559 Rothbaum (2014)). Thus, the authors argue for considering these variables as part of  
560 standard assessment at the beginning of treatment.

#### 561 **Strengths and limitations**

562 First of all, this study demonstrates once more the usefulness and applicability of  
563 machine learning procedures such as ENR in clinical psychology and supports current  
564 directions in PTSD research (Deisenhofer et al., 2018).

565 As a specific strength, the study follows rigorously the upcoming trend of integrating  
566 translational machine learning approaches into clinical psychology research in order to  
567 overcome limitations of traditional statistical methods (by enhancing generalizability of  
568 results and thus increasing external validity). Specifically, generalizability was increased due  
569 to the large sample size, sound pre-processing modules to clean up data in order to reach a  
570 higher degree of homogeneity, and the sample split into a training and testing set.

571 Importantly, the study uses a disorder-specific self-report measures in PTSD, i.e., the IES-R  
572 as the main outcome measure. Lack of such an instrument was a shortcoming of previous  
573 studies (e.g., Deisenhofer et al., 2018). Although this study used a symptom-specific  
574 measurement of PTSD, the IES-R only assesses 3 clusters of PTSD as opposed to the 4  
575 clusters appearing in the DSM-5 (Association American Psychiatric, 2015). Future  
576 naturalistic practice-based studies in Germany should therefore use internationally  
577 established DSM-5 based diagnostic tools such as the PCL-5 (Kr uger-Gottschalk et al.,  
578 2017; Weathers et al., 2013) or PDS-5 (Foa et al., 2016). Further, all reported medical  
579 diagnoses were included in the analyses to satisfy the complex pathology of a severely  
580 disabled, inpatient group of PTSD patients. In addition, making use of the extensive

581 sociodemographic and clinical data, the sample could be characterized in detail; this  
582 documentation goes beyond the sample characteristics of other naturalistic routine clinical  
583 studies. As discussed above, in psychopathologic key features, this sample is comparable to  
584 those of other studies (Deisenhofer et al., 2018; Kratzer et al., 2018). Thus, the model  
585 addresses the heterogeneity and complexity of PTSD inpatients by incorporating multivariate  
586 data of several candidate predictors. Further, it seems important to assess treatment outcome,  
587 not only in terms of symptom-specific change, but also in terms of consequences of the  
588 disorder (e.g., general functional level, satisfaction with life).

589 The fact that the explained variances were comparable in training and testing sets  
590 supports the assumption that overfitting to the data was not an issue and enhances the  
591 chances of finding similar predictors in a comparable dataset. On the downside, these results  
592 must be interpreted cautiously for the following reasons:

593 This study can be considered as a Phase IV-study with high external validity but the risk  
594 of a lower internal validity. Generally, interpretation of the results is therefore limited by the  
595 exploratory and naturalistic nature of the study, i.e., the use of retrospective data, the lack of  
596 a strict manual for the PTSD treatment in the referred clinics and no assessment of therapist  
597 effects (e.g., treatment fidelity). Therefore, we do not know exactly which strategies and  
598 techniques were used in which intensity and which setting (individual or group) reducing  
599 internal validity and generalizability aspects. That is, it remains to some extent unclear,  
600 whether the predictors reliably predict outcome for other CBT protocols, delivered in  
601 different formats (e.g., group vs. individual) over different situations. Also, there was no  
602 follow-up assessment, so it remains unclear whether the predictor variable were significant  
603 outcome predictors when the patients were discharged from the clinic. Further, our study did  
604 not include a control group and can therefore not account for the impact of the natural course  
605 of symptoms. Since PTSD symptoms wax and wane in severity, often in response to a life

606 stressor, the reduction of symptoms could have been caused by both, the PTSD treatment  
607 and the fact that patients were removed from their everyday environments and placed in a  
608 low-stress inpatient environment. Of note, another limitation of the naturalistic setting lies in  
609 the problem that the diagnoses and observer-rated treatment outcomes were assessed by the  
610 respective therapist and not by an independent “blinded” evaluator. Also, within the routine  
611 clinical care setting, a large percentage of patients had to be excluded due to complete  
612 missing values on the outcome measures, because these measures were not routinely  
613 collected across all clinics and wards. Moreover, the generalizability of our findings is  
614 restricted to the nature of the treatment, that is, patients were treated in an inpatient setting  
615 and received a treatment bundle rather than a specific treatment component with associated  
616 problems such as potential influences of nonindependence created by the group treatments  
617 and multiple therapists. Although this is often done in Germany, elsewhere in the world most  
618 patients are treated in an outpatient setting. Thus, the identified predictors need to be  
619 replicated under standardized circumstances (such as RCTs to enhance internal validity),  
620 preferably in an outpatient setting (to enhance generalizability to other countries) and with a  
621 follow-up-assessment (to ensure long-term effects), also in order to differentiate those  
622 prescriptive variables which are important for one specific treatment approach.

623 Specifically, since the baseline score of the respective outcome seems to be an important  
624 predictor, ceiling and floor effects as well as effects due to regression toward the mean were  
625 difficult to estimate and might have contributed to the findings. Nevertheless, this study tried  
626 to identify those additional predictors that might be important to consider for further studies.  
627 Indeed, a strength of a naturalistic design is its closer approximation to routine clinical  
628 practice/care. Treatment selection was not randomized, however; thus, unknown patient and  
629 clinician characteristics that contributed to treatment choice may have relevance (and may  
630 also have compromised, to a degree, internal validity aspects). In our analyses, we used a

## Predictors of treatment outcome in PTSD

28

631 form of statistical replication to examine the stability of the effects across subsamples, which  
632 enhances the chance of replication of stable effects in further similar-designed studies. In  
633 addition, the use of the GAF may be criticized as a single-item instrument. However, using  
634 simple one-time global scales may be useful, given that the clinical synthesis of  
635 heterogeneous components seems often to discriminate better than complex multi-item  
636 measures (Fava, Evins, Dorer, & Schoenfeld, 2003). Despite the relevance of the ICD, the  
637 GAF is more closely able to capture relevant domains of the ICF, i.e., disability of mental  
638 disorders (Linden, 2017). Finally, this study did not assess or consider  
639 psychopharmacological interventions and therefore was not able to control for medication.  
640

**641 Conclusion and future research directions**

642 The vision of current clinical psychology research is to identify determinants of what  
643 works for whom in order to tailor interventions to specific characteristics and needs of  
644 patients. In this context, practice-oriented research aims to provide therapists with  
645 scientifically obtained criteria for their decisions as to which patient receives which  
646 treatment, whereby these criteria are in turn generated from practice, preferably through big  
647 data sets (Castonguay, Barkham, Lutz, & McAleavey, 2013). Our study tries to help close  
648 this practice-research-gap by identifying relevant predictors which could support tailoring  
649 interventions to the specific needs of PTSD patients. The proposed models might help  
650 clinicians to identify those PTSD patients at risk for poor response, that is, these identified  
651 predictors should be addressed during treatment planning (e.g., asthma monitoring, inclusion  
652 of depression elements) as a means of optimizing treatment outcome. In future research,  
653 these predictors should be replicated in confirmative prospective studies (e.g., in RCTs).  
654 Since the variance explained by the fitted models is only moderate, there might be other  
655 important predictors of treatment outcome, e.g., expectations, childhood maltreatment, and  
656 the level of treatment resistance. In addition, further research should consider incorporating  
657 brain structure and function via biological assessments (e.g., Colvonen et al., 2017) to guide  
658 treatment selection in an early phase of clinical decision and thereby derive the optimal  
659 treatment at the best time. The identification of relevant negative predictors should support  
660 tailoring interventions to the specific problems and needs of PTSD patients. In this regard,  
661 studies should differentiate between patients with PTSD and cPTSD and use adequate  
662 diagnostic tools to capture the complexity of a person's specific symptomatology and to  
663 determine treatment outcome. In future practice-based studies, it could be promising to  
664 additionally integrate the ITQ (Cloitre et al., 2018) as a cPTSD-specific instrument in the  
665 standard assessment at intake. The proposed models might help clinicians to identify those



666 PTSD patients at risk for poor response, that is, these identified symptom-specific predictors  
667 should be addressed during treatment planning (e.g., asthma monitoring, inclusion of  
668 depression elements) as a means of optimizing treatment outcome. Finding answers for the  
669 question, “What works for whom?” remain difficult, but by incorporating new statistical  
670 techniques (such as machine learning) and including previously omitted predictors (such as  
671 expectations), profound steps might be taken towards personalized, evidence-based  
672 psychotherapies and, thus, improved treatment outcomes for individual patients.

673 **Disclosure of interest**

674 The authors report no conflict of interest.

## References

- Association American Psychiatric. (2015). Diagnostisches und Statistisches Manual Psychischer Störungen DSM-5®. In P. Falkai & H.-U. Wittchen (Eds.), *Diagnostisches und Statistisches Manual Psychischer Störungen DSM-5®*.  
<https://doi.org/10.1026/02803-000>
- Averill, P. M., & Beck, J. G. (2000, March 1). Posttraumatic stress disorder in older adults: A conceptual review. *Journal of Anxiety Disorders*, Vol. 14, pp. 133–156.  
[https://doi.org/10.1016/S0887-6185\(99\)00045-6](https://doi.org/10.1016/S0887-6185(99)00045-6)
- Barman, R., Detweiler, M. B., & Kim, K. Y. (2016). PTSD in Combat Veterans With Cognitive Decline. *Federal Practitioner : For the Health Care Professionals of the VA, DoD, and PHS*, 33(Suppl 2), 17S-21S. Retrieved from  
<http://www.ncbi.nlm.nih.gov/pubmed/30766207>
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory-II*. San Antonio, TX: Harcourt Assessment Inc.
- Bisson, J. I., Ehlers, A., Matthews, R., Pilling, S., Richards, D., & Turner, S. (2007, February 2). Psychological treatments for chronic post-traumatic stress disorder: Systematic review and meta-analysis. *British Journal of Psychiatry*, Vol. 190, pp. 97–104. <https://doi.org/10.1192/bjp.bp.106.021402>
- Bonwick, R. J., & Morris, P. L. P. (1996, December 1). Post-traumatic stress disorder in elderly war veterans. *International Journal of Geriatric Psychiatry*, Vol. 11, pp. 1071–1076. [https://doi.org/10.1002/\(SICI\)1099-1166\(199612\)11:12<1071::AID-GPS511>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1099-1166(199612)11:12<1071::AID-GPS511>3.0.CO;2-B)
- Böttche, M., Ehring, T., Krüger-Gottschalk, A., Rau, H., Schäfer, I., Schellong, J., ... Knaevelsrud, C. (2018). Testing the ICD-11 proposal for complex PTSD in trauma-exposed adults: factor structure and symptom profiles. *European Journal of*

- Psychotraumatology*, 9(1), 1512264. <https://doi.org/10.1080/20008198.2018.1512264>
- Bradley, R., Greene, J., Russ, E., Dutra, L., & Westen, D. (2005). A Multidimensional Meta-Analysis of Psychotherapy for PTSD. *American Journal of Psychiatry*, 162(2), 214–227. <https://doi.org/10.1176/appi.ajp.162.2.214>
- Bzdok, D., Karrer, T. M., Habel, U., & Schneider, F. (2017). Big data approaches in psychiatry: examples in depression research. *Der Nervenarzt*, 89(8), 869–874. <https://doi.org/10.1007/s00115-017-0456-2>
- Castonguay, L. G., Barkham, M., Lutz, W., & McAleavey, A. A. (2013). Practice-oriented research: Approaches and application. In M. J. Lambert (Ed.), *Bergin & Garfield's handbook of psychotherapy and behavior change* (6th edition). New York: Wiley.
- Cho, S., Kim, K., Kim, Y. J., Lee, J.-K., Cho, Y. S., Lee, J.-Y., ... Park, T. (2010). Joint Identification of Multiple Genetic Variants via Elastic-Net Variable Selection in a Genome-Wide Association Analysis. *Annals of Human Genetics*, 74(5), 416–428. <https://doi.org/10.1111/j.1469-1809.2010.00597.x>
- Chung, M. C., Rudd, H., & Wall, N. (2012). Posttraumatic stress disorder following asthma attack (post-asthma attack PTSD) and psychiatric co-morbidity: The impact of alexithymia and coping. *Psychiatry Research*, 197(3), 246–252. <https://doi.org/10.1016/j.psychres.2012.01.008>
- Cloitre, M., Shevlin, M., Brewin, C. R., Bisson, J. I., Roberts, N. P., Maercker, A., ... Hyland, P. (2018). The International Trauma Questionnaire: development of a self-report measure of ICD-11 PTSD and complex PTSD. *Acta Psychiatrica Scandinavica*, 138(6), 536–546. <https://doi.org/10.1111/acps.12956>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates.

- Cohen, Z. D., Kim, T. T., Van, H. L., Dekker, J. J. M., & Driessen, E. (2019). A demonstration of a multi-method variable selection approach for treatment selection: Recommending cognitive-behavioral versus psychodynamic therapy for mild to moderate adult depression. *Psychotherapy Research*, 1–14.  
<https://doi.org/10.1080/10503307.2018.1563312>
- Colvonen, P. J., Glassman, L. H., Crocker, L. D., Buttner, M. M., Orff, H., Schiehsler, D. M., ... Afari, N. (2017). Pretreatment biomarkers predicting PTSD psychotherapy outcomes: A systematic review. *Neuroscience & Biobehavioral Reviews*, 75, 140–156.  
<https://doi.org/10.1016/j.neubiorev.2017.01.027>
- Cook, J. M., Schnurr, P. P., & Foa, E. B. (2004, December). Bridging the gap between posttraumatic stress disorder research and clinical practice: The example of exposure therapy. *Psychotherapy*, Vol. 41, pp. 374–387. <https://doi.org/10.1037/0033-3204.41.4.374>
- Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: an inhibitory learning approach. *Behaviour Research and Therapy*, 58, 10–23. <https://doi.org/10.1016/j.brat.2014.04.006>
- Cusack, K., Jonas, D. E., Forneris, C. A., Wines, C., Sonis, J., Middleton, J. C., ... Gaynes, B. N. (2016). Psychological treatments for adults with posttraumatic stress disorder: A systematic review and meta-analysis. *Clinical Psychology Review*, 43(290), 128–141.  
<https://doi.org/10.1016/j.cpr.2015.10.003>
- Deisenhofer, A.-K., Delgadillo, J., Rubel, J. A., Böhnke, J. R., Zimmermann, D., Schwartz, B., & Lutz, W. (2018). Individual treatment selection for patients with posttraumatic stress disorder. *Depression and Anxiety*, 35(6), 541–550.  
<https://doi.org/10.1002/da.22755>

- Dunlop, B., Kaye, J., Youngner, C., & Rothbaum, B. (2014). Assessing treatment-resistant Posttraumatic Stress Disorder: The Emory Treatment Resistance Interview for PTSD (E-TRIP). *Behavioral Sciences, 4*(4), 511–527. <https://doi.org/10.3390/bs4040511>
- Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology, 14*(1), 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>
- 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*(2), 269–276. <https://doi.org/10.1016/j.cpr.2009.12.001>
- Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder : Theory and therapy. *Behaviour Research and Therapy, 38*, 319–345.
- Ehlers, A., Grey, N., Wild, J., Stott, R., Liness, S., Deale, A., ... Clark, D. M. (2013). Implementation of Cognitive Therapy for PTSD in routine clinical care: Effectiveness and moderators of outcome in a consecutive sample. *Behaviour Research and Therapy, 51*(11), 742–752. <https://doi.org/10.1016/j.brat.2013.08.006>
- Ehlers, A., Hackmann, A., Grey, N., Wild, J., Liness, S., Albert, I., ... Clark, D. M. (2014). A randomized controlled trial of 7-day intensive and standard weekly cognitive therapy for PTSD and emotion-focused supportive therapy. *American Journal of Psychiatry, 171*(3), 294–304. <https://doi.org/10.1176/appi.ajp.2013.13040552>
- Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale. *Archives of General Psychiatry, 33*(6), 766. <https://doi.org/10.1001/archpsyc.1976.01770060086012>
- Equit, M., Maurer, S., Michael, T., & Köllner, V. (2018). Konfrontation oder Stabilisierung:

Wie planen Verhaltenstherapeuten die Behandlung bei Posttraumatischer Belastungsstörung? *Verhaltenstherapie*, 28(1), 7–14.

<https://doi.org/10.1159/000477418>

Fava, M., Evins, A. E., Dorer, D. J., & Schoenfeld, D. A. (2003). The problem of the placebo response in clinical trials for psychiatric disorders: Culprits, possible remedies, and a novel study design approach. *Psychotherapy and Psychosomatics*, 72(3), 115–127. <https://doi.org/10.1159/000069738>

Foa, E. B., Hembree, E., & Rothbaum, B. (2015). Prolonged Exposure Therapy for PTSD: Therapist Guide. In *Prolonged Exposure Therapy for PTSD: Therapist Guide*.

<https://doi.org/10.1093/med:psych/9780195308501.001.0001>

Foa, E. B., McLean, C. P., Zang, Y., Zhong, J., Powers, M. B., Kauffman, B. Y., ...

Knowles, K. (2016). Psychometric properties of the posttraumatic diagnostic scale for DSM-5 (PDS-5). *Psychological Assessment*, 28(10), 1166–1171.

<https://doi.org/10.1037/pas0000258>

Friedman, J., Hastie, T., & Tibshirani, R. (2015). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1–22.

<https://doi.org/10.18637/jss.v033.i01>

Golino, H. F., & Gomes, C. M. A. (2016). Random forest as an imputation method for education and psychology research: its impact on item fit and difficulty of the Rasch model. *International Journal of Research and Method in Education*, 39(4), 401–421.

<https://doi.org/10.1080/1743727X.2016.1168798>

Goodwin, R. D., Fischer, M. E., & Goldberg, J. (2007). A twin study of post-traumatic stress disorder symptoms and asthma. *American Journal of Respiratory and Critical Care Medicine*, 176(10), 983–987. <https://doi.org/10.1164/rccm.200610-1467OC>

- Gräfe, K., Zipfel, S., Herzog, W., & Löwe, B. (2004). Screening psychischer störungen mit dem "Gesundheitsfragebogen für Patienten (PHQ-D)". Ergebnisse der Deutschen validierungsstudie. *Diagnostica, 50*(4), 171–181. <https://doi.org/10.1026/0012-1924.50.4.171>
- Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *Journal of Educational Statistics, 6*(2), 107–128. <https://doi.org/10.3102/10769986006002107>
- Hoerl, A., & Kennard, R. (1988). Ridge regression. In *Encyclopedia of Statistical Sciences* (Vol. 8, pp. 129–136). New York: Wiley.
- Hung, Y. H., Cheng, C. M., Lin, W. C., Bai, Y. M., Su, T. P., Li, C. T., ... Chen, M. H. (2019). Post-traumatic stress disorder and asthma risk: A nationwide longitudinal study. *Psychiatry Research, 276*, 25–30. <https://doi.org/10.1016/j.psychres.2019.04.014>
- Hyer, L., & Sohnle, S. J. (2001). *Trauma among older people : issues and treatment*. Philadelphia: Brunner-Routledge.
- Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., ... Uher, R. (2016). Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *Journal of Psychiatric Research, 78*, 94–102. <https://doi.org/10.1016/j.jpsychires.2016.03.016>
- Karatzias, T., Murphy, P., Cloitre, M., Bisson, J., Roberts, N., Shevlin, M., ... Hutton, P. (2019). Psychological interventions for ICD-11 complex PTSD symptoms: systematic review and meta-analysis. *Psychological Medicine, 49*(11), 1761–1775. <https://doi.org/10.1017/S0033291719000436>
- Karlin, B. E., & Cross, G. (2014). From the laboratory to the therapy room: National dissemination and implementation of evidence-based psychotherapies in the U.S.

- Department of Veterans Affairs Health Care System. *American Psychologist*, 69(1), 19–33. <https://doi.org/10.1037/a0033888>
- Kaup, B. A., Ruskin, P. E., & Nyman, G. (1994). Significant life Events and PTSD in Elderly World War II Veterans. *American Journal of Geriatric Psychiatry*, 2(3), 239–243. <https://doi.org/10.1097/00019442-199400230-00008>
- Kessler, E.-M., & Tegeler, C. (2018). Psychotherapeutisches Arbeiten mit alten und sehr alten Menschen. *Psychotherapeut*, 63(6), 501–518. <https://doi.org/10.1007/s00278-018-0315-z>
- Kratzer, L., Heinz, P., Schennach, R., Schiepek, G. K., Padberg, F., & Jobst, A. (2018). Inpatient Treatment of Complex PTSD Following Childhood Abuse: Effectiveness and Predictors of Treatment Outcome. *PPmP Psychotherapie Psychosomatik Medizinische Psychologie*. <https://doi.org/10.1055/a-0591-3962>
- Krüger-Gottschalk, A., Knaevelsrud, C., Rau, H., Dyer, A., Schäfer, I., Schellong, J., & Ehring, T. (2017). The German version of the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): psychometric properties and diagnostic utility. *BMC Psychiatry*, 17(1), 379. <https://doi.org/10.1186/s12888-017-1541-6>
- Kube, T., Siebers, V. H. A., Herzog, P., Glombiewski, J. A., Doering, B. K., & Rief, W. (2018). Integrating situation-specific dysfunctional expectations and dispositional optimism into the cognitive model of depression – A path-analytic approach. *Journal of Affective Disorders*, 229. <https://doi.org/10.1016/j.jad.2017.12.082>
- Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>
- Lapp, L. K., Agbokou, C., & Ferreri, F. (2011, August). PTSD in the elderly: The interaction between trauma and aging. *International Psychogeriatrics*, Vol. 23, pp. 858–868.



<https://doi.org/10.1017/S1041610211000366>

- Laurin, C., Boomsma, D., & Lubke, G. (2016). The use of vector bootstrapping to improve variable selection precision in Lasso models. *Statistical Applications in Genetics and Molecular Biology*, *15*(4), 305–320. <https://doi.org/10.1515/sagmb-2015-0043>
- Linden, M. (2017). Definition and Assessment of Disability in Mental Disorders under the Perspective of the International Classification of Functioning Disability and Health (ICF). *Behavioral Sciences and the Law*, *35*(2), 124–134. <https://doi.org/10.1002/bsl.2283>
- Linehan, M. M. (1993). *Cognitive behavioral therapy of borderline personality disorder*. New York: Guilford Press.
- Lonergan, M. (2014). Cognitive Behavioral Therapy for PTSD: The Role of Complex PTSD on Treatment Outcome. *Journal of Aggression, Maltreatment & Trauma*, *23*(5), 494–512. <https://doi.org/10.1080/10926771.2014.904467>
- Maercker, A., & Schützwohl, M. (1998). Erfassung von psychischen Belastungsfolgen: Die Impact of Event Skala-revidierte Version. *Diagnostica*, *44*(3), 130–141. <https://doi.org/10.1037/t01210-000>
- Maercker, Andreas, Brewin, C. R., Bryant, R. A., Cloitre, M., van Ommeren, M., Jones, L. M., ... Reed, G. M. (2013). Diagnosis and classification of disorders specifically associated with stress: proposals for ICD-11. *World Psychiatry*, *12*(3), 198–206. <https://doi.org/10.1002/wps.20057>
- Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(4), 417–473.
- Miller, K., Driscoll, D., Smith, L. M., & Ramaswamy, S. (2017). The Role of Inflammation in Late-Life Post-Traumatic Stress Disorder. *Military Medicine*, *182*(11), e1815–e1818.

<https://doi.org/10.7205/milmed-d-17-00073>

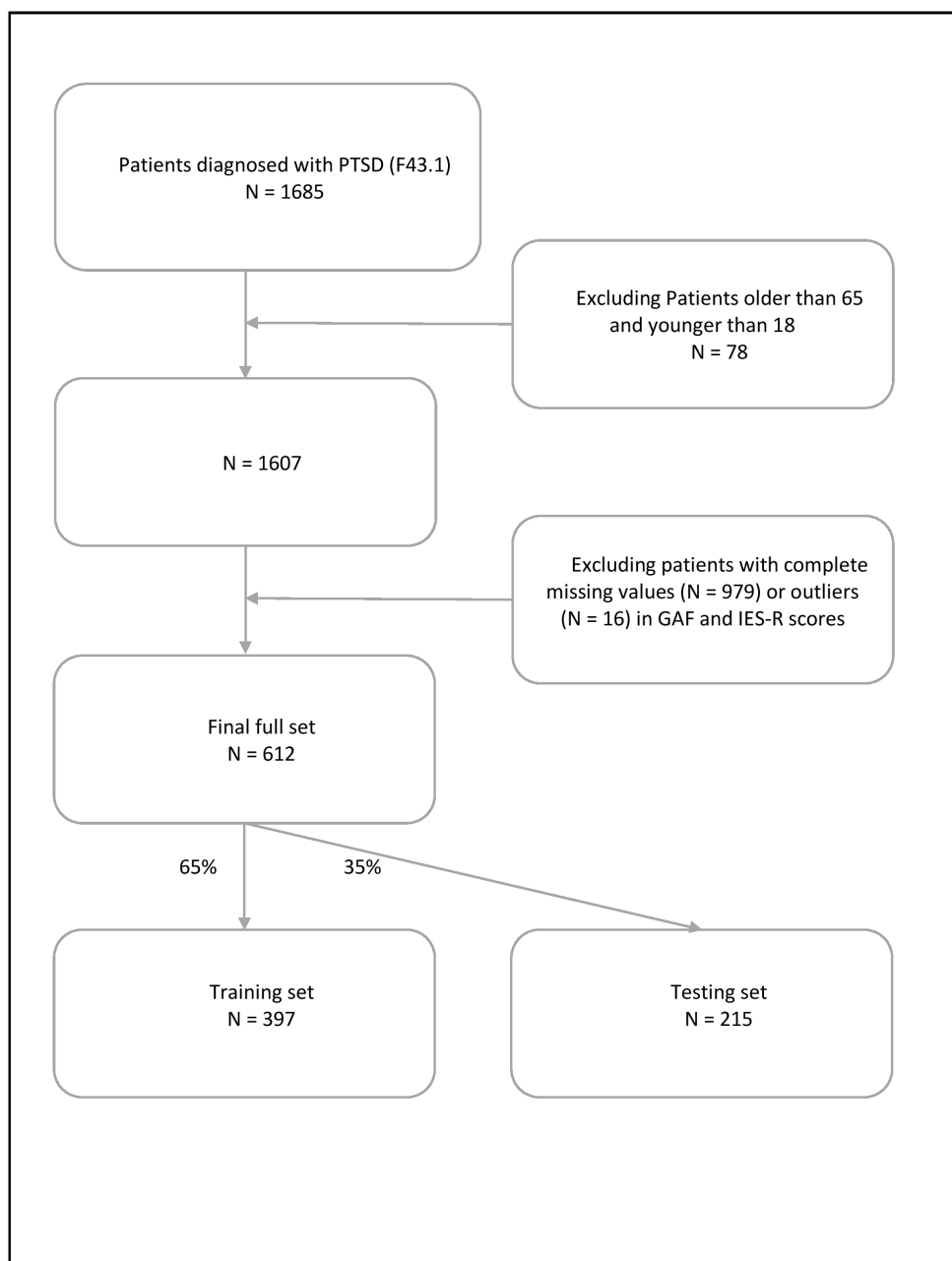
- Mindlis, I., Morales-Raveendran, E., Goodman, E., Xu, K., Vila-Castelar, C., Keller, K., ... Wisnivesky, J. P. (2017). Post-traumatic stress disorder dimensions and asthma morbidity in World Trade Center rescue and recovery workers. *Journal of Asthma*, *54*(7), 723–731. <https://doi.org/10.1080/02770903.2016.1263650>
- Mittal, D., Torres, R., Abashidze, A., & Jimerson, N. (2001). Worsening of Post-Traumatic Stress Disorder Symptoms with Cognitive Decline: Case Series. *Journal of Geriatric Psychiatry and Neurology*, *14*(1), 17–20. <https://doi.org/10.1177/089198870101400105>
- Mota, N., Tsai, J., Kirwin, P. D., Harpaz-Rotem, I., Krystal, J. H., Southwick, S. M., & Pietrzak, R. H. (2016). Late-Life Exacerbation of PTSD Symptoms in US Veterans. *The Journal of Clinical Psychiatry*, *77*(03), 348–354. <https://doi.org/10.4088/JCP.15m10101>
- Najavits, L. M. (2015). The problem of dropout from “gold standard” PTSD therapies. *F1000Prime Reports*, *7*(April), 1–8. <https://doi.org/10.12703/P7-43>
- National Institute for Clinical Excellence (NICE). (2005). *Post-Traumatic Stress Disorder: The Management of PTSD in Adults and Children in Primary and Secondary Care: National Collaborating Centre for Mental Health* (National Institute for Clinical Excellence (UK) & National Collaborating Centre for Mental Health (UK), Ed.). <https://doi.org/10.1192/pb.30.9.357-a>
- O’Connor, M. (2010). PTSD in older bereaved people. *Aging and Mental Health*, *14*(6), 670–678. <https://doi.org/10.1080/13607860903311725>
- Park, M. Y., & Hastie, T. (2008). Penalized logistic regression for detecting gene interactions. *Biostatistics*, *9*(1), 30–50. <https://doi.org/10.1093/biostatistics/kxm010>
- Pless Kaiser, A., Cook, J. M., Glick, D. M., & Moye, J. (2019, August 8). Posttraumatic

- Stress Disorder in Older Adults: A Conceptual Review. *Clinical Gerontologist*, Vol. 42, pp. 359–376. <https://doi.org/10.1080/07317115.2018.1539801>
- Port, C. L., Engdahl, B., Frazier, P., & Eberly, R. (2002). Factors Related to the Long-Term Course of PTSD in Older Ex-Prisoners of War. *Journal of Clinical Geropsychology*, 8(3), 203–214. <https://doi.org/10.1023/A:1015996211452>
- R.-Mercier, A., Masson, M., Bussi eres, E.-L., & Cellard, C. (2018). Common transdiagnostic cognitive deficits among people with psychiatric disorders exposed to childhood maltreatment: A meta-analysis. *Cognitive Neuropsychiatry*, 23(3), 180–197. <https://doi.org/10.1080/13546805.2018.1461617>
- R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna(Austria). Retrieved from <https://www.r-project.org/>
- Richardson, J. D., Contractor, A. A., Armour, C., St. Cyr, K., Elhai, J. D., & Sareen, J. (2014). Predictors of long-term treatment outcome in combat and peacekeeping veterans with military-related PTSD. *Journal of Clinical Psychiatry*, 75(11), e1299–e1305. <https://doi.org/10.4088/JCP.13m08796>
- Rizvi, S., Vogt, D., & Resick, P. (2012). Cognitive and affective predictors of treatment outcome in cognitive processing therapy and prolonged exposure for posttraumatic stress disorder. *Behavioral Research and Therapy*, 47(9), 737–743. <https://doi.org/10.1016/j.brat.2009.06.003>. Cognitive
- Sch afer, I., Gast, U., Hofmann, A., Knaevelsrud, C., Lampe, A., Liebermann, P., ... W oller, W. (2019). *S3-Leitlinie Posttraumatische Belastungsst orung*. Springer Berlin / Heidelberg.
- Schottenbauer, M. A., Glass, C. R., Arnkoff, D. B., Tendick, V., & Gray, S. H. (2008).

- Nonresponse and Dropout Rates in Outcome Studies on PTSD: Review and Methodological Considerations. *Psychiatry: Interpersonal and Biological Processes*, 71(2), 134–168. <https://doi.org/10.1521/psyc.2008.71.2.134>
- Shapiro, F. 1948-, Kierdorf, T., Höhr, H., & Mallett, D. (2013). *EMDR - Grundlagen und Praxis Handbuch zur Behandlung traumatisierter Menschen*. Junfermann.
- Shiner, B., D'Avolio, L. W., Nguyen, T. M., Zayed, M. H., Young-Xu, Y., Desai, R. A., ... Watts, B. V. (2013). Measuring use of evidence based psychotherapy for posttraumatic stress disorder. *Administration and Policy in Mental Health and Mental Health Services Research*, 40(4), 311–318. <https://doi.org/10.1007/s10488-012-0421-0>
- Shiratori, Y., & Samuelson, K. W. (2012). Relationship between posttraumatic stress disorder and asthma among New York area residents exposed to the World Trade Center disaster. *Journal of Psychosomatic Research*, 73(2), 122–125. <https://doi.org/10.1016/j.jpsychores.2012.05.003>
- Smucker, M., Reschke, K., & Kögel, B. (2008). *Imagery rescripting & reprocessing therapy: Behandlungsmanual für Typ I Trauma*. Shaker.
- Sobell, L. C. (1996). Bridging the gap between scientists and practitioners: The challenge before us. *Behavior Therapy*, 27(3), 297–320. [https://doi.org/10.1016/S0005-7894\(96\)80019-0](https://doi.org/10.1016/S0005-7894(96)80019-0)
- Sonne, C., Carlsson, J., Bech, P., Vindbjerg, E., Mortensen, E. L., & Elklit, A. (2016). Psychosocial predictors of treatment outcome for trauma-affected refugees. *European Journal Of Psychotraumatology*, 7, 30907. <https://doi.org/10.3402/ejpt.v7.30907>
- Spitzer, R., Kroenke, K., Williams, J., & Lowe B. (2006). A brief measure for assessing generalized anxiety disorder. *Arch Intern Med*, 166, 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>

- Spottswood, M., Davydow, D. S., & Huang, H. (2017). The prevalence of posttraumatic stress disorder in primary care: A systematic review. *Harvard Review of Psychiatry*, Vol. 25, pp. 159–169. <https://doi.org/10.1097/HRP.0000000000000136>
- Stekhoven, D. J., & Bühlmann, P. (2012). Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118. <https://doi.org/10.1093/bioinformatics/btr597>
- Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2), 111–147.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Boston: Pearson Education.
- Tarrier, N. (2010). The Cognitive and Behavioral Treatment of PTSD, What Is Known and What Is Known to Be Unknown: How Not to Fall Into the Practice Gap. *Clinical Psychology: Science and Practice*, 17(2), 134–143. <https://doi.org/10.1111/j.1468-2850.2010.01203.x>
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. In *Source: Journal of the Royal Statistical Society. Series B (Methodological)* (Vol. 58). Retrieved from <https://www.jstor.org/stable/pdf/2346178.pdf?refreqid=excelsior%3A19da901b23ea06f4bb03d798a533c0cb>
- Uher, R., Tansey, K. E., Malki, K., & Perlis, R. H. (2012). Biomarkers predicting treatment outcome in depression. *Pharmacogenomics*, 13(2), 233–240. <https://doi.org/10.2217/pgs.11.161>
- Underner, M., Goutaudier, N., Peiffer, G., Perriot, J., Harika-Germaneau, G., & Jaafari, N. (2019). Influence du trouble stress post-traumatique sur l'asthme. *La Presse Médicale*, 48(5), 488–502. <https://doi.org/10.1016/j.lpm.2019.03.002>

- Van Minnen, A., Arntz, A., & Keijsers, G. P. J. (2002). Prolonged exposure in patients with chronic PTSD: Predictors of treatment outcome and dropout. *Behaviour Research and Therapy, 40*(4), 439–457. [https://doi.org/10.1016/S0005-7967\(01\)00024-9](https://doi.org/10.1016/S0005-7967(01)00024-9)
- Vanderbilt, D., Young, R., MacDonald, H. Z., Grant-Knight, W., Saxe, G., & Zuckerman, B. (2008). Asthma severity and PTSD symptoms among inner city children: A pilot study. *Journal of Trauma and Dissociation, 9*(2), 191–207. <https://doi.org/10.1080/15299730802046136>
- Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics, 7*, 91. <https://doi.org/10.1186/1471-2105-7-91>  
PM - 16504092 M4 - Citavi
- Wagner, E. H., Hoeltherhoff, M., & Chung, M. C. (2017). Posttraumatic stress disorder following asthma attack: the role of agency beliefs in mediating psychiatric morbidity. *Journal of Mental Health, 26*(4), 342–350. <https://doi.org/10.1080/09638237.2017.1340628>
- Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). The PTSD Checklist for DSM-5 (PCL-5). *National Center for PTSD, 5*(August), 2002. <https://doi.org/10.1037/t02622-000>
- Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science, 12*(6), 1100–1122.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>

*Figure 1.* Patient flow diagram

## PREDICTORS OF TREATMENT OUTCOME IN PTSD

45

Table 1  
*Sample characteristics of 612 inpatients with PTSD*

Characteristics	Patients (n=612)
Age at entry <i>M</i> (SD)	42.3 (11.6)
Sex <i>n</i> (%)	
male	149 (24.3)
female	463 (75.7)
Number of mental disorders <i>M</i> (SD) <sup>a</sup>	2.7 (1.1)
Most frequent mental comorbidities <i>n</i> (%) <sup>a</sup>	
Depressive episode (F32)	125 (20.4)
Recurrent depressive disorder (F33)	423 (69.1)
Pain disorder related to psychological factors (F45.4)	64 (10.5)
Personality disorder (F60)	63 (10.3)
Housing Situation <i>n</i> (%)	
Alone	125 (20.4)
Co-living with partner or family	269 (44.0)
Co-living with parents	31 (5.1)
Institutional placement	4 (.7)
Shared flat, private flat, furnished room	162 (26.5)
No fixed household	1 (.3)
Flat-sharing community	18 (2.9)
Marital status <i>n</i> (%)	
Married	262 (42.8)
Divorced	118 (19.3)
Widowed	18 (2.9)
Single	214 (35.0)
In a relationship <i>n</i> (%)	292 (47.7)
Educational score <i>n</i> (%) <sup>b</sup>	3.0 (0.9)
Occupational status <i>n</i> (%)	
Unemployed	132 (21.6)
Retired	160 (26.1)
Student, in training, home care	52 (8.5)
Working full time	170 (27.8)
Working half time	84 (13.7)
Working occasionally	7 (1.1)
Ability to work <i>n</i> (%)	131 (21.4)
Wish to retire <i>n</i> (%)	96 (15.7)
First inpatient treatment <i>n</i> (%)	217 (42.8)
Outpatient psychotherapy <i>n</i> (%)	420 (68.6)
Outpatient psychiatric treatment <i>n</i> (%)	460 (75.3)
IES-R score at baseline <i>M</i> (SD) <sup>c</sup>	77.8 (16.8)
PHQ-9 score at baseline <i>M</i> (SD) <sup>d</sup>	17.2 (5.2)
BDI-II score at baseline <i>M</i> (SD) <sup>e</sup>	34.9 (11.1)
GAF score at baseline <i>M</i> (SD) <sup>f</sup>	44.5 (8.3)
BSI mean score at baseline <sup>g</sup>	1.9 (.7)
SWLS score at baseline <sup>h</sup>	13.4 (6.3)

*Note.* *M*=mean, *SD*=Standard deviation, *n*=number

<sup>a</sup> Diagnosis as given by practitioners according to ICD-10; <sup>b</sup> Based on the German school system; scale from 0 (no degree) to 4 (general qualification for university entrance); <sup>c</sup> IES-R = Impact of Events Scale – Revised: 22 items, scale 0-110; <sup>d</sup> PHQ-9 = Patient Health Questionnaire-9: 9 items scale 0-27; <sup>e</sup> BDI-II = Beck Depression Inventory-II: 21 Items, scale 0-63; <sup>f</sup> Global Assessment of Functioning: scale 0-100; <sup>g</sup> Brief Symptom Inventory: 53 items, scale 0-4; <sup>h</sup> Satisfaction With Life Scale: 5 Items, scale 5-35



## PREDICTORS OF TREATMENT OUTCOME IN PTSD

46

Table 2  
*Prediction of change in symptom specific and functional scores*

Predictors	Criterion: IES-R improvement			Criterion: GAF improvement		
	$\beta$	%sel	%pos	$\beta$	%sel	%pos
<b>Baseline IES-R</b>	<b>.207</b>	98.0	<b>100**</b>	.	12.7	78.7
<b>(BSI) Psychoticism</b>	<b>-.110</b>	93.5	<b>.1**</b>	<b>-.047</b>	73.8	<b>0.3**</b>
<b>(IES-R) Avoidance</b>	<b>.088</b>	96.7	<b>100**</b>	.	38.4	73.4
<b>Wish to retire</b>	<b>-.078</b>	97.4	<b>.4**</b>	.	44.1	66.2
<b>Baseline BDI-II</b>	<b>-.068</b>	87.3	<b>.2**</b>	.	22.1	17.1
<b>Number of comorbid diagnoses</b>	<b>-.057</b>	96.2	<b>.2**</b>	<b>.011</b>	68.0	<b>97.9*</b>
<b>Age at baseline</b>	<b>-.038</b>	94.6	<b>.1**</b>	.	32.8	58.2
<b>Bronchial Asthma<sup>1</sup></b>	<b>-.024</b>	86.0	<b>1.3*</b>	<b>-.047</b>	88.2	<b>.0**</b>
(PHQ-15) Physical symptoms	-.015	65.5	3.5	.	42.7	7.5
Outpatient psychiatric care	-.014	74.6	8.6	.	44.6	78.9
Children	.010	80.4	9.9	.	61.5	0.3
<b>Being retired</b>	<b>-.007</b>	87.5	<b>.2**</b>	.	31.7	31.9
<b>Work disability last year</b>	<b>-.005</b>	73.9	<b>1.2*</b>	.	24.6	43.1
Outpatient psychotherapy	-.004	77.8	4.8	.	45.2	29.2
(BSI) Somatization	-.003	50.7	13.6	.	37.3	13.4
(GAF) General Functioning	.	68.8	88.1	-.167	99.9	.
<b>Endocrine, nutritional and metabolic disease<sup>2</sup></b>	.	70.3	96.7	<b>-.045</b>	84.7	<b>.2**</b>
<b>(BSI) Anxiety</b>	.	37.1	12.7	<b>-.036</b>	67.3	<b>.1**</b>
<b>(BSI) PST Score<sup>3</sup></b>	.	34.9	76.5	<b>-.027</b>	52.2	<b>1.0*</b>
<b>Number of Children</b>	.	47.1	73.7	<b>-.020</b>	66.3	<b>2.4*</b>
<b>Personality Disorder<sup>4</sup></b>	.	38.8	13.7	<b>-.016</b>	53.6	<b>2.1*</b>
<b>(SWLS) Quality of Life</b>	.	66.0	9.5	<b>-.014</b>	69.1	<b>1.7*</b>
Duration of work disability	.	48.6	44.7	-.002	65.4	95.4
Obesity <sup>5</sup>	.	52.8	21.4	-.001	70.3	96.7

*Note.* <sup>1</sup> Section J45 according to ICD-10; <sup>2</sup> Section E according to ICD-10; <sup>3</sup> PST Score = Positive Symptom Total = Total Number of distressing symptoms; <sup>4</sup> Section F60 according to ICD-10; <sup>5</sup> Diagnosis E66.0 according to ICD-10; BDI-II = Beck Depression Inventory II; BSI = Brief Symptom Inventory; IES-R = Impact of Events Scale – Revised; PHQ = Patient Health Questionnaire; pst = positive symptom total scale of BSI; SWLS = Satisfaction With Life; GAF = Global Assessment of Functioning; Outpatient  $\beta$  = elastic net coefficients; variables with zero  $\beta$  coefficients were not included in final model; %sel = percentage of bootstrapped subsamples, in which the predictor was selected; %pos = percentage of positive coefficients in bootstrapped subsamples in which the predictor was selected; \*/\*\* = consistent sign in more than 97.5%/99.5% of the models a predictor was selected in.

## PREDICTORS OF TREATMENT OUTCOME IN PTSD

47

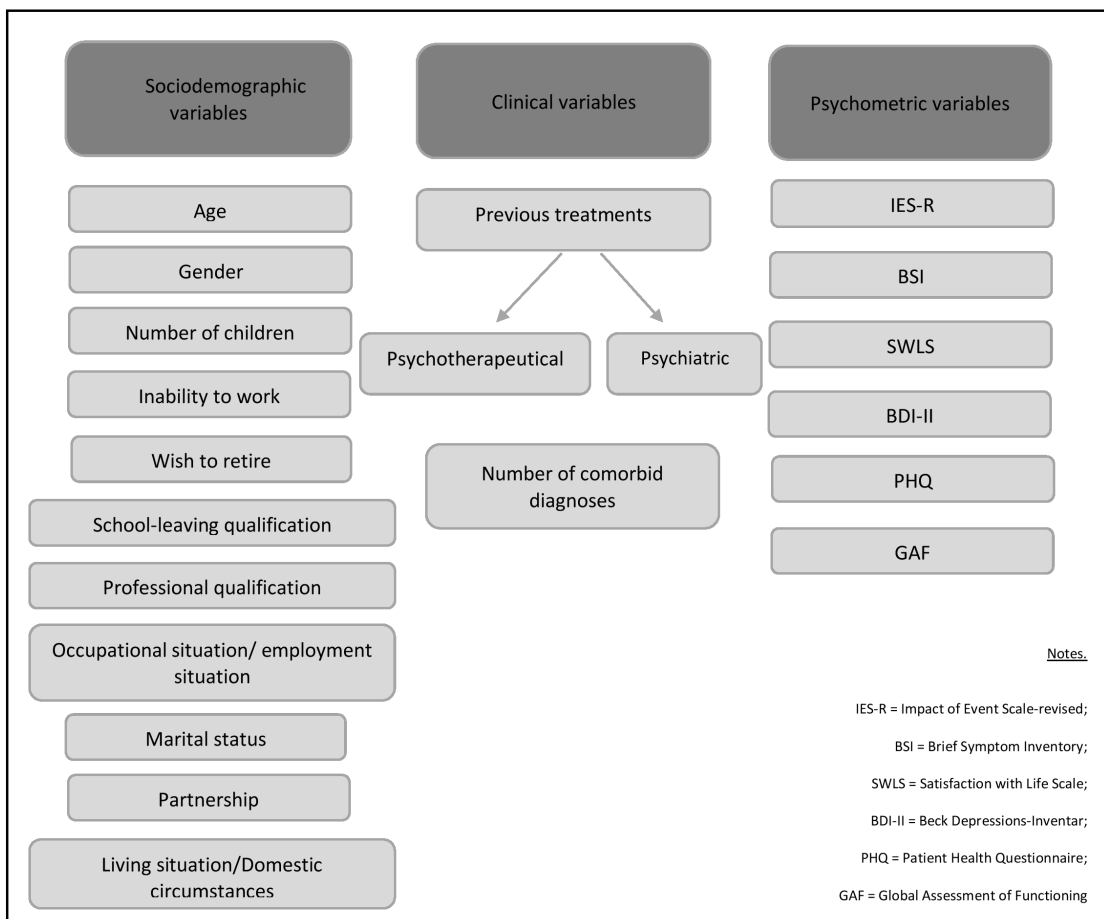
Table 3  
*Prediction quality on difference scores of clinical and functional outcome*

Prediction	Criterion IES-R			Criterion: GAF		
	$R^2$	$MAE$	$RMSE$	$R^2$	$MAE$	$RMSE$
On training set	.17	.69	.91	.11	.68	.89
On testing set	.16	.77	.95	.12	.82	1.08
Cross prediction on testing set	.00	.81	1.02	.00	.88	1.15

*Note.*  $R^2$  = coefficient of determination;  $MAE$  = Mean average error;  $RMSE$  = Root mean squared error

Supplementary Material

Figure 2. Candidate predictors



**Description of Candidate predictors (psychometric variables)**

The Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983) consists of 53 items and is the short version of the SCL-90-R (Franke, 1995). This instrument assesses nine primary symptom dimensions: Somatization (SOM), Obsessive-Compulsiveness (OBS), Interpersonal Sensitivity (INS), Depression (DEP), Anxiety (ANX), Hostility (HOS), Phobic Anxiety (PHOB), Paranoid Ideation (PAR), and Psychoticism (PSY). Additionally, global indices of distress can be calculated, most importantly the General Symptom Index (GSI). Validation studies (e.g., Franke, 1997) of the German version have shown a satisfactory test-retest reliability.

The Satisfaction With Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985) assesses satisfaction with one's life in general. The five-item scale was designed to measure a person's global judgment of life satisfaction and does not allow judgments in life domains (e.g., health, finances). The SWLS has shown good psychometric properties, high internal consistency, high temporal stability, good convergent validity with other scales of subjective well-being, and some evidence of its discriminant validity with scales of emotional well-being (Glaesmer, Grande, Braehler, & Roth, 2011). Moreover, the SWLS has sufficient sensitivity to be useful in detecting changes in life satisfaction during the course of treatment (Pavot, 1993).

The PHQ (Gräfe et al., 2004) is a self-administered questionnaire consisting of the PHQ-9 assessing depression, the GAD-7 measuring generalized anxiety disorder, and the PHQ-15 administering somatic symptoms.

**Summary table of the behavioral treatment program***Three stages of the behavioral treatment program in routine inpatient care*

<b>Stage 1: Motivation and Psychoeducation</b>	<b>Stage 2: Exposure therapy</b>	<b>Stage 3: Transfer und relapse prevention</b>
<b>Individual psychotherapy</b>	<b>Individual psychotherapy</b>	<b>Individual psychotherapy</b>
<b>Trauma-focused psychoeducation group</b>	<b>Trauma-focused psychoeducation group</b>	<b>Home Exposition</b>
<b>Emotional skills Training</b>	<b>Exposure therapy (accompanied + self- controlled)</b>	<b>Self-controlled expositions and behavior training</b>
<b>Interactional ward groups</b>	<b>Interactional ward groups</b>	<b>Interactional ward groups</b>
<b>Social Skills Training / Assertiveness Traing</b>		<b>Inclusion of relatives and partners</b>
<b>Access to (according to individual indication): Mindfulness training, relaxation training, sports- and body therapy groups, trauma-focused art therapy, social counseling, Biofeedback, symptom-specific group therapy for comorbid disorders (e.g. depression or social phobia)</b>		

## Anhang C: Studie 3

## PREDICTORS OF TREATMENT OUTCOME IN OCD

1

1

2 **Title:** Determining prognostic variables of treatment outcome in obsessive-compulsive  
3 disorder: Effectiveness and its predictors in routine clinical care

4

**Short title:**

5

Predictors of treatment outcome in OCD

6

Philipp Herzog\*,

7

Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,

8

Gutenbergstraße 18, D-35032 Marburg, Germany,

9

Bernhard Osen,

10

Birkenweg 10, Psychosomatic Clinic,

11

Birkenweg 10, D-24576 Bad Bramstedt, Germany

12

Christian Stierle,

13

Birkenweg 10, Psychosomatic Clinic,

14

Birkenweg 10, D-24576 Bad Bramstedt, Germany

15

Thomas Middendorf,

16

Schön-Klinik Bad Arolsen, Psychosomatic Clinic,

17

Hofgarten 10, D-34454 Bad Arolsen, Germany

18

Ulrich Voderholzer,

19

Schön-Klinik Roseneck, Psychosomatic Clinic,

20

Am Roseneck 6, D-83209 Prien am Chiemsee, Germany

PREDICTORS OF TREATMENT OUTCOME IN OCD

2

21

Stefan Koch,

22

Schön-Klinik Roseneck, Psychosomatic Clinic,

23

Am Roseneck 6, D-83209 Prien am Chiemsee, Germany

24

Matthias Feldmann,

25

Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,

26

Gutenbergstraße 18, D-35032 Marburg, Germany,

27

Winfried Rief,

28

Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,

29

Gutenbergstraße 18, D-35032 Marburg, Germany,

30

and Eva-Lotta Brakemeier,

31

Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,

32

Gutenbergstraße 18, D-35032 Marburg, Germany

33

\*Correspondence to be addressed to: Philipp Herzog, Philipps-University of Marburg,

34

Department of Clinical Psychology and Psychotherapy, Gutenbergstraße 18, D-35032

35

Marburg, Germany, Email: philipp.herzog@staff.uni-marburg.de, Telephone: +49 6421

36

2824061

37

**Acknowledgments**

38

None.

39

**Contributions**

40

All authors materially participated in the research and/or article preparation. All authors have

41

approved the final article.

42

**Declarations of interest**

43

The authors report no conflict of interest.

44

**Funding**

45

The authors received no funding from an external source.

PREDICTORS OF TREATMENT OUTCOME IN OCD

3

46

**Highlights**

- 47 ○ The naturalistic effectiveness of inpatient psychotherapy for OCD is large.
- 48 ○ This might serve as a benchmark of OCD treatment in routine inpatient care.
- 49 ○ Social support and washing compulsions are positive general predictors of outcome.
- 50 ○ There are specific predictor profiles for changes in obsessions and compulsions.
- 51 ○ The identification of those prognostic variables helps to optimize treatments.



52

**Abstract**

53 The objectives of this study were to investigate the naturalistic effectiveness of routine  
54 inpatient treatment for patients with obsessive-compulsive disorder (OCD) and to identify  
55 predictors of treatment outcome. A routinely collected data set of 1,596 OCD inpatients ( $M$   
56 = 33.9 years,  $SD$  = 11.7; 60.4% female) having received evidence-based psychotherapy  
57 based on cognitive-behavioral therapy (CBT) in five German clinics was analyzed. Effect  
58 sizes (Hedges'  $g$ ) were calculated for several outcome variables to determine effectiveness.  
59 Predictor analyses were performed on a subsample ( $N$  = 514;  $M$  = 34.3 years,  $SD$  = 12.2;  
60 60.3% female). For this purpose, the number of potential predictors was reduced using factor  
61 analysis, followed by multiple regression analysis to identify robust predictors. Effect sizes  
62 of various outcome variables could be classified as large ( $g$ =1.34 of OCD-symptom change).  
63 Predictors of changes in OCD and depressive symptoms were symptom severity at  
64 admission and general psychopathological distress. In addition, patients with higher social  
65 support and more washing compulsions benefited more from treatment. Subgroup analyses  
66 showed a distinct predictor profile of changes in compulsions and obsessions. The results  
67 indicate that an evidence-based psychotherapy program for OCD can be effectively  
68 implemented in routine inpatient care. In addition to well-established predictors, social  
69 support and washing compulsions in particular were identified as important positive  
70 predictors. Specific predictor profiles for changes in obsessions and compulsions are  
71 discussed.

72

**Keywords**

73 Obsessive-compulsive disorder, effectiveness, predictors, patient characteristics, prognostic  
74 variables

75 **Introduction**

76 Obsessive-compulsive disorder (OCD) is a serious and debilitating disease (Ruscio et al.,  
77 2010; Stengler-Wenzke et al., 2006). Patients suffering from OCD are usually treated with  
78 cognitive-behavioral therapies (CBT) which usually refer to exposure with response  
79 prevention (ERP), cognitive therapy (CT) and combinations thereof. CBT have been shown  
80 to be very effective, with large effect sizes (ES) ranging from 1.31 compared with waiting-  
81 list, 1.33 compared with placebo conditions, and .55 compared with antidepressant  
82 medication (Öst et al., 2015). No significant differences have been found between ERP and  
83 CT (Öst et al., 2015), suggesting that they alleviate symptoms in OCD patients to the same  
84 extent. In general, a combination of CBT and medication appears to be no more effective  
85 than CBT plus a placebo (Öst et al., 2015), although better results can be achieved for severe  
86 OCD with a combination treatment (Skapinakis et al., 2016).

87 These findings lead to the conclusion that ERP and CT can be considered the “gold  
88 standard” in the treatment of OCD. However, although CBT is considered a first line,  
89 evidence-based psychological treatment in OCD, its integration into daily clinical practice is  
90 hampered by several problems, e.g., patients’ lack of motivation, limited numbers of  
91 clinicians qualified to provide it, and treatment delivery challenges such as alleged  
92 organizational difficulties (McKay et al., 2015; Moritz et al., 2019). This research-practice  
93 gap is supported by studies showing that less than half of therapists regularly use ERP (Kütz  
94 et al., 2010) and about 21% of psychotherapists trained in CBT never or rarely use ERP in  
95 the outpatient treatment of OCD (Roth et al., 2004). In line with this, from the patients’  
96 perspective, two fifths of OCD patients in one study reported that they had never received  
97 CBT with ERP (Voderholzer et al., 2015). Thus, the consequent use of ERP in an inpatient  
98 setting appears to be a promising approach, since a higher treatment dose can be delivered by  
99 several therapists.

## PREDICTORS OF TREATMENT OUTCOME IN OCD

6

100 Throughout Germany, inpatient treatment is provided to OCD patients who are  
101 considered to be the most severely treatment-resistant – a difficult-to-treat population – and  
102 for whom no outpatient treatment is available. However, the question of whether CBT for  
103 OCD can be effectively used in routine clinical care has not been fully clarified. One study  
104 has shown that CBT in a routine clinical inpatient setting can achieve large effect sizes for  
105 multi-morbid, pre-treated OCD patients (Gönner et al., 2012). Furthermore, the mean  
106 dropout rate from OCD treatment in controlled studies is about 15% (Eddy, Dutra, Bradley,  
107 & Westen, 2004), with CT exhibiting the lowest rate (11.4 %) and ERP (19.1%) and  
108 combined ERP and antidepressant medication (32.0%) showing the highest (Öst et al.,  
109 2015). Refusal rates vary widely, depending on the study (Öst et al., 2015). Even where  
110 patients receive the “gold standard,” about 30% appear not to benefit from treatment in terms  
111 of non-response (Schruiers et al., 2005). Only 50%–60% of OCD patients experience  
112 clinically meaningful changes, while about 75% report residual symptoms (Fisher & Wells,  
113 2005). Therefore, in addition to the above-mentioned effectiveness question, it would seem  
114 promising to identify stable predictors of treatment outcome (i.e., prognostic variables) to  
115 determine patients at risk for non-response.

116 Two reviews have identified some important predictors which, at first glance, lead to  
117 similar results. In the first, symptom severity, symptom subtype, severe depression, presence  
118 of comorbid personality disorders, family dysfunction, and the therapeutic alliance were  
119 found to be among the most consistent and important predictors (Keeley et al., 2008). The  
120 other review predicted poorer treatment outcomes as a result of hoarding pathology, higher  
121 anxiety and symptom severity, certain symptom subtypes, unemployment, and being  
122 single/not married (Knopp et al., 2013). Similar findings were reported by Keeley et al.  
123 (2008).

## PREDICTORS OF TREATMENT OUTCOME IN OCD

7

124 In contrast with the above-mentioned reviews, original studies sometimes produce  
125 different results because they investigate different variables as predictors (which in turn  
126 leads to the garbage-in-garbage-out or apple-with-oranges problems of meta-analyses and  
127 reviews).

**128 Sociodemographic Variables as Predictors**

129 The literature on important sociodemographic predictors appears sparse and need to be  
130 validated by further studies. However, with regard to such variables, being married appears  
131 to positively predict treatment outcome (Steketee et al., 1999). In line with this, in another  
132 study being married or cohabiting consistently predicted better outcomes in an inpatient  
133 setting (Boschen et al., 2010). Other studies have argued that this result depends on the  
134 quality of social support: patients who perceived their relatives to be either critical or hostile  
135 had more severe OCD symptoms (Van Noppen & Steketee, 2009).

**136 Clinical Variables as Predictors**

137 **Symptom Severity.** With regard to important clinical variables, lower OCD symptom  
138 severity at admission has consistently been found across studies to be a positive predictor of  
139 treatment outcome (Farrell & Boschen, 2011; Kempe et al., 2007; Kyrios et al., 2015; Raffin  
140 et al., 2009; Steketee et al., 1999, 2011). Moreover, higher scores on psychoticism, the  
141 presence of obsessions, and high general psychological distress have been found to be  
142 predictors of non-response (Kohls et al., 2002).

143 **OCD subtypes and comorbid disorders.** Furthermore, some inconsistencies occur  
144 when other relevant clinical variables (for example, OCD subtype and the presence of  
145 comorbid Axis I and Axis II diagnoses) are examined more closely.

146 Hoarding symptoms are consistently found to be predictors of worse treatment outcomes  
147 (Abramowitz et al., 2003; Eisen et al., 2013; Maher et al., 2012; Veale et al., 2016), and in  
148 the long run too (Jakubovski et al., 2013). Yet, while one study found that the presence of

## PREDICTORS OF TREATMENT OUTCOME IN OCD

8

149 sexual obsessions predicted poorer treatment response (Farrell & Boschen, 2011), another  
150 indicated conversely that the presence of sexual (but not religious) OCD symptoms predicted  
151 improvement (Steketee et al., 2011).

152 Some studies have found that more Axis I comorbid diagnoses (mainly major depressive  
153 disorder and anxiety disorders) and more depressive symptoms in general are linked to a  
154 higher chance of improvement (Kempe et al., 2007; Steketee et al., 2011). In contrast, OCD  
155 patients with higher depressive symptom severity have been found to show a moderate  
156 treatment outcome but significantly less improvement compared with that of less depressed  
157 patients (Abramowitz et al., 2000; Abramowitz & Foa, 2000). In accordance with this, the  
158 presence of at least one comorbid disorder and a depressive disorder also appears to predict  
159 poorer long-term outcomes (Jakubovski et al., 2013). In particular, patients with generalized  
160 anxiety disorder (GAD) and/or panic disorder comorbidity showed worse treatment  
161 outcomes post-treatment (Hansen et al., 2007). Similarly, GAD was also found to be a  
162 predictor for dropout, whereas major depressive disorder (MDD) and all three personality  
163 clusters predicted worse post-treatment outcomes, as did MDD and GAD in the longer-term  
164 (Steketee et al., 2001). An obsessive-compulsive personality disorder (OCPD) diagnosis and  
165 greater OCPD severity was also observed to predict worse treatment outcomes (Eisen et al.,  
166 2013; Pinto et al., 2011) as well as positive schizotypal symptoms did (Moritz et al., 2004).  
167 Conversely, other studies have found that personality disorders generally do not influence  
168 treatment outcome (Dreessen et al., 1997) or long-term treatment outcome (Fricke et al.,  
169 2006). Another study also found that neither the presence nor the number of pre-treatment  
170 comorbid disorders predicted symptom severity, treatment response, remission, or clinically  
171 significant change post-treatment (Storch et al., 2010). And finally, in another comorbid  
172 alcohol misuse appeared to predict worse outcomes (Veale et al., 2016). In summary, the  
173 role of OCD subtypes, comorbid disorders (Axis I and II), and in particular depressive

## PREDICTORS OF TREATMENT OUTCOME IN OCD

9

174 symptoms, as positive or negative predictors remains contentious and consistent findings  
175 lacking.

176 **Chronicity and Onset.** Looking at other clinical variables (for example, chronicity and  
177 onset), a longer duration of illness (Eisen et al., 2013) has been found to predict worse  
178 treatment outcomes, while early age of onset has been linked to a higher chance of remission  
179 (Kempe et al., 2007). One study identified different predictors of treatment outcome in early  
180 and late OCD (EOCD, LOCD): in EOCD, high motivation and high initial symptom severity  
181 were associated with better treatment outcomes, while higher age, longer duration of  
182 psychiatric inpatient treatment before assessment, and low level of social functioning were  
183 associated with worse outcomes; for LOCD, living in a stable relationship, high motivation,  
184 and completing treatment predicted better treatment outcomes, while a low level of  
185 psychological functioning and a longer duration of inpatient psychiatric treatment before  
186 assessment was associated with more undesirable outcomes (Langner et al., 2009).

187 **Obsessions vs. Compulsions.** Moreover, closer inspection suggests a pronounced  
188 predictor profile of poorer outcomes in obsessions and compulsions: greater initial severity  
189 of complaints and depression predicted worse outcomes for compulsions; while greater  
190 initial severity of complaints, higher levels of depression, longer problem duration, poorer  
191 motivation for treatment, and dissatisfaction with the therapeutic relationship predicted  
192 worse outcomes for obsessions (Keijsers et al., 1994).

193 In summary, a number of studies have identified specific predictors of outcome, but to  
194 date no comprehensive models of robust predictors of treatment outcome that include  
195 sociodemographic, clinical, and psychological variables have been identified. Although a  
196 significant number of studies report a relationship between demographic and clinical  
197 variables and treatment outcome, the results are to some extent inconsistent, leading to the  
198 conclusion that reliable evidence on predictors of treatment outcome in OCD is still lacking.

## PREDICTORS OF TREATMENT OUTCOME IN OCD

10

199 The main limitations of previous studies are their lack of power due to small sample sizes,  
200 and their focus on either sociodemographic or clinical or psychological variables alone.  
201 Also, research indicates that reducing the pool of potential predictors prior to data analyses  
202 might be useful because an increased number of predictors could lead to more uncertainty  
203 regarding an individual predictor, especially but not limited to standard linear regression  
204 models (Riley, Snell, Ensor, Burke, Harrell Jr, et al., 2019; Riley, Snell, Ensor, Burke,  
205 Harrell, et al., 2019). In this study, we try to overcome both these shortcomings. Therefore,  
206 the aims of the current explorative study are: (1) to determine the effectiveness of inpatient  
207 OCD treatment in routine clinical care; and (2) to examine the predictive relationship of  
208 sociodemographic, clinical, and psychological variables in a single model of treatment  
209 outcome by using novel statistical approaches to reduce the number of variables.

**210 Methods**

211 This study was conducted in accordance with the ethical standards as laid down in the  
212 1964 Declaration of Helsinki and its subsequent amendments.

**213 Sample**

214 For this study, we were able to analyze data routinely collected between 2013 and 2017  
215 from OCD inpatients treated in five German clinics offering specialized inpatient treatment.  
216 In the German mental healthcare system, specialized inpatient treatment is given to OCD  
217 patients when outpatient treatment fails or is not available. The patients in these clinics  
218 completed various self-report questionnaires at admission and at discharge and were  
219 interviewed by raters. All patients gave informed consent to anonymous evaluations of their  
220 routinely collected data. The present study included all patients consecutively diagnosed  
221 with OCD (F42 according to ICD-10; Dilling, Mombour, & Schmidt, 2010) by practitioners  
222 between 2013 and 2017, who met the following inclusion and exclusion criteria. The  
223 inclusion criteria were having a primary diagnosis of OCD, the focus of treatment being on

## PREDICTORS OF TREATMENT OUTCOME IN OCD

11

224 OCD, and being 18–65 years of age. The exclusion criteria were having comorbid substance  
225 abuse and/or a comorbid schizophrenia. The reliability of the diagnosis of OCD was assured  
226 by a trained therapist and Y-BOCS scores <16 for F42.2 and <10 for F42.0 and F42.1, and  
227 further reassured by the head of the unit. To deal with the heterogeneity of OCD patients, no  
228 further exclusion criteria were defined a priori.

229 A total of 1,595 consecutive patients from four clinics met the inclusion criteria and were  
230 used to calculate effect size. A subsample of only 514 patients from one clinic could be  
231 included in the final analysis, since data on the relevant potential predictors were not  
232 available (in particular, there was no data on the Obsessive-Compulsive Inventory -  
233 Revised)<sup>1</sup> for many of the patients. At first glance, it appeared that the subsample was  
234 different on several important sociodemographic and clinical variables (e.g., Y-BOCS  
235 baseline score, OCD subtype), as indicated by significant differences. However, most of  
236 these differences could be explained by the large sample size and further visual inspection  
237 revealed no clinically relevant differences. Thus, the subsample's sociodemographic and  
238 clinical variables did not differ in a clinically meaningful manner from those of the full  
239 sample (see Table 1). The patient flow diagram is shown in Figure 1.

240 Insert Figure 1 here.

241 Relevant sociodemographic and clinical characteristics of the full and the OCI-R sample  
242 are displayed in Table 1. Exemplarily, the relevant sociodemographic and clinical features of  
243 the OCI-R sample are described below:

244 The OCI-R sample consisted of 514 patients with a mean age of 34.3 years (SD=12.2),  
245 60.3% of whom were female. Almost half the sample were co-habiting with a partner or  
246 family (44.0%), about one quarter shared a flat (26.5%) and one fifth lived alone (20.4%).

---

<sup>1</sup> This subsample is called the "OCI-R sample" in the following paragraphs.



## PREDICTORS OF TREATMENT OUTCOME IN OCD

12

247 The mean education score according to the German school system was 3.3, indicating a  
248 relatively highly-educated patient sample. While 60.5% patients were not married, nearly  
249 half (48.9%) were in a permanent relationship. One quarter (25.1%) were unemployed,  
250 30.5% patients worked full-time, and only 45.7% were able to work. For one-third (33.9%),  
251 this was their first inpatient treatment. Before beginning this inpatient treatment, 72.6% had  
252 had one kind of outpatient psychotherapy and 67.0% had had one outpatient psychiatric  
253 treatment.

254 According to the ICD-10 and regarding OCD subtype a, 68.9% of the sample had mixed  
255 thoughts and actions. On average, patients exhibited 1.4 mental comorbidities, depressive  
256 disorders being among the most frequent. The Y-BOCS mean score was 24.7 and the OCI-R  
257 score 30.2, both indicating a moderate to severely disabled sample. In addition, the PHQ-9  
258 mean score was 12.5, indicating moderate depressive symptoms, while general  
259 psychopathology was high as indicated by a GSI (mean BSI score) of 1.35 ( $SD = .70$ )  
260 compared with the mean GSI score of inpatient psychiatric patients ( $M = 1.19$ ;  $SD = .86$ )  
261 (Franke, 2000). A mean GAF score of 45.7 at admission indicated severely impaired general  
262 functioning as a result of serious symptoms. Overall satisfaction with life as measured by the  
263 SWLS was rather low, at 14.7.

264 Insert Table 1 here.

**265 Routine clinical care treatment**

266 The multiple clinics are part of one large clinic group and all offer a multimodal  
267 symptom-specific CBT treatment program for OCD based on national and international  
268 guidelines of CBT for OCD (Hohagen et al., 2015; National Institute for Health and Clinical  
269 Excellence (NICE), 2006). These guidelines also refer to modern transdiagnostic approaches  
270 like Acceptance Commitment Therapy (ACT; e.g. Twohig et al., 2010). In Germany,  
271 outpatient treatments for OCD very often fall short of offering the necessary amount and

## PREDICTORS OF TREATMENT OUTCOME IN OCD

13

272 quality of therapist-accompanied ERPs (Külz et al., 2010; Voderholzer et al., 2015).  
273 Therefore, OCD-specific inpatient treatment programs aim to deliver a maximum of  
274 therapist-accompanied ERPs as a crucial change factor in the treatment of OCD. The  
275 inpatient treatment in these clinics is carried out on specific wards for the treatment of OCD  
276 (i.e. in community with other OCD inpatients), and comprises three stages in individual and  
277 group therapy: ‘motivation and psychoeducation’ (1-2 weeks), ‘exposure with response  
278 prevention (ERP)’ (4-6 weeks) and ‘transfer and relapse prevention’ (1-2 weeks). The  
279 therapeutic elements of this CBT treatment program are as follows:

280 The first stage of ‘psychoeducation’ includes OCD-specific behavior analyses based on  
281 the cognitive-behavioral model of OCD (Salkovskis, 1999), the identification of individual  
282 safety behavior and avoidance strategies, the rationale of expositions, the development of an  
283 individual hierarchy of critical situations and cognitive therapy (e.g. with focus on the  
284 appraisal of individual thoughts as threatening, probability bias and thought-action fusion) as  
285 well as a detailed functional assessment of OCD (e.g. Abramowitz, Deacon, & Whiteside,  
286 2019). In the second stage of ‘exposure with response prevention (ERP)’, patients receive a  
287 minimum of three accompanied ERPs within the individual psychotherapy plus a minimum  
288 of four ERPs delivered by specifically trained nurses in individual- or group-setting and  
289 weekly protocol groups including behavior analyses of critical situations and review of self-  
290 controlled ERPs. Additionally, about two hours per day are reserved for self-controlled  
291 ERPs, behavioral training of norm behavior accompanied by therapeutic contracts defining  
292 ‘reference dates’ to strengthen commitment to norm behavior after expositions, and  
293 behavioral training of alternative coping strategies. In the third stage of ‘transfer and relapse  
294 prevention’, the focus lies on home exposition (including videoconference-based therapeutic  
295 supervision; cf. Netter et al., 2018), inclusion of relatives and partners into individual  
296 psychotherapy and peer-support groups with former OCD-patients in remission.

## PREDICTORS OF TREATMENT OUTCOME IN OCD

14

297 Multimodal and transdiagnostic interventions include weekly interactional ward groups  
298 (2 hours per week; e.g. elaboration of alternative social skills and coping with emotions,  
299 clarification of ‘normal’ behavior, identification and treatment of individual functions of  
300 OCD), weekly mindfulness groups (1 hour per week), sports and body therapy groups (3  
301 hours per week) as well as access to social skills training and/or symptom-specific group  
302 therapy for comorbid disorders (e.g. depression or social phobia) (2 hours per week), and  
303 relaxation training, art therapy group, biofeedback and social counseling (by trained social  
304 workers) (2 hours per week) according to individual indication and psychiatric  
305 comorbidities. Additionally, modules of ACT with a focus on the application of acceptance-  
306 based strategies on coping with obsessions and compulsive behavior are included into  
307 individual and group therapies (2 hours per week in group therapy). Therapy is delivered  
308 under weekly team supervision and a weekly visit of a senior physician.

309 In total, patients receive at least 1 hour per week individual psychotherapy, 7 hours per  
310 week specific OCD-therapy plus a minimum of 12 hours per week of multimodal and  
311 transdiagnostic interventions over 8-10 weeks leading to a total of 160-200 therapy hours.  
312 In our sample, the average length of stay was  $M = 54.88$  days ( $SD = 18.02$ ) for the full  
313 sample and  $M = 65.84$  days ( $SD = 22.50$ ) for the OCI-R sample.

314 For the patient-staff ratio on each OCD ward, a minimum of two licensed  
315 psychotherapists with specific training in OCD treatment are available (3-5 years of training  
316 in CBT; one medical doctor and one certified clinical psychologist), 1-2 therapists in  
317 advanced CBT training, and at least 1,5 certified nursing staff member, through which the  
318 adherence to OCD-specific behavior therapy including therapist-accompanied ERPs could  
319 be guaranteed. During the weekends, emergency care of the clinic with medical and nursing  
320 personal is warranted.

321 Psychopharmacological treatment was administered when needed and indicated, mainly  
322 to treat psychiatric comorbidities, in compliance with current national and international  
323 guidelines for OCD treatment and in accordance with clinical expert supervision (Hohagen  
324 et al., 2015; National Institute for Health and Clinical Excellence (NICE), 2006).  
325 Nevertheless, the major treatment focus lied on psychotherapeutic work in single and group  
326 settings; medication was therefore continued or reduced to former psychopharmacological  
327 treatment before intake.

### 328 **Measures**

329 Primary treatment outcome in terms of symptom-specific change was assessed at  
330 admission and discharge using the following measures: the Yale-Brown Obsessive-  
331 Compulsive Scale (Y-BOCS) and the revised Obsessive-Compulsive Inventory (OCI-R). As  
332 for secondary treatment outcomes, the following measures were assessed at admission and  
333 discharge: Beck's Depression Inventory-II (BDI-II), the Brief Symptom Inventory (BSI), the  
334 Patient Health Questionnaire-9 (PHQ-9), the Patient Health Questionnaire-15 (PHQ-15), the  
335 General Anxiety Disorder-7 (GAD-7), the General Assessment of Functioning (GAF), and  
336 the Satisfaction with Life Scale (SWLS). For the symptom-specific primary outcome Y-  
337 BOCS-SR and the secondary outcome PHQ-9, prediction analyses were computed.

### 338 **Primary outcome: obsessive-compulsive symptoms**

339 The Yale-Brown Obsessive Compulsive Scale (Y-BOCS) (Goodman et al., 1989) was  
340 originally developed as a clinician-rated interview measuring the severity of obsessive-  
341 compulsive disorder. It consists of 10 items, each rated from 0 (no symptoms) to 4 (extreme  
342 symptoms). Separate sub-scores for the severity of obsessions and compulsions can be  
343 calculated. A self-rating scale of the Y-BOCS (Y-BOCS-SR) was developed; while one  
344 study showed only a moderate relationship between the interview and the self-report version  
345 of the Y-BOCS in a clinical sample of OCD patients (Federici et al., 2010), another study

## PREDICTORS OF TREATMENT OUTCOME IN OCD

16

346 has shown excellent internal consistency, test-retest reliability, and convergent validity  
347 (Steketee et al., 1996).

348 For further analyses, the revised Obsessive-Compulsive Inventory (OCI-R) (Foa et al.,  
349 2002) was used. The OCI-R is a self-report instrument measuring various symptom domains  
350 of obsessive-compulsive disorder. It contains 18 items over six subscales: washing,  
351 checking, ordering, obsessing, hoarding, neutralizing. Items are rated on a five-point Likert-  
352 scale indicating the distress caused by symptoms, ranging from 0 (not at all) to 4  
353 (extremely). The OCI-R has shown good to excellent internal consistency, test-retest  
354 reliability, and convergent reliability (Foa et al., 2002).

355 **Secondary outcomes: depression, anxiety, somatic symptoms, psychological distress,**  
356 **general functioning, quality of life**

357 The Patient Health Questionnaire (PHQ-D) is a self-administered questionnaire  
358 consisting of the PHQ-9 assessing depression, the GAD-7 measuring generalized anxiety  
359 disorder, and the PHQ-15 measuring somatic symptoms. In addition, data from the BSI  
360 (Derogatis & Melisaratos, 1983), BDI-II (Beck et al., 1996), GAF (Endicott et al., 1976), and  
361 SWLS (Diener et al., 1985) were used in the analyses. Among all secondary outcomes, the  
362 PHQ-9 was chosen for predictor analyses due to the high rate of comorbid depression in this  
363 sample and fewer missing data compared to the BDI-II.

364 **Other measures**

365 At admission, patients answered sociodemographic questions in a self-report  
366 questionnaire covering age, sex, number of children, marital status, living conditions, ability  
367 to work, wish to retire, and so forth. In addition, clinical characteristics such as mental and  
368 medical diagnoses, previous psychotherapeutic and/or psychiatric treatment, and number of  
369 comorbidities were assessed at this stage.

370

371 **Statistical analysis**

372 All statistical analyses were conducted in the statistical processing language R (Ihaka &  
373 Gentleman, 1996). Time-event data was log-transformed. Univariate outlier values were  
374 removed by a  $p < .001$  criterion (Tabachnick & Fidell, 2013) adapted to the large sample size  
375 on the z distribution for all continuous variables, post-treatment values of the outcomes, and  
376 change scores. Missing data imputation was performed by random forest imputation using  
377 the R package *missForest* (Stekhoven & Bühlmann, 2012), a non-parametric imputation  
378 algorithm based on random forests that results in fewer imputation and prediction errors  
379 compared with other commonly used imputation techniques (Waljee et al., 2013).

380 In order to determine the effectiveness of routine clinical care treatment, effect sizes for  
381 the full and the OCI-R sample were calculated using Hedge's  $g$ . In addition, despite criticism  
382 of its use in clinical trials (Lachin, 2016), the Last Observation Carried Forward (LOCF)  
383 method was used as a more conservative method for estimating effect size. Since this study  
384 was naturalistic in nature, LOCF seemed to be a satisfactory way of comparing effect sizes  
385 with uncorrected effect sizes in order to get a more holistic estimation of the real effect size.

386 In order to reduce the dimensionality of the data, sparse principal component analysis  
387 (SPCA) ( $\alpha = 0.005$ ) was performed on the predictors using the package *sparsepca*  
388 (Erichson et al., 2018). SPCA has emerged as a powerful data analysis technique with better  
389 interpretability than that of traditional principal component analysis (PCA). Because of  
390 ambiguous scree plot results, Horn's parallel analysis (Horn, 1965) was performed with the  
391 function *fa.parallel* in the R package *psych* (Revelle, 2017) to determine the optimal number  
392 of components. For each outcome (Y-BOCS-SR as primary and PHQ-9 as secondary), a  
393 stepwise multiple regression model was fitted. In the first step, the baseline sum score of the  
394 two outcomes was used as the single predictor. In the second step, the components derived  
395 from the SPCA were included as well.

396

**Results****397 Effectiveness of the inpatient treatment**

398 In the full sample, the mean difference between the Y-BOCS scores pre-treatment (M =  
399 25.5, SD = 5.3) and post-treatment (M = 16.0, SD = 7.2) was 9.5 points (SD = 7.1). For the  
400 primary treatment outcome, Hedge's *g* of the symptom change score in the Y-BOCS total  
401 was 1.34 (95% CI [1.26, 1.43]) while the LOCF-corrected Hedge's *g* was 1.10 (95% CI  
402 [1.03, 1.17]), indicating large effect sizes. Even with the LOCF-corrected estimation, effect  
403 sizes for the Y-BOCS pre-/post change remained large. With regard to the subscales,  
404 Hedge's *g* of the symptom change in Y-BOCS behavior was 1.29 (95% CI [1.21, 1.37]),  
405 while the LOCF-corrected ES was 1.10 (95% CI [1.00, 1.15]); Hedge's *g* of the symptom  
406 change in Y-BOCS thoughts was to some extent lower, i.e., 1.10 (95% CI [1.00, 1.18]), and  
407 for LOCF-corrected .93 (95% CI [.85, 1.00]).

408 In the OCI-R sample, the pre-treatment (M = 24.7, SD = 5.5) and post-treatment scores  
409 (M = 15.6, SD = 7.0) were slightly lower than in the full sample, as was the mean difference  
410 between pre- and post-treatment scores (M = 9.0, SD = 7.2). The effect sizes of all outcomes  
411 are shown in Table 2.

412 Even in this subsample, the effect sizes with respect to the change in Y-BOCS symptoms  
413 could be considered large: Hedge's *g* of the symptom change score in the OCI-R was 1.17  
414 (95% CI [1.02, 1.31]), which indicates large effects overall, while the subscales showed  
415 differences. Most of the effect sizes in subscales were large, with two exceptions: Hedge's *g*  
416 of symptom changes in OCI-R hoarding .46 (95% CI [.32, .60]) and neutralizing .52 (95%  
417 CI [.38, .67]) could be considered medium. The effect sizes of other secondary outcomes  
418 ranged between medium for the SWLS and PHQ-15, to large for the BDI-II, PHQ-9, BSI  
419 (GSI), GAD-7, and GAF (see Table 2).

420

Insert Table 2 here.

421 **Dimensional reduction**

422 The SPCA revealed 10 distinct factors to best represent the underlying data of 54 potential  
423 predictors. These ten factors, interpreted by the variables with the highest factor loading and  
424 considering the side loadings, were labelled as follows: “Distress,” “Somatic disorders,”  
425 “Obsessing,” “Social support,” “Ordering,” “Chronicity of depression,” “Comorbid  
426 depression,” “Academic,” “Functional disability,” and “Washing behaviour.” These factors  
427 were then used in the subsequent regression analyses.

428 The results of the SPCA, including factor loadings and communalities as well as the scree  
429 plot, are presented in the Supplemental material (see Supplemental material 1 and 2).

430 **Predictors of primary treatment outcome**

431 The regression results for the Y-BOCS total score as the criterion are shown in Table 3,  
432 the regression results for the Y-BOCS sub-scores “Obsessions” and “Compulsions” in Table  
433 4.

434 Insert Table 3 here.

435 For the Y-BOCS total score post-treatment as the criterion, the baseline Y-BOCS score  
436 as predictor in the first step explained 13.3 % of the variance and had a significant effect ( $\beta =$   
437  $.37, p < .001$ ). The second set of predictors added another 8.7% of explained variance (95%  
438 CI [.04, .13],  $p < .01$ ), so that overall 22% of the variance could be explained. In this step  
439 and additionally for the Y-BOCS baseline score ( $\beta = .33, p < .001$ ), the factors “distress” ( $\beta$   
440  $= .21, p < .001$ ), “social support” ( $\beta = -.10, p = .036$ ), and “washing behavior” ( $\beta = -.21, p <$   
441  $.001$ ) had significant effects.

442 Insert Table 4 here.

443 The baseline Y-BOCS Obsessions score as predictor in the first step explained 20.6% of  
444 the variance and significantly predicted Y-BOCS Obsessions scores post-treatment ( $\beta = .45,$   
445  $p < .001$ ). The second series of predictors contributed another 8.8% of the explained variance



## PREDICTORS OF TREATMENT OUTCOME IN OCD

20

446 (95% CI [.04, .13],  $p < .01$ ), so that overall 29.4% of the variance could be explained. In this  
 447 step and additionally for the Y-BOCS Obsessions baseline score ( $\beta = .37, p < .001$ ), the  
 448 factors “distress” ( $\beta = .24, p < .001$ ), “obsessing” ( $\beta = .14, p = .002$ ), and “washing  
 449 behavior” ( $\beta = -.24, p < .001$ ) had significant effects on Y-BOCS Obsessions scores post-  
 450 treatment.

451 For the Y-BOCS Compulsions score post-treatment as the criterion, the baseline Y-  
 452 BOCS Compulsions score as predictor in the first step explained a significant proportion of  
 453 variance (13.0%) and had a significant effect ( $\beta = .36, p < .001$ ). The second group of  
 454 predictors added another 5.9% of the explained variance (95% CI [.02, .10],  $p < .01$ ), so that  
 455 overall 18.9% of the variance could be explained. In this step and additionally for the Y-  
 456 BOCS Compulsions baseline score ( $\beta = .32, p < .001$ ), the factors “distress” ( $\beta = .12, p =$   
 457 .035) and “ordering” ( $\beta = .12, p = .028$ ) significantly predicted Y-BOCS Compulsion scores  
 458 post-treatment.

459 All other potential predictors had no significant effect either on the total score or on the  
 460 scores of the subscales post-treatment ( $p > .05$ ).

#### 461 Predictors of secondary treatment outcome

462 The regression results for the PHQ-9 score post-treatment as the criterion are listed in  
 463 Table 5.

464 Insert Table 5 here.

465 The baseline PHQ-9 score as predictor in the first step explained 32% of the variance and  
 466 had a significant effect on PHQ-9 score post-treatment ( $\beta = .57, p < .001$ ). Another 7.1% of  
 467 the explained variance was added by the second set of predictors (95% CI [.03, .11],  $p <$   
 468 .01), so that overall 39.1% of the variance could be explained. In this step and additionally  
 469 for the PHQ-9 baseline score ( $\beta = .28, p < .001$ ), the factors “distress” ( $\beta = .31, p < .001$ ),  
 470 “social support” ( $\beta = -.12, p = .002$ ), and “disability” ( $\beta = .12, p = .003$ ) significantly

471 predicted PHQ-9 scores post-treatment. All other potential predictors had no significant  
472 effect ( $p > .05$ ).

#### 473 **Discussion**

474 This study sheds some light on the routine clinical care of OCD patients: first, by  
475 determining the effectiveness of an inpatient OCD treatment on symptom-specific as well as  
476 functional outcomes under everyday health care conditions; and secondly, by identifying  
477 important predictors of treatment outcome in a large sample of OCD patients being treated in  
478 German psychosomatic clinics. Encouragingly, the overall large effect sizes indicate that  
479 OCD treatment is delivered effectively in routine clinical inpatient care, despite the fact that  
480 the pre-/post-effect sizes were marginally lower than those reported in other studies (Grøtte  
481 et al., 2018) and a COCHRANE review (Gava et al., 2007). Notwithstanding this, given the  
482 naturalistic nature of the study this finding speaks in favor of the widespread implementation  
483 of effective OCD treatment in daily practice in Germany. Concerning the effectiveness vs.  
484 efficacy discussion (cf. Wampold & Imel, 2015), the effect sizes found in this study could be  
485 interpreted as being very good. Due to the large and well-described inpatient OCD sample,  
486 this effect sizes might be seen as a benchmark for OCD treatment. Furthermore, a number of  
487 relevant predictors were identified, replicating and building on earlier research results  
488 through a study combining a large, well-described sample, novel state-of-the-art statistical  
489 procedures, and the inclusion of sociodemographic, clinical, and psychometric variables in a  
490 single model.

#### 491 **Main findings and clinical implications**

492 The sample in this study seems to be typical of a German inpatient clinic: in another  
493 sample from two German inpatient clinics, a similar mean score for age, gender distribution  
494 and Y-BOCS was reported, albeit with a slightly higher OCI-R score and a higher  
495 percentage of comorbid personality disorders (Voderholzer et al., 2013). However,

## PREDICTORS OF TREATMENT OUTCOME IN OCD

22

496 Voderholzer and colleagues' sample did not include patients with a comorbid depressive  
497 disorder, while a high percentage of our study's sample showed depressive comorbidity.  
498 Differences between the pre-treatment and post-treatment Y-BOCS scores were slightly  
499 lower than the mean difference of 10.7 points (95% CI: 9.8–11.5) computed in a recent  
500 meta-analysis (Veale et al., 2016). However, as pre-treatment OCD severity appears to be a  
501 predictor of treatment outcome (Veale et al., 2016), the lower difference scores might be a  
502 result of relatively less OCD severity compared with the meta-analytic mean. This could, in  
503 turn, be partly a result of using a cut-off score of 10 for the Y-BOCS on diagnosis F42.0 and  
504 F42.1 instead of an overall cut-off score of 16.

505 Overall, and in line with previous studies (Kohls et al., 2002; Kyrios et al., 2015; Raffin  
506 et al., 2009; Steketee et al., 1999), pre-treatment OCD symptom severity and general distress  
507 appear to predict post-treatment symptom severity. In general, for symptom-specific change,  
508 OCD patients with lower general distress, higher social support, and higher scores on  
509 washing behavior appear to benefit most from the treatment.

510 Two conclusions might be drawn from this result: First, a lack of social support appears  
511 to diminish treatment outcome. This finding is in line with the results of a recent systematic  
512 review (Palardy et al., 2018) in which the authors concluded that negative social support  
513 appeared to be associated with more severe symptoms whereas positive social support could  
514 be beneficial for OCD patients. This is also in accordance with another study that  
515 highlighted the quality of social support (Van Noppen & Steketee, 2009). Hence, our  
516 findings once again underscore the role of relatives in supporting OCD patients recovering  
517 from their illness. There is no doubt that living with a patient who suffers from severe OCD  
518 can be a challenge for family members, as the compulsions can manifest themselves in daily  
519 interactions and affect relationships. In particular, living together can become a burden if  
520 relatives do not know how to deal with such behavior. It could therefore be beneficial either

## PREDICTORS OF TREATMENT OUTCOME IN OCD

23

521 to involve relatives of OCD patients in the treatment process by providing them with useful  
522 information, e.g., psychoeducation on how to deal with the disease and its symptoms in  
523 everyday life, or to strengthen the quality of relationships through systemic or family  
524 interventions (Sydow & Borst, 2018). This might also lead to a reduction in comorbid  
525 depressive symptoms, as our results highlight the positive role of social support in reducing  
526 depressive symptoms in OCD patients.

527       Second, our analyses indicate that, surprisingly, OCD patients who exhibit more washing  
528 behavior benefit more from treatment than do other subgroups of OCD patients. This may be  
529 because washing behavior during CBT (respectively ERP) is both easier to treat and  
530 treatable in the inpatient setting compared with other compulsive behaviors, such as  
531 hoarding which is more a problem in patients' daily lives at home, or checking which is  
532 sometimes difficult to stop in patients' minds. Future studies conducted in inpatient settings  
533 could provide more evidence for this finding.

534       Sub-analyses revealed distinct predictor profiles for each of the symptom-specific  
535 reductions (obsessions and compulsions): interestingly, the results suggest that OCD patients  
536 with higher scores on obsessing thoughts appear to benefit less from inpatient treatment in  
537 terms of a reduction in obsessions. The relatively short CBT treatment involved may not be  
538 sufficient to reduce pure obsessions. Hence, for patients with high scores on obsession scales  
539 it may be useful to integrate some metacognitive elements (Andouz et al., 2012) from  
540 metacognitive therapy (Fisher & Wells, 2008; Wells, 2009) into the CBT to improve their  
541 treatment outcome. Recently, several research groups have applied metacognitive theories to  
542 OCD (Grøtte et al., 2016; Hansmeier et al., 2016; Hauschildt et al., 2016; Mavrogiorgou et  
543 al., 2016; van der Heiden et al., 2016). On the other hand, OCD patients with higher scores  
544 on washing behavior appear to benefit more from the treatment through a reduction in their  
545 obsessions. This result adds to the evidence base for CBT in treating OCD patients with high

546 levels of washing behavior, and stresses the crucial role of changing debilitating obsessions  
547 in such patients.

548 Also noteworthy is the finding that OCD patients with higher scores on ordering appear  
549 to benefit less from the treatment in terms of reducing compulsions. For these patients,  
550 traditional CBT treatment may not be effective in reducing their compulsions. Practitioners  
551 might therefore consider incorporating other elements, such as home-visits, into treatment  
552 planning, as has been suggested in specialized CBT treatment for patients with compulsive  
553 hoarding (Steketee et al., 2010).

#### 554 **Strengths and limitations**

555 A major strength of the current study is the large sample size, combined with the use of  
556 the “gold-standard” of disorder-specific self-report measures for OCD, i.e., the Y-BOCS-SR,  
557 as the primary outcome. In addition, the sample was well characterized by sociodemographic  
558 and clinical data. Furthermore, the dimensional reduction of sociodemographic, clinical, and  
559 psychological variables to interpretable factors was rendered “state-of-the-art” by using a  
560 data-driven, novel, statistical technique—sparse principal component analysis—which  
561 proved to be superior to previous factor analysis techniques.

562 Within a Phase IV-study, our study exhibits high external validity while internal validity  
563 aspects might be compromised. The interpretation of the results is therefore limited by the  
564 naturalistic and thus uncontrolled and non-randomized nature of the study, i.e., the lack of a  
565 control group and no assessment of treatment integrity nor patient engagement/compliance.  
566 Treatment selection was not random, so unknown patient and/or clinician characteristics that  
567 contributed to treatment choice may have been relevant. Since there was no strict CBT  
568 treatment manual, practitioners may have chosen different techniques (possibly no CBT-  
569 specific strategies) to treat the specific needs of individual patients. Furthermore, we have no  
570 information on the number of therapists, their competence, their level of training or their

571 therapeutic experience. On the other hand, one of the strengths of a naturalistic design is its  
572 closer approximation to clinical practice and therefore higher external validity—even if this  
573 is at the expense of internal validity. Another major limitation is the lack of data on  
574 medication and psychopharmacological interventions. Moreover, this study lacks long-  
575 term/follow-up outcome assessment, which could be of importance because of the controlled  
576 nature of the residential environment. A further limitation is the reliance on self-report  
577 measures (Federici et al., 2010), i.e., Y-BOCS-SR and OCI-R, which were used because they  
578 are cost-effective to use in daily patient care. This is insofar problematic as it reduces the  
579 reliability with which the diagnosis of OCD was assigned. That aside, various outcomes  
580 were considered, and the findings of the Y-BOCS-SR were supported by the results of the  
581 OCI-R. However, future studies should also include observer-rated instruments in large  
582 samples, e.g., by developing new, cost-effective, therapist-rated instruments to assess OCD  
583 symptoms. Lastly, there is a significant loss of patients from effectiveness analysis to  
584 predictor analyses. Yet, the authors decided to include the OCI-R as a crucial and potentially  
585 informative candidate predictor.

#### 586 **Future research directions**

587 With regard to naturalistic studies, future research could include routine outcome  
588 monitoring (ROM; Lambert & Harmon, 2018; Lambert, Whipple, & Kleinstäuber, 2018)  
589 with the aim of further benchmarking OCD treatments, predicting early changes (early  
590 responder, sudden gains; Lutz et al., 2014; Lutz, Stulz, & Köck, 2009), and correcting  
591 negative developments (sudden losses; Lutz et al., 2013). Despite the promising findings of  
592 this study, it remains a future research question whether these predictors, more specifically  
593 prognostic variables, can be replicated under more standardized conditions (such as RCTs).  
594 Therefore, these predictors should be validated in prospective controlled studies, focusing on  
595 long-term outcomes in order to build on the findings presented here. In addition, there may

## PREDICTORS OF TREATMENT OUTCOME IN OCD

26

596 be other important predictors that were not included in this study, e.g., the therapeutic  
597 relationship (Vogel et al., 2006), motivational processes (de Haan et al., 1997; Reid et al.,  
598 2017; Steketee et al., 2011), and patient adherence (Maher et al., 2012; Simpson et al.,  
599 2011), childhood maltreatment (R.-Mercier et al., 2018), and expectations (Rief et al., 2015).  
600 Finally, to guide treatment selection procedures, it is necessary to identify not only  
601 prognostic but also prescriptive variables that indicate the relative efficacy of one treatment  
602 over another (Cohen & DeRubeis, 2018). Such research, which is also referred to as  
603 "practice-oriented research" or "patient-focused research" (Lutz et al., 2015) and includes  
604 process measures, has become more and more a focus of interest in psychotherapy research,  
605 whereby large datasets from routine care, sometimes in the context of ROM, are used to  
606 generate useful statements that can be applied directly in clinical practice. The aim is to  
607 provide therapists with helpful feedback information to facilitate the decision-making  
608 process before and during the course of treatment, and thus to be able to treat patients in a  
609 more personalized and thus effective way.

610

**References**

- 611 Abramowitz, J. S., Deacon, B. J., & Whiteside, S. P. H. (2019). *Exposure therapy for*  
612 *anxiety : principles and practice*. Guilford Press.
- 613 Abramowitz, J. S., & Foa, E. B. (2000). Does comorbid major depressive disorder influence  
614 outcome of exposure and response prevention for OCD? *Behavior Therapy, 31*(4), 795–  
615 800. [https://doi.org/10.1016/S0005-7894\(00\)80045-3](https://doi.org/10.1016/S0005-7894(00)80045-3)
- 616 Abramowitz, J. S., Franklin, M. E., Schwartz, S. A., & Furr, J. M. (2003). Symptom  
617 Presentation and Outcome of Cognitive-Behavioral Therapy for Obsessive-Compulsive  
618 Disorder. *Journal of Consulting and Clinical Psychology, 71*(6), 1049–1057.  
619 <https://doi.org/10.1037/0022-006X.71.6.1049>
- 620 Abramowitz, J. S., Franklin, M. E., Street, G. P., Kozak, M. J., & Foa, E. B. (2000). Effects  
621 of comorbid depression on response to treatment for obsessive-compulsive disorder.  
622 *Behavior Therapy, 31*(3), 517–528. [https://doi.org/10.1016/S0005-7894\(00\)80028-3](https://doi.org/10.1016/S0005-7894(00)80028-3)
- 623 Andouz, Z., Dolatshahi, B., Moshtagh, N., & Dadkhah, A. (2012). The efficacy of  
624 metacognitive therapy on patients suffering from pure obsession. *Iranian Journal of*  
625 *Psychiatry, 7*(1), 11–21. <http://www.ncbi.nlm.nih.gov/pubmed/23056112>
- 626 Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory-II*. Harcourt  
627 Assessment Inc.
- 628 Boschen, M. J., Drummond, L. M., Pillay, A., & Morton, K. (2010). Predicting outcome of  
629 treatment for severe, treatment resistant OCD in inpatient and community settings.  
630 *Journal of Behavior Therapy and Experimental Psychiatry, 41*(2), 90–95.  
631 <https://doi.org/10.1016/j.jbtep.2009.10.006>
- 632 Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment Selection in Depression. *Annual Review*  
633 *of Clinical Psychology, 14*, 15.1-15.28. <https://doi.org/10.1146/annurev-clinpsy-050817>
- 634 de Haan, E., van Oppen, P., van Balkom, A. J., Spinhoven, P., Hoogduin, K. A., & Van  
635 Dyck, R. (1997). Prediction of outcome and early vs. late improvement in OCD patients  
636 treated with cognitive behaviour therapy and pharmacotherapy. *Acta Psychiatrica*  
637 *Scandinavica, 96*(5), 354–361. <http://www.ncbi.nlm.nih.gov/pubmed/9395153>
- 638 Derogatis, L. R., & Melisaratos, N. (1983). The brief symptom inventory: an introductory



## PREDICTORS OF TREATMENT OUTCOME IN OCD

28

- 639 report. *Psychological Medicine*, 13(03), 595–605.
- 640 Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life  
641 Scale. *Journal of Personality Assessment*, 49(1), 71–75.  
642 [https://doi.org/10.1207/s15327752jpa4901\\_13](https://doi.org/10.1207/s15327752jpa4901_13)
- 643 Dilling, H., Mombour, W., & Schmidt, M. H. (Eds.). (2015). *Internationale Klassifikation*  
644 *psychischer Störungen: ICD-10 Kapitel V (F); Klinisch-diagnostische Leitlinien* (10.  
645 Aufl.). Hogrefe.
- 646 Dreessen, L., Hoekstra, R., & Arntz, A. (1997). Personality disorders do not influence the  
647 results of cognitive and behavior therapy for obsessive compulsive disorder. *Journal of*  
648 *Anxiety Disorders*, 11(5), 503–521. [https://doi.org/10.1016/S0887-6185\(97\)00027-3](https://doi.org/10.1016/S0887-6185(97)00027-3)
- 649 Eddy, K. T., Dutra, L., Bradley, R., & Westen, D. (2004). A multidimensional meta-analysis  
650 of psychotherapy and pharmacotherapy for obsessive-compulsive disorder. *Clinical*  
651 *Psychology Review*, 24(8), 1011–1030. <https://doi.org/10.1016/j.cpr.2004.08.004>
- 652 Eisen, J. L., Sibrava, N. J., Boisseau, C. L., Mancebo, M. C., Stout, R. L., Pinto, A., &  
653 Rasmussen, S. A. (2013). Five-year course of obsessive-compulsive disorder:  
654 Predictors of remission and relapse. *The Journal of Clinical Psychiatry*, 74(3), 233–  
655 239. <https://doi.org/10.4088/JCP.12m07657>
- 656 Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale.  
657 *Archives of General Psychiatry*, 33(6), 766.  
658 <https://doi.org/10.1001/archpsyc.1976.01770060086012>
- 659 Erichson, N. B., Zheng, P., Manohar, K., Brunton, S. L., Kutz, J. N., & Aravkin, A. Y.  
660 (2018). *Sparse Principal Component Analysis via Variable Projection*.  
661 <http://arxiv.org/abs/1804.00341>
- 662 Farrell, L. J., & Boschen, M. (2011). Treatment outcome in adult OCD: Predictors and  
663 processes of change. *Asia Pacific Journal of Counselling and Psychotherapy*, 2(1), 82–  
664 97. <https://doi.org/10.1080/21507686.2010.536915>
- 665 Federici, A., Summerfeldt, L. J., Harrington, J. L., McCabe, R. E., Purdon, C. L., Rowa, K.,  
666 & Antony Martin M., M. M. (2010). Consistency between self-report and clinician-  
667 administered versions of the Yale-Brown Obsessive-Compulsive Scale. *Journal of*

## PREDICTORS OF TREATMENT OUTCOME IN OCD

29

- 668 *Anxiety Disorders*, 24(7), 729–733. <https://doi.org/10.1016/j.janxdis.2010.05.005>
- 669 Fisher, P. L., & Wells, A. (2005). How effective are cognitive and behavioral treatments for  
670 obsessive–compulsive disorder? A clinical significance analysis. *Behaviour Research*  
671 *and Therapy*, 43(12), 1543–1558. <https://doi.org/10.1016/j.brat.2004.11.007>
- 672 Fisher, P. L., & Wells, A. (2008). Metacognitive therapy for obsessive-compulsive disorder:  
673 A case series. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(2), 117–  
674 132. <https://doi.org/10.1016/j.jbtep.2006.12.001>
- 675 Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P.  
676 M. (2002). The Obsessive-Compulsive Inventory: Development and validation of a  
677 short version. *Psychological Assessment*, 14(4), 485–496.  
678 <http://www.ncbi.nlm.nih.gov/pubmed/12501574>
- 679 Franke, H. (2000). *Brief Symptom Inventory (BSI) - German Version. Manual*. Beltz Test.
- 680 Fricke, S., Moritz, S., Andresen, B., Jacobsen, D., Kloss, M., Rufer, M., & Hand, I. (2006).  
681 Do personality disorders predict negative treatment outcome in obsessive–compulsive  
682 disorders? A prospective 6-month follow-up study. *European Psychiatry*, 21(5), 319–  
683 324. <https://doi.org/10.1016/j.eurpsy.2005.03.010>
- 684 Gava, I., Barbui, C., Aguglia, E., Carlino, D., Churchill, R., De Vanna, M., & McGuire, H.  
685 (2007). Psychological treatments versus treatment as usual for obsessive compulsive  
686 disorder (OCD). *Cochrane Database of Systematic Reviews*, 2.  
687 <https://doi.org/10.1002/14651858.CD005333.pub2>
- 688 Gönner, S., Limbacher, K., & Ecker, W. (2012). Stationäre kognitive Verhaltenstherapie bei  
689 Zwangsstörungen: Effektivität und Erfolgsprädiktoren in der Routineversorgung.  
690 *Verhaltenstherapie*, 22(1), 17–26. <https://doi.org/10.1159/000335776>
- 691 Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C.  
692 L., Heninger, G. R., & Charney, D. S. (1989). The Yale-Brown Obsessive Compulsive  
693 Scale. *Archives of General Psychiatry*, 46(11), 1006.  
694 <https://doi.org/10.1001/archpsyc.1989.01810110048007>
- 695 Grøtte, T., Hansen, B., Haseeth, S., Vogel, P. A., Guzey, I. C., & Solem, S. (2018). Three-  
696 Week Inpatient Treatment of Obsessive-Compulsive Disorder: A 6-Month Follow-Up

## PREDICTORS OF TREATMENT OUTCOME IN OCD

30

- 697 Study. *Frontiers in Psychology*, 9, 620. <https://doi.org/10.3389/fpsyg.2018.00620>
- 698 Grøtte, T., Solem, S., Myers, S. G., Hjemdal, O., Vogel, P. A., Güzey, I. C., Hansen, B.,  
699 Nordahl, H. M., & Fisher, P. (2016). Metacognitions in obsessive-compulsive disorder:  
700 A psychometric study of the Metacognitions Questionnaire-30. *Journal of Obsessive-*  
701 *Compulsive and Related Disorders*, 11(May), 82–90.  
702 <https://doi.org/10.1016/j.jocrd.2016.09.002>
- 703 Hansen, B., Vogel, P. A., Stiles, T. C., & Gunnar Gøtestam, K. (2007). Influence of Co-  
704 Morbid Generalized Anxiety Disorder, Panic Disorder and Personality Disorders on the  
705 Outcome of Cognitive Behavioural Treatment of Obsessive-Compulsive Disorder.  
706 *Cognitive Behaviour Therapy*, 36(3), 145–155.  
707 <https://doi.org/10.1080/16506070701259374>
- 708 Hansmeier, J., Exner, C., Rief, W., & Glombiewski, J. A. (2016). A test of the metacognitive  
709 model of obsessive-compulsive disorder. *Journal of Obsessive-Compulsive and Related*  
710 *Disorders*, 10, 42–48. <https://doi.org/10.1016/j.jocrd.2016.05.002>
- 711 Hauschildt, M., Schröder, J., & Moritz, S. (2016). Randomized-controlled trial on a novel  
712 (meta-)cognitive self-help approach for obsessive-compulsive disorder (“myMCT”).  
713 *Journal of Obsessive-Compulsive and Related Disorders*, 10, 26–34.  
714 <https://doi.org/10.1016/j.jocrd.2016.04.010>
- 715 Hohagen, F., Wahl-Kordon, A., Lotz-Rambaldi, W., & Muche-Borowski, C. (2015). *S3-*  
716 *Leitlinie Zwangsstörungen*. Springer-Verlag. [https://doi.org/10.1007/978-3-662-44138-](https://doi.org/10.1007/978-3-662-44138-1)  
717 1
- 718 Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis.  
719 *Psychometrika*, 30(2), 179–185. <https://doi.org/10.1007/BF02289447>
- 720 Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal*  
721 *of Computational and Graphical Statistics*, 5(3), 299–314.  
722 <https://doi.org/10.1080/10618600.1996.10474713>
- 723 Jakubovski, E., Diniz, J. B., Valerio, C., Fossaluza, V., Belotto-Silva, C., Gorenstein, C.,  
724 Miguel, E., & Shavitt, R. G. (2013). Clinical predictors of long-term outcome in  
725 obsessive-compulsive disorder. *Depression and Anxiety*, 30(8), 763–772.  
726 <https://doi.org/10.1002/da.22013>

## PREDICTORS OF TREATMENT OUTCOME IN OCD

31

- 727 Keeley, M. L., Storch, E. A., Merlo, L. J., & Geffken, G. R. (2008). Clinical predictors of  
728 response to cognitive-behavioral therapy for obsessive-compulsive disorder. *Clinical*  
729 *Psychology Review*, 28(1), 118–130. <https://doi.org/10.1016/j.cpr.2007.04.003>
- 730 Keijsers, G. P. J., Hoogduin, C. A. L., & Schaap, C. P. D. R. (1994). Predictors of Treatment  
731 Outcome in the Behavioural Treatment of Obsessive-Compulsive Disorder. *British*  
732 *Journal of Psychiatry*, 165(06), 781–786. <https://doi.org/10.1192/bjp.165.6.781>
- 733 Kempe, P. T., van Oppen, P., de Haan, E., Twisk, J. W. R., Sluis, A., Smit, J. H., van Dyck,  
734 R., & van Balkom, A. J. L. M. (2007). Predictors of course in obsessive-compulsive  
735 disorder: logistic regression versus Cox regression for recurrent events. *Acta*  
736 *Psychiatrica Scandinavica*, 116(3), 201–210. [https://doi.org/10.1111/j.1600-](https://doi.org/10.1111/j.1600-0447.2007.00997.x)  
737 [0447.2007.00997.x](https://doi.org/10.1111/j.1600-0447.2007.00997.x)
- 738 Knopp, J., Knowles, S., Bee, P., Lovell, K., & Bower, P. (2013). A systematic review of  
739 predictors and moderators of response to psychological therapies in OCD: Do we have  
740 enough empirical evidence to target treatment? *Clinical Psychology Review*, 33(8),  
741 1067–1081. <https://doi.org/10.1016/j.cpr.2013.08.008>
- 742 Kohls, S., Bents, H., & Pietrowsky, R. (2002). Prädiktoren für Erfolg und Misserfolg in der  
743 verhaltenstherapeutischen Behandlung von Zwangspatienten. *Verhaltenstherapie*,  
744 12(2), 98–106. <https://doi.org/10.1159/000064373>
- 745 Külz, A. K., Hassenpflug, K., Riemann, D., Linster, H. W., Dornberg, M., & Voderholzer,  
746 U. (2010). Ambulante psychotherapeutische Versorgung bei Zwangserkrankungen:  
747 Ergebnisse einer anonymen Therapeutenbefragung. *PPmP Psychotherapie*  
748 *Psychosomatik Medizinische Psychologie*, 60(6), 194–201. [https://doi.org/10.1055/s-](https://doi.org/10.1055/s-0029-1202837)  
749 [0029-1202837](https://doi.org/10.1055/s-0029-1202837)
- 750 Kyrios, M., Hordern, C., & Fassnacht, D. B. (2015). Predictors of response to cognitive  
751 behaviour therapy for obsessive-compulsive disorder. *International Journal of Clinical*  
752 *and Health Psychology*, 15(3), 181–190. <https://doi.org/10.1016/j.ijchp.2015.07.003>
- 753 Lachin, J. M. (2016). Fallacies of last observation carried forward analyses. *Clinical Trials*,  
754 13(2), 161–168. <https://doi.org/10.1177/1740774515602688>
- 755 Lambert, M. J., & Harmon, K. L. (2018). The merits of implementing routine outcome  
756 monitoring in clinical practice. *Clinical Psychology: Science and Practice*, 25(4),

## PREDICTORS OF TREATMENT OUTCOME IN OCD

32

- 757 e12268. <https://doi.org/10.1111/cpsp.12268>
- 758 Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering  
759 progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*,  
760 55(4), 520–537. <https://doi.org/10.1037/pst0000167>
- 761 Langner, J., Laws, M., Röper, G., Zaudig, M., Hauke, W., & Piesbergen, C. (2009).  
762 Predicting Therapy Outcome in Patients with Early and Late Obsessive-Compulsive  
763 Disorder (EOCD and LOCD). *Behavioural and Cognitive Psychotherapy*, 37(05), 485.  
764 <https://doi.org/10.1017/S1352465809990294>
- 765 Lutz, W., De Jong, K., & Rubel, J. (2015). Patient-focused and feedback research in  
766 psychotherapy: Where are we and where do we want to go? *Psychotherapy Research*,  
767 25(6), 625–632. <https://doi.org/10.1080/10503307.2015.1079661>
- 768 Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M.-A., Jorasz, C., Mocanu, S.,  
769 Vocks, S., Schulte, D., & Tschitsaz-Stucki, A. (2013). The ups and downs of  
770 psychotherapy: Sudden gains and sudden losses identified with session reports.  
771 *Psychotherapy Research*, 23(1), 14–24. <https://doi.org/10.1080/10503307.2012.693837>
- 772 Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., Woods, S.  
773 W., & Barlow, D. H. (2014). Patterns of early change and their relationship to outcome  
774 and early treatment termination in patients with panic disorder. *Journal of Consulting  
775 and Clinical Psychology*, 82(2), 287–297. <https://doi.org/10.1037/a0035535>
- 776 Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to  
777 outcome and follow-up among patients with major depressive disorders. *Journal of  
778 Affective Disorders*, 118(1–3), 60–68. <https://doi.org/10.1016/j.jad.2009.01.019>
- 779 Maher, M. J., Wang, Y., Zuckoff, A., Wall, M. M., Franklin, M., Foa, E. B., & Simpson, H.  
780 B. (2012). Predictors of patient adherence to cognitive-behavioral therapy for  
781 obsessive-compulsive disorder. *Psychotherapy and Psychosomatics*, 81(2), 124–126.  
782 <https://doi.org/10.1159/000330214>
- 783 Mavrogiorgou, P., Bethge, M., Luksnat, S., Nalato, F., Juckel, G., & Brüne, M. (2016).  
784 Social cognition and metacognition in obsessive-compulsive disorder: An explorative  
785 pilot study. *European Archives of Psychiatry and Clinical Neuroscience*, 266(3), 209–  
786 216. <https://doi.org/10.1007/s00406-016-0669-6>

## PREDICTORS OF TREATMENT OUTCOME IN OCD

33

- 787 McKay, D., Sookman, D., Neziroglu, F., Wilhelm, S., Stein, D. J., Kyrios, M., Matthews, K.,  
788 & Veale, D. (2015). Efficacy of cognitive-behavioral therapy for obsessive-compulsive  
789 disorder. *Psychiatry Research*, 227(1), 104–113.  
790 <https://doi.org/10.1016/j.psychres.2014.11.058>
- 791 Moritz, S., Fricke, S., Jacobsen, D., Kloss, M., Wein, C., Rufer, M., Katenkamp, B.,  
792 Farhumand, R., & Hand, I. (2004). Positive schizotypal symptoms predict treatment  
793 outcome in obsessive–compulsive disorder. *Behaviour Research and Therapy*, 42(2),  
794 217–227. [https://doi.org/10.1016/S0005-7967\(03\)00120-7](https://doi.org/10.1016/S0005-7967(03)00120-7)
- 795 Moritz, S., Külz, A., Voderholzer, U., Hillebrand, T., McKay, D., & Jelinek, L. (2019).  
796 “Phobie à deux” and other reasons why clinicians do not apply exposure with response  
797 prevention in patients with obsessive–compulsive disorder. *Cognitive Behaviour*  
798 *Therapy*, 48(2), 162–176. <https://doi.org/10.1080/16506073.2018.1494750>
- 799 National Institute for Health and Clinical Excellence (NICE). (2006). *Obsessive-Compulsive*  
800 *Disorder: Core Interventions in the Treatment of Obsessive-Compulsive Disorder and*  
801 *Body Dysmorphic Disorder*. British Psychological Society.
- 802 Netter, A. L., Hessler, J. B., Diedrich, A., Lehr, D., Berking, M., Pfeuffer, S., &  
803 Voderholzer, U. (2018). Usability and naturalness of videoconference-based exposure  
804 and response prevention for obsessive- compulsive disorder at the patients’ homes: A  
805 pilot study. *Journal of Cognitive Psychotherapy*, 32(4), 303–314.  
806 <https://doi.org/10.1891/0889-8391.32.4.303>
- 807 Öst, L. G., Havnen, A., Hansen, B., & Kvale, G. (2015). Cognitive behavioral treatments of  
808 obsessive-compulsive disorder. A systematic review and meta-analysis of studies  
809 published 1993-2014. In *Clinical Psychology Review* (Vol. 40, pp. 156–169).  
810 Pergamon. <https://doi.org/10.1016/j.cpr.2015.06.003>
- 811 Palardy, V., El-Baalbaki, G., Fredette, C., Rizkallah, E., & Guay, S. (2018). Social Support  
812 and Symptom Severity Among Patients With Obsessive-Compulsive Disorder or Panic  
813 Disorder With Agoraphobia: A Systematic Review. *Europe’s Journal of Psychology*,  
814 14(1), 254–286. <https://doi.org/10.5964/ejop.v14i1.1252>
- 815 Pinto, A., Liebowitz, M. R., Foa, E. B., & Simpson, H. B. (2011). Obsessive compulsive  
816 personality disorder as a predictor of exposure and ritual prevention outcome for

## PREDICTORS OF TREATMENT OUTCOME IN OCD

34

- 817 obsessive compulsive disorder. *Behaviour Research and Therapy*, 49(8), 453–458.  
818 <https://doi.org/10.1016/j.brat.2011.04.004>
- 819 R.-Mercier, A., Masson, M., Bussi eres, E.-L., & Cellard, C. (2018). Common  
820 transdiagnostic cognitive deficits among people with psychiatric disorders exposed to  
821 childhood maltreatment: A meta-analysis. *Cognitive Neuropsychiatry*, 23(3), 180–197.  
822 <https://doi.org/10.1080/13546805.2018.1461617>
- 823 Raffin, A. L., Guimar aes Fachel, J. M., Ferr ao, Y. A., Pasquoto de Souza, F., & Cordioli, A.  
824 V. (2009). Predictors of response to group cognitive-behavioral therapy in the treatment  
825 of obsessive-compulsive disorder. *European Psychiatry: The Journal of the*  
826 *Association of European Psychiatrists*, 24(5), 297–306.  
827 <https://doi.org/10.1016/j.eurpsy.2008.12.001>
- 828 Reid, A. M., Garner, L. E., Van Kirk, N., Gironde, C., Kropfing, J. W., Brennan, B. P.,  
829 Mathes, B. M., Monaghan, S. C., Tiff, E. D., Andr e, M.-C., Cattie, J., Crosby, J. M., &  
830 Elias, J. A. (2017). How willing are you? Willingness as a predictor of change during  
831 treatment of adults with obsessive-compulsive disorder. *Depression and Anxiety*, May,  
832 1–8. <https://doi.org/10.1002/da.22672>
- 833 Revelle, W. R. (2017). *psych: Procedures for Personality and Psychological Research*.  
834 [https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-](https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research)  
835 [personality-and-psychological-research](https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research)
- 836 Rief, W., Glombiewski, J. A., Gollwitzer, M., Schub o, A., Schwarting, R., & Thorwart, A.  
837 (2015). Expectancies as core features of mental disorders. *Current Opinion in*  
838 *Psychiatry*, 28(5), 378–385. <https://doi.org/10.1097/YCO.0000000000000184>
- 839 Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E., Moons, K. G. M., &  
840 Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction  
841 model: Part I - Continuous outcomes. *Statistics in Medicine*, 38(7), 1262–1275.  
842 <https://doi.org/10.1002/sim.7993>
- 843 Riley, R. D., Snell, K. I., Ensor, J., Burke, D. L., Harrell Jr, F. E., Moons, K. G., & Collins,  
844 G. S. (2019). Minimum sample size for developing a multivariable prediction model:  
845 PART II - binary and time-to-event outcomes. *Statistics in Medicine*, 38(7), 1276–1296.  
846 <https://doi.org/10.1002/sim.7992>

## PREDICTORS OF TREATMENT OUTCOME IN OCD

35

- 847 Roth, C., Siegl, J., Aufdermauer, N., & Reinecker, H. (2004). Therapie von Angst- und  
 848 Zwangspatienten in der verhaltenstherapeutischen Praxis. *Verhaltenstherapie, 14*(1),  
 849 16–21. <https://doi.org/10.1159/000078027>
- 850 Ruscio, A. M., Stein, D. J., Chiu, W. T., & Kessler, R. C. (2010). The epidemiology of  
 851 obsessive-compulsive disorder in the National Comorbidity Survey Replication.  
 852 *Molecular Psychiatry, 15*(1), 53–63. <https://doi.org/10.1038/mp.2008.94>
- 853 Salkovskis, P. M. (1999). Understanding and treating obsessive-compulsive disorder.  
 854 *Behaviour Research and Therapy, 37*(SUPPL. 1), 29–52.  
 855 [https://doi.org/10.1016/S0005-7967\(99\)00049-2](https://doi.org/10.1016/S0005-7967(99)00049-2)
- 856 Schruers, K., Koning, K., Luermans, J., Haack, M. J., & Griez, E. (2005). Obsessive-  
 857 compulsive disorder: A critical review of therapeutic perspectives. *Acta Psychiatrica*  
 858 *Scandinavica, 111*(4), 261–271. <https://doi.org/10.1111/j.1600-0447.2004.00502.x>
- 859 Simpson, H. B., Maher, M. J., Wang, Y., Bao, Y., Foa, E. B., & Franklin, M. (2011). Patient  
 860 adherence predicts outcome from cognitive behavioral therapy in obsessive-compulsive  
 861 disorder. *Journal of Consulting and Clinical Psychology, 79*(2), 247–252.  
 862 <https://doi.org/10.1037/a0022659>
- 863 Skapinakis, P., Caldwell, D. M., Hollingworth, W., Welton, N. J., Fineberg, N., Salkovskis,  
 864 P., Churchill, R., & Lewis, G. (2016). Network meta-analyses and treatment  
 865 recommendations for obsessive-compulsive disorder - Authors' reply. *The Lancet*  
 866 *Psychiatry, 3*(10), 921–922. [https://doi.org/10.1016/S2215-0366\(16\)30282-6](https://doi.org/10.1016/S2215-0366(16)30282-6)
- 867 Steketee, G., Chambless, D. L., & Tran, G. Q. (2001). Effects of axis I and II comorbidity on  
 868 behavior therapy outcome for obsessive-compulsive disorder and agoraphobia.  
 869 *Comprehensive Psychiatry, 42*(1), 76–86. <https://doi.org/10.1053/comp.2001.19746>
- 870 Steketee, G., Eisen, J., Dyck, I., Warshaw, M., & Rasmussen, S. (1999). Predictors of course  
 871 in obsessive compulsive disorder. *Psychiatry Research, 89*(3), 229–238.  
 872 [https://doi.org/10.1016/S0165-1781\(99\)00104-3](https://doi.org/10.1016/S0165-1781(99)00104-3)
- 873 Steketee, G., Frost, R., & Bogart, K. (1996). The YALE-BROWN obsessive compulsive  
 874 scale: Interview versus self-report. *Behaviour Research and Therapy, 34*(8), 675–684.  
 875 [https://doi.org/10.1016/0005-7967\(96\)00036-8](https://doi.org/10.1016/0005-7967(96)00036-8)



## PREDICTORS OF TREATMENT OUTCOME IN OCD

36

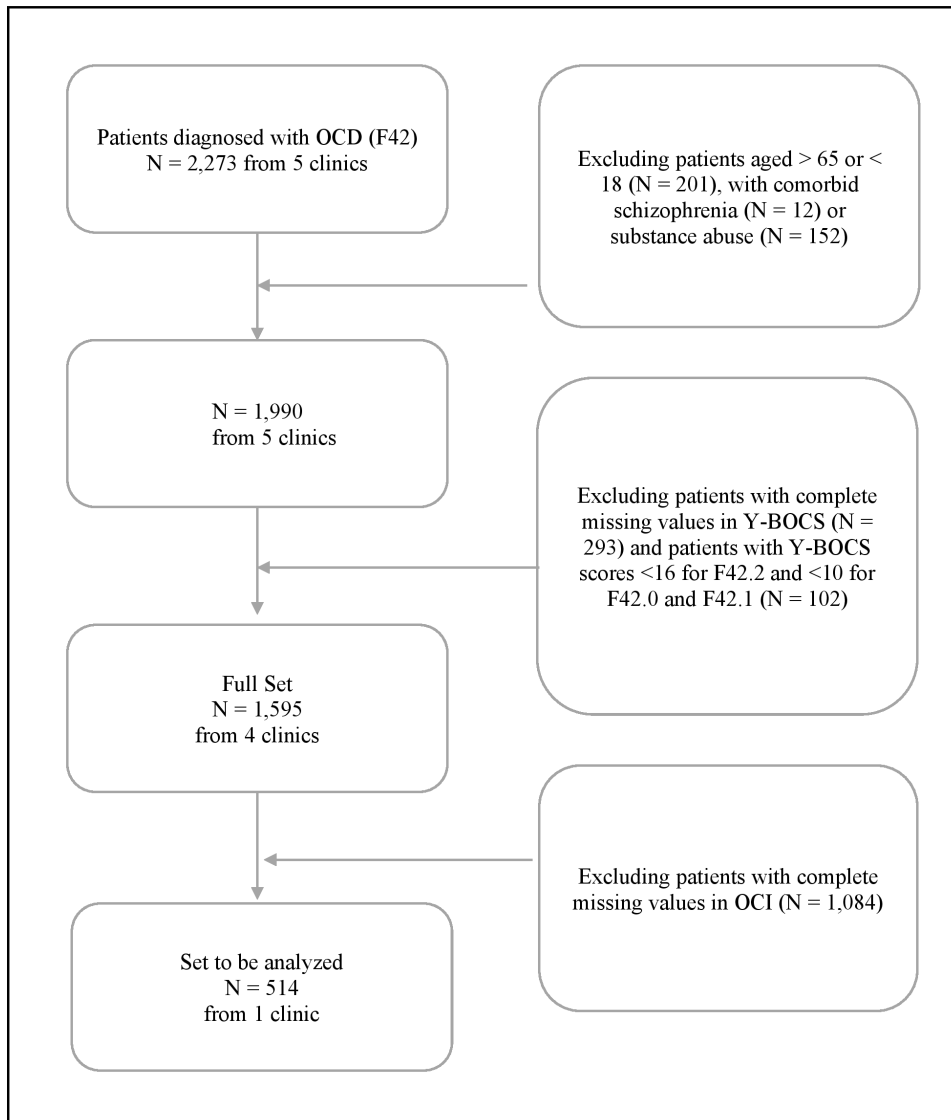
- 876 Steketee, G., Frost, R. O., Tolin, D. F., Rasmussen, J., & Brown, T. A. (2010). Waitlist-  
 877 controlled trial of cognitive behavior therapy for hoarding disorder. *Depression and*  
 878 *Anxiety*, 27(5), 476–484. <https://doi.org/10.1002/da.20673>
- 879 Steketee, G., Siev, J., Fama, J. M., Keshaviah, A., Chosak, A., & Wilhelm, S. (2011).  
 880 Predictors of treatment outcome in modular cognitive therapy for obsessive-compulsive  
 881 disorder. *Depression and Anxiety*, 28(4), 333–341. <https://doi.org/10.1002/da.20785>
- 882 Stekhoven, D. J., & Bühlmann, P. (2012). Missforest-Non-parametric missing value  
 883 imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118.  
 884 <https://doi.org/10.1093/bioinformatics/btr597>
- 885 Stengler-Wenzke, K., Kroll, M., Matschinger, H., & Angermeyer, M. C. (2006). Subjective  
 886 quality of life of patients with obsessive–compulsive disorder. *Social Psychiatry and*  
 887 *Psychiatric Epidemiology*, 41(8), 662–668. <https://doi.org/10.1007/s00127-006-0077-8>
- 888 Storch, E. A., Lewin, A. B., Farrell, L., Aldea, M. A., Reid, J., Geffken, G. R., & Murphy, T.  
 889 K. (2010). Does cognitive-behavioral therapy response among adults with obsessive–  
 890 compulsive disorder differ as a function of certain comorbidities? *Journal of Anxiety*  
 891 *Disorders*, 24(6), 547–552. <https://doi.org/10.1016/j.janxdis.2010.03.013>
- 892 Sydow, K. von, & Borst, U. (2018). *Systemische Therapie in der Praxis*. Beltz.  
 893 [https://www.beltz.de/fachmedien/psychologie/buecher/produkt\\_produktdetails/34680-](https://www.beltz.de/fachmedien/psychologie/buecher/produkt_produktdetails/34680-systemische_therapie_in_der_praxis.html)  
 894 [systemische\\_therapie\\_in\\_der\\_praxis.html](https://www.beltz.de/fachmedien/psychologie/buecher/produkt_produktdetails/34680-systemische_therapie_in_der_praxis.html)
- 895 Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Pearson Education.
- 896 Twohig, M. P., Hayes, S. C., Plumb, J. C., Pruitt, L. D., Collins, A. B., Hazlett-Stevens, H.,  
 897 & Woidneck, M. R. (2010). A randomized clinical trial of acceptance and commitment  
 898 therapy versus progressive relaxation training for obsessive-compulsive disorder.  
 899 *Journal of Consulting and Clinical Psychology*, 78(5), 705–716.  
 900 <https://doi.org/10.1037/a0020508>
- 901 van der Heiden, C., van Rossen, K., Dekker, A., Damstra, M., & Deen, M. (2016).  
 902 Metacognitive therapy for obsessive-compulsive disorder: A pilot study. *Journal of*  
 903 *Obsessive-Compulsive and Related Disorders*, 9, 24–29.  
 904 <https://doi.org/10.1016/j.jocrd.2016.02.002>

## PREDICTORS OF TREATMENT OUTCOME IN OCD

37

- 905 Van Noppen, B., & Steketee, G. (2009). Testing a conceptual model of patient and family  
906 predictors of obsessive compulsive disorder (OCD) symptoms. *Behaviour Research*  
907 *and Therapy*, 47(1), 18–25. <https://doi.org/10.1016/j.brat.2008.10.005>
- 908 Veale, D., Naismith, I., Miles, S., Gledhill, L. J., Stewart, G., & Hodsoll, J. (2016).  
909 Outcomes for residential or inpatient intensive treatment of obsessive-compulsive  
910 disorder: A systematic review and meta-analysis. *Journal of Obsessive-Compulsive and*  
911 *Related Disorders*, 8, 38–49. <https://doi.org/10.1016/j.jocrd.2015.11.005>
- 912 Voderholzer, U., Schlegl, S., Diedrich, A., Külz, A. K., Thiel, N., Hertenstein, E., Schwartz,  
913 C., Rufer, M., Herbst, N., Nissen, C., Hillebrand, T., Osen, B., Stengler, K., Jelinek, L.,  
914 & Moritz, S. (2015). Versorgung Zwangserkrankter mit kognitiver Verhaltenstherapie  
915 als Behandlungsmethode erster Wahl. *Verhaltenstherapie*, 25(3), 183–190.  
916 <https://doi.org/10.1159/000438717>
- 917 Voderholzer, U., Schwartz, C., Freyer, T., Zurovski, B., Thiel, N., Herbst, N., Wahl, K.,  
918 Kordon, A., Hohagen, F., & Kuelz, A. K. (2013). Cognitive functioning in medication-  
919 free obsessive-compulsive patients treated with cognitive-behavioural therapy. *Journal*  
920 *of Obsessive-Compulsive and Related Disorders*, 2(3), 241–248.  
921 <https://doi.org/10.1016/j.jocrd.2013.03.003>
- 922 Vogel, P. A., Hansen, B., Stiles, T. C., & Gøtestam, K. G. (2006). Treatment motivation,  
923 treatment expectancy, and helping alliance as predictors of outcome in cognitive  
924 behavioral treatment of OCD. *Journal of Behavior Therapy and Experimental*  
925 *Psychiatry*, 37(3), 247–255. <https://doi.org/10.1016/j.jbtep.2005.12.001>
- 926 Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., Marrero, J.,  
927 Zhu, J., & Higgins, P. (2013). Comparison of imputation methods for missing  
928 laboratory data in medicine. *BMJ Open*, 3(8), e002847.  
929 <https://doi.org/10.1136/bmjopen-2013-002847>
- 930 Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for*  
931 *what makes psychotherapy work*. Routledge.
- 932 Wells, A. (2009). *Metacognitive therapy for anxiety and depression*. Guilford Press.
- 933

934 **Figure 1.** Patient flow diagram



## PREDICTORS OF TREATMENT OUTCOME IN OCD

39

935 Table 1  
 936 *Sample characteristics of the full sample (n = 1,595) and OCI-R (n = 514) subsample*

Characteristics	OCI-R Sample (n=514)	Full Sample (n=1595)	p
Age at admission <i>M</i> (SD)	34.3 (12.2)	33.9 (11.7)	.304
Sex <i>n</i> (%)			.986
Male	204 (39.7)	631 (39.6)	
Female	310 (60.3)	964 (60.4)	
Educational score <i>M</i> (SD) <sup>a</sup>	3.3 (0.8)	3.3 (0.8)	.862
OCD subtype <i>n</i> (%) <sup>b</sup>			<.001
Predominantly obsessional thoughts (F42.0)	28 (5.4)	101 (6.3)	
Predominantly compulsive actions (F42.1)	128 (24.9)	243 (15.2)	
Mixed thoughts and actions (F42.2)	354 (68.9)	1247 (78.2)	
Unspecified (F42.9)	4 (0.8)	4.0 (0.3)	
Number of mental comorbidities <i>M</i> (SD) <sup>b</sup>	1.4 (1.0)	1.3 (1.0)	.002
Most frequent mental comorbidities <i>n</i> (%) <sup>b</sup>			
Depressive episode (F32)	142 (27.6)	491 (30.8)	.068
Recurrent depressive disorder (F33)	231 (44.9)	659 (41.3)	.049
Phobic disorder (F40)	73 (14.2)	188 (11.8)	.048
Personality disorder (F6)	81 (15.7)	224 (14.0)	.200
Housing situation <i>n</i> (%)			<.001
Alone	125 (20.4)	220 (13.8)	
Co-living with partner or family	269 (44.0)	474 (29.7)	
Co-living with parents	31 (5.1)	321 (20.1)	
Institutional placement	4 (0.7)	10 (0.6)	
Shared flat, private flat, furnished room	162 (26.5)	509 (31.9)	
No fixed household	1 (0.3)	8 (0.5)	
Flat-sharing community	18 (2.9)	29 (1.8)	
Marital status <i>n</i> (%)			<.001
Married	184 (35.8)	493 (30.9)	
Divorced	15 (2.9)	65 (4.1)	
Widowed	4 (0.1)	5 (0.3)	
Single	311 (60.5)	1032 (64.7)	
In a relationship <i>n</i> (%)	251 (48.9)	790 (49.6)	
Occupational status <i>n</i> (%)			.357
Unemployed	125 (25.1)	374 (23.5)	
Retired	57 (11.4)	173 (10.9)	
Student, in training, home care	99 (19.9)	328 (20.6)	
Working full time	152 (30.5)	498 (30.7)	
Working half time	43 (8.6)	144 (9.0)	
Working occasionally	19 (3.8)	41 (2.6)	
Ability to work <i>n</i> (%)	235 (45.7)	708 (44.5)	.523
First inpatient treatment <i>n</i> (%)	319 (74.7)	927 (79.1)	.006
Outpatient psychotherapy <i>n</i> (%)	373 (72.6)	1033 (64.8)	<.001
Outpatient psychiatric treatment <i>n</i> (%)	343 (67.0)	1000 (62.9)	.022
Y-BOCS score at baseline <i>M</i> (SD) <sup>c</sup>	24.7 (5.5)	25.5 (5.6)	<.001
OCI-R mean score at baseline <i>M</i> (SD) <sup>d</sup>	30.2 (12.3)	30.2 (12.3)	
PHQ-9 score at baseline <i>M</i> (SD) <sup>e</sup>	12.5 (6.0)	12.7 (6.0)	.425
GAF score at baseline <i>M</i> (SD) <sup>f</sup>	45.7 (9.4)	45.7 (8.5)	.976
BSI mean score at baseline <i>M</i> (SD) <sup>g</sup>	1.35 (.7)	1.3 (0.7)	.103
SWLS score at baseline <i>M</i> (SD) <sup>h</sup>	14.7 (6.8)	15.1 (6.8)	.132

937 *Note.* *M*=mean, *SD*=Standard deviation, *n*=number, *p*=*p*-value from tests for difference between samples with and without OCI-R values in full set; *t*-  
 938 test for continuous variables, chi-squared-test for dichotomous variables and Fisher's test for categorical variables with more than two levels; <sup>a</sup> Based  
 939 on the German school system; scale from 0 (no degree) to 4 (general qualification for university entrance); <sup>b</sup> Diagnosis as given by practitioners  
 940 according to ICD-10; <sup>c</sup> Y-BOCS = Yale-Brown Obsessive Compulsive Scale: 10 items, scale 0–40; <sup>d</sup> OCI-R = Obsessive Compulsive Inventory-  
 941 Revised: 18 items, scale 0–72; <sup>e</sup> PHQ-9 = Patient Health Questionnaire-9: 9 items scale 0–27; <sup>f</sup> GAF = Global Assessment of Functioning: Scale  
 942 0–100; <sup>g</sup> BSI = Brief Symptom Inventory: 53 items, scale 0–4; <sup>h</sup> Satisfaction With Life Scale: 5 items, scale 5–35.

## PREDICTORS OF TREATMENT OUTCOME IN OCD

40

943 Table 2  
 944 Means (*M*), standard deviations (*SD*), and effect sizes (*ES*) at pre- and post-treatment for the  
 945 OCR-I subsample (*N* = 514)

	<i>Pre</i>		<i>Post</i>		<i>t</i>	<i>p</i>	<i>ES</i>	95% CI   <i>ES</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
OCI-R total	1.68	.68	.98	.63	23.80	<.001	1.17	1.02; 1.31
Washing	2.06	1.48	1.15	1.14	16.87	<.001	.84	.70; .99
Obsessing	2.60	1.18	1.72	1.16	16.72	<.001	.83	.69; .98
Hoarding	.88	1.05	.53	.80	9.11	<.001	.46	.32; .60
Ordering	1.55	1.34	.91	1.11	13.53	<.001	.67	.53; .82
Checking	1.99	1.33	1.05	1.00	18.66	<.001	.92	.78; 1.06
Neutralizing	.94	1.12	.51	.82	10.26	<.001	.52	.38; .67
Y-BOCS total	24.66	5.53	15.55	7.04	25.77	<.001	1.25	1.10; 1.40
Compulsions	2.57	.64	1.58	.80	25.07	<.001	1.20	1.05; 1.34
Obsessions	2.42	.72	1.60	.81	19.88	<.001	.99	.84; 1.13
BDI-II	26.54	11.40	15.03	11.15	19.86	<.001	1.12	.95; 1.29
GSI (BSI)	121.22	60.80	77.24	55.99	18.24	<.001	.86	.72; 1.00
PHQ-9	12.52	5.98	7.85	5.23	18.11	<.001	.86	.72; 1.00
GAD-7	1.59	.66	.97	.62	22.17	<.001	1.06	.92; 1.20
PHQ-15	.74	.40	.54	.35	14.66	<.001	.70	.56; .83
GAF	45.69	9.42	58.78	11.67	27.21	<.001	1.24	1.38; 1.10
SWLS	14.70	6.81	18.15	6.99	12.08	<.001	.57	.71; .44

946 *Note.* OCI-R = Obsessive Compulsive Inventory-Revised; Y-BOCS = Yale-Brown Obsessive Compulsive Scale;  
 947 BDI-II = Beck Depression Inventory-II; GSI (BSI) = Global Severity Index of the Brief Symptom Inventory;  
 948 PHQ-9 = Patient Health Questionnaire-9; GAD-7 = General Anxiety Disorder-7; PHQ-15 = Patient Health  
 949 Questionnaire-15; SWLS = Satisfaction with Life Scale; GAF = Global Assessment of Functioning.

## PREDICTORS OF TREATMENT OUTCOME IN OCD

41

950 Table 3  
 951 Regression results using Y-BOCS total score post-treatment as the criterion for the OCR-I  
 952 subsample ( $N = 514$ )  
 953

Predictor	<i>beta</i>	<i>beta</i> 95% CI	<i>p</i>	<i>sr</i> <sup>2</sup>	<i>sr</i> <sup>2</sup> 95% CI	<i>r</i>	Fit <i>R</i> <sup>2</sup>	Difference $\Delta R^2$
(Intercept)								
Baseline	.37**	[0.28, 0.45]	<.001	.13	[NA, NA]	.37**		
							.133**	
(Intercept)								
Baseline	.33**	[0.21, 0.45]	<.001	.06	[.02, .10]	.37**		
Distress	.21**	[0.10, 0.32]	<.001	.03	[-.00, .05]	.33**		
Somatic disorders	.03	[-0.06, 0.12]	.512	.00	[-.00, .01]	.11*		
Obsessing	.04	[-0.05, 0.13]	.417	.00	[-.00, .01]	.18**		
Social support	-.10*	[-0.19, -0.01]	.036	.01	[-.01, .02]	-.07		
Ordering	.07	[-0.04, 0.17]	.201	.00	[-.01, .01]	.26**		
Chronic depression	.01	[-0.08, 0.10]	.825	.00	[-.00, .00]	.04		
Depression	.01	[-0.08, 0.10]	.792	.00	[-.00, .00]	.10*		
Academic	.02	[-0.07, 0.10]	.695	.00	[-.00, .00]	.03		
Disability	.08	[-0.01, 0.18]	.083	.01	[-.01, .02]	.23**		
Washing behavior	-.21**	[-0.33, -0.10]	<.001	.03	[-.00, .05]	.12*		
							.220**	.087**
								95% CI
								[.04, .13]

954 Note. A significant beta-weight indicates that semi-partial correlations are also significant. *beta* indicates the  
 955 standardized regression weights. *sr*<sup>2</sup> represents the semi-partial correlation squared. *r* represents the zero-order  
 956 correlation. \* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

957

958

## PREDICTORS OF TREATMENT OUTCOME IN OCD

42

959 Table 4  
 960 Regression results using Y-BOCS sub-scores Obsessions and Compulsions post-treatment as  
 961 the criterion for the OCR-I subsample ( $N = 514$ )  
 962

Y-BOCS Obsessions as criterion								
Predictor	<i>beta</i>	<i>beta</i> 95% CI	<i>p</i>	<i>sr</i> <sup>2</sup>	<i>sr</i> <sup>2</sup> 95% CI	<i>r</i>	Fit <i>R</i> <sup>2</sup>	Difference $\Delta R^2$
(Intercept)								
Baseline	.45**	[0.37, 0.54]	<.001	.21		.45**		
							.206**	
(Intercept)								
Baseline	.37**	[0.27, 0.47]	<.001	.10	[.05, .15]	.45**		
Distress	.24**	[0.14, 0.35]	<.001	.04	[.01, .07]	.29**		
Somatic disorders	.03	[-0.06, 0.12]	.471	.00	[-.00, .01]	.10*		
Obsessing	.14**	[0.05, 0.23]	.002	.02	[-.00, .04]	.30**		
Social support	-.03	[-0.12, 0.05]	.446	.00	[-.00, .01]	-.02		
Ordering	-.00	[-0.10, 0.09]	.919	.00	[-.00, .00]	.12*		
Chronic depression	-.00	[-0.08, 0.08]	.999	.00	[-.00, .00]	.00		
Depression	-.01	[-0.09, 0.08]	.879	.00	[-.00, .00]	.05		
Academic	.02	[-0.06, 0.11]	.605	.00	[-.00, .00]	.02		
Disability	.06	[-0.04, 0.15]	.241	.00	[-.01, .01]	.20**		
Washing behavior	-.24**	[-0.33, -0.14]	<.001	.04	[.01, .07]	.01	.294**	.088** 95% CI [.04, .13]
963 Y-BOCS Compulsions as criterion								
Predictor	<i>beta</i>	<i>beta</i> 95% CI	<i>p</i>	<i>sr</i> <sup>2</sup>	<i>sr</i> <sup>2</sup> 95% CI	<i>r</i>	Fit <i>R</i> <sup>2</sup>	Difference $\Delta R^2$
(Intercept)								
Baseline	0.36**	[0.27, 0.45]	<.001	.13	[NA, NA]	.36**		
							.130**	
(Intercept)								
Baseline	0.32**	[0.19, 0.46]	<.001	.04	[.01, .07]	.36**		
Distress	0.12*	[0.01, 0.23]	.035	.01	[-.01, .02]	.26**		
Somatic disorders	0.01	[-0.09, 0.10]	.880	.00	[-.00, .00]	.08		
Obsessing	-0.08	[-0.17, 0.01]	.083	.01	[-.01, .02]	-.02		
Social support	-0.09	[-0.18, 0.01]	.065	.01	[-.01, .02]	-.07		
Ordering	0.12*	[0.01, 0.23]	.028	.01	[-.01, .03]	.31**		
Chronic depression	0.03	[-0.06, 0.11]	.550	.00	[-.00, .01]	.04		
Depression	0.05	[-0.04, 0.14]	.282	.00	[-.01, .01]	.13**		
Academic	0.00	[-0.09, 0.09]	.992	.00	[-.00, .00]	.04		
Disability	0.06	[-0.03, 0.15]	.204	.00	[-.01, .01]	.17**		
Washing behavior	-0.10	[-0.23, 0.04]	.171	.00	[-.01, .01]	.24**	.198**	.059** 95% CI [.02, .10]

964 Note. A significant beta-weight indicates that semi-partial correlations are also significant. *beta* indicates the  
 965 standardized regression weights. *sr*<sup>2</sup> represents the semi-partial correlation squared. *r* represents the zero-order  
 966 correlation. \* indicates  $p < .05$ . \*\* indicates  $p < .01$ .  
 967

## PREDICTORS OF TREATMENT OUTCOME IN OCD

43

968 Table 5  
 969 Regression results using PHQ-9 score post-treatment as the criterion for the OCR-I  
 970 subsample ( $N = 514$ )  
 971

Predictor	<i>beta</i>	<i>beta</i> 95% CI	<i>p</i>	<i>sr</i> <sup>2</sup>	<i>sr</i> <sup>2</sup> 95% CI	<i>r</i>	Fit <i>R</i> <sup>2</sup>	Difference $\Delta R^2$
(Intercept)								
Baseline	.57**	[0.49, 0.64]	<.001	.32		.57**		
							.320**	
(Intercept)								
Baseline	.28**	[0.15, 0.42]	<.001	.06	[-.00, .05]	.57**		
Distress	.31**	[0.17, 0.44]	<.001	.03	[-.00, .05]	.57**		
Somatic disorders	-.01	[-0.09, 0.07]	.794	.00	[-.00, .00]	.09		
Obsessing	.01	[-0.07, 0.09]	.778	.00	[-.00, .00]	.13**		
Social support	-.12**	[-0.20, -0.04]	.002	.01	[-.00, .03]	-.12**		
Ordering	.04	[-0.04, 0.13]	.303	.00	[-.00, .01]	.29**		
Chronic depression	.05	[-0.02, 0.13]	.158	.00	[-.00, .01]	.09		
Depression	.00	[-0.08, 0.08]	.967	.00	[-.00, .00]	.18**		
Academic	-.02	[-0.10, 0.05]	.592	.00	[-.00, .00]	-.00		
Disability	.12**	[0.04, 0.20]	.003	.01	[-.00, .03]	.29**		
Washing behavior	-.08	[-0.16, 0.01]	.073	.03	[-.01, .01]	.20**		
							.391**	.071**
								95% CI
								[.03, .11]

972 Note. A significant beta-weight indicates that semi-partial correlations are also significant.  
 973 *beta* indicates the standardized regression weights. *sr*<sup>2</sup> represents the semi-partial correlation  
 974 squared. *r* represents the zero-order correlation. \* indicates  $p < .05$ . \*\* indicates  $p < .01$ .



PREDICTORS OF TREATMENT OUTCOME IN OCD

975

**Supplementary material**

976

**Supplemental material 1.** Factor loadings and communalities

977 Table 2

978 *Factor loadings and communalities for the clinical sample (N=514): pattern matrix.*

Variable	Factor										Communality
	Distress	Somatic Disorders	Obsessing	Social Support	Ordering	Chronicity Depression	Comorbid Depression	Academic	Functional Disability	Washing Behavior	
Age		.03		.50	.09	.01					.70
Gender					-.15					.46	.39
First inpatient treatment			-.01		.03				-.06	.04	.06
Ability to work									<b>-.59</b>		.58
Outpatient psychotherapy				.03	-.16			-.05			.10
Psychiatric treatment			.02				.09		.27	-.04	.22
Children				<b>.56</b>							.68
SWLS			.01	.19	-.10				-.07	-.12	.40
BSI somatization	.18	.01									.53
BSI obsession compulsion	.15				.11						.67
BSI interpersonal sensitivity	.27		-.03								.67
BSI depression	.26										.69
BSI hostility	.21										.51
BSI phobic anxiety	.12		.02						.03		.40
BSI paranoid ideation	.23		-.12								.58
BSI psychoticism	.31										.69
BSI pst	<b>.43</b>										.80
BSI psdi	.37										.69
PHQ mood	.28										.72
PHQ anxiety	.26		.03								.67
PHQ physical symptoms	.25										.59
BSI anxiety	.28										.70
Number of Diagnosis		<b>.61</b>									.75
GAF		.05							-.42	-.06	.40
OCI-R washing										<b>.60</b>	.62
OCI-R obsessing	.11		.32								.43
OCI-R hoarding					.40						.41

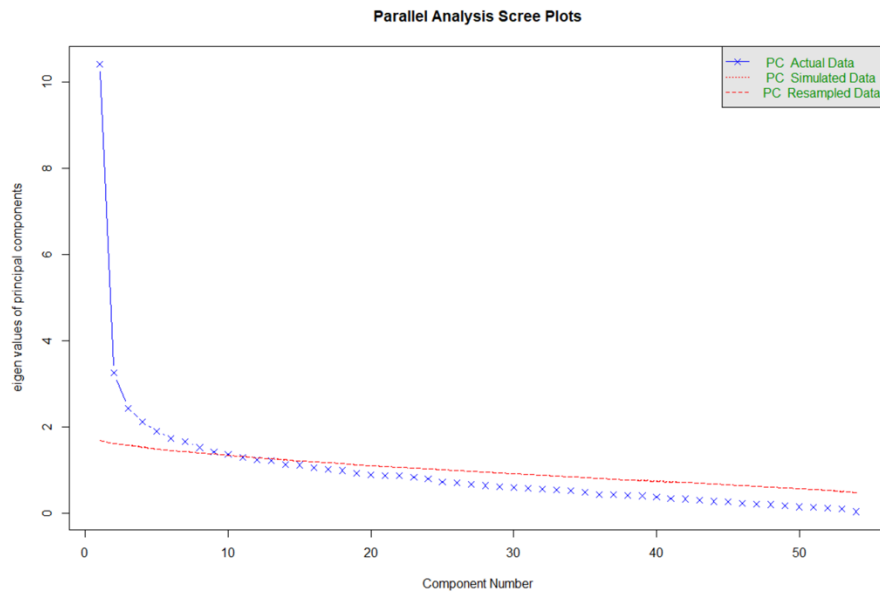
PREDICTORS OF TREATMENT OUTCOME IN OCD

Variable	Factor										Communality
	Distress	Somatic Disorders	Obsessing	Social Support	Ordering	Chronicity Depression	Comorbid Depression	Academic	Functional Disability	Washing Behavior	
OCI-R ordering			-.02		<b>.50</b>						.59
OCI-R checking					.43				-.01		.44
OCI-R neutralizing			.07		.30						.32
Y-BOCS Behavior					.16					.56	.63
Y-BOCS Thoughts	.01		.18					-.02	.13	.07	.29
Diagnosis F32.1		.23	-.07						.02		.23
Diagnosis F33.1						<b>-.69</b>	.17				.81
Diagnosis E						.65	.20				.80
Diagnosis F40		.01					<b>.73</b>				.80
Diagnosis F42.1									.16		.10
Diagnosis F42.2			-.57								.85
Diagnosis F6			<b>.65</b>								.83
Diagnosis G				-.04	.01				.24		.18
Diagnosis I		.19								.13	.21
Diagnosis M5		.22		.01	.12						.28
Diagnosis M62.89		.35									.44
Diagnosis M		.26	.06		-.09		.03				.29
Diagnosis N		.49					-.56				.68
Educational level			.05		-.18		.12		-.12	.04	.23
Living alone			-.07	-.01	.02						.14
Job trained											.74
Academic Degree									<b>-.66</b>		.74
Unemployed					.15				<b>.66</b>		.28
Retired			-.05	.13		.08				.37	.41
Married				.49							.55
In a relationship					.03	-.04				-.01	.05

979 Note. Extraction method: sparse principal component analysis. BSI: Brief Symptom Inventory; pst: positive symptom total; psdi: positive symptom distress index; SWLS:  
980 Satisfaction With Live Scale; OCI-R: Obsessive Compulsive Inventory-Revised; Y-BOCS: Yale-Brown Obsessive Compulsive Scale (Self Report); GAF: General  
981 Assessment of Functioning.

982  
983

### Supplemental material 2. Scree plot



984

985

986

987

988

989

#### 990 **Data availability statement**

991 Data sharing is not applicable to this article as only secondary analyses were performed on a  
992 routinely collected data set with permission to use for research purposes but without explicit  
993 permission of data sharing by the patients.

## Anhang D: Studie 4

Running head: Predictors of change of inpatient psychotherapy for depression

1

1

2 **Title:** Inpatient psychotherapy for depression in a large routine psychosomatic care sample:  
3 A Bayesian approach to examining effectiveness and predictors of change

4

**Short title:**

5

Effectiveness and predictors of change of inpatient psychotherapy for depression

6

7

Philipp Herzog\*,

8

Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,  
9 Gutenbergstraße 18, D-35032 Marburg, Germany

10

Matthias Feldmann,

11

Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,  
12 Gutenbergstraße 18, D-35032 Marburg, Germany,

13

Tobias Kube,

14

Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,  
15 Gutenbergstraße 18, D-35032 Marburg, Germany,

16

Gernot Langs,

17

Birkenweg 10, Psychosomatic Clinic,

18

Birkenweg 10, D-24576 Bad Bramstedt, Germany

19

Thomas Gärtner,

20

Schön-Klinik Bad Arolsen, Psychosomatic Clinic,

21

Hofgarten 10, D-34454 Bad Arolsen, Germany

22

Elisabeth Rauh,

23

Schön-Klinik Bad Staffelstein, Psychosomatic Clinic,

24

Am Kurpark 11, D-96231 Bad Staffelstein, Germany

## Predictors of change of inpatient psychotherapy for depression

2

- 25 Robert Doerr,  
26 Schön-Klinik Berchtesgadener Land, Psychosomatic Clinic,  
27 Malterhöh 1, D-83471 Schönau am Königssee, Germany
- 28 Andreas Hillert,  
29 Schön-Klinik Roseneck, Psychosomatic Clinic,  
30 Am Roseneck 6, D-83209 Prien am Chiemsee, Germany
- 31 Ulrich Voderholzer,  
32 Schön-Klinik Roseneck, Psychosomatic Clinic,  
33 Am Roseneck 6, D-83209 Prien am Chiemsee, Germany
- 34 Winfried Rief,  
35 Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,  
36 Gutenbergstraße 18, D-35032 Marburg, Germany,
- 37 Dominik Endres,  
38 Philipps-University of Marburg, Department of Theoretical Neuroscience, Gutenbergstraße  
39 18, D-35032 Marburg, Germany,  
40 and Eva-Lotta Brakemeier,  
41 Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy,  
42 Gutenbergstraße 18, D-35032 Marburg, Germany
- 43 \*Correspondence to be addressed to: Philipp Herzog, Philipps-University of Marburg,  
44 Department of Clinical Psychology and Psychotherapy, Gutenbergstraße 18, D-35032  
45 Marburg, Germany, Email: philipp.herzog@staff.uni-marburg.de, Telephone: +49 6421  
46 2824061

Predictors of change of inpatient psychotherapy for depression

3

47 **Conflict of interest statement**

48 The authors declare no conflict of interest.

49 **Funding**

50 The authors received no funding from an external source for this work.

51 **Acknowledgement**

52 The authors thank Dr. Stefan Koch (Schön Clinic Roseneck, Prien am Chiemsee, Germany)  
53 for his assistance to the description of the treatment program.

54 **Author Contributions**

55 PH and ELB conceptualized the study. PH originally drafted and edited the manuscript, ELB  
56 supervised, reviewed and edited the manuscript, MF computed the statistical analyses, DE  
57 supervised the statistical analyses, TK revisited the manuscript critically for important  
58 intellectual content, WR gave substantial feedback for the conception and design of this  
59 study and GL, TG, ER, RD, AH and UV routinely collected data and were significantly  
60 involved in the acquisition of this routine data for the present work.

61 **Statement of Ethics**

62 Hereby, we confirm that we complied with the guidelines for human studies and that the  
63 research was conducted ethically in accordance with the [World Medical Association](#)  
64 [Declaration of Helsinki](#). Yet, ethical approval was not obtained for the present retrospective  
65 study because the data was routinely collected as a standard clinic diagnostic procedure to  
66 ensure quality assurance and did therefore not involve prospective evaluation. Nevertheless,  
67 all patients gave informed consent to anonymous evaluations of their routinely collected  
68 data.

Predictors of change of inpatient psychotherapy for depression

4

69

### Abstract

70 *Objective:* A routinely collected dataset was analyzed (1) to determine the effectiveness of  
71 inpatient psychotherapy for depression in routine psychosomatic care, and (2) to identify  
72 potential predictors of change.

73 *Method:* In a large sample of 22,681 inpatients with depression ( $M = 47.3$  years [ $SD = 12.7$ ],  
74 58.1% female), pre-post and pre-follow-up effect sizes were computed for various outcome  
75 variables to determine effectiveness (Hedges'  $g$ ). To build a probabilistic model of predictors  
76 of change, an independent component analysis generated components from demographic and  
77 clinical data, and Bayesian EFA extracted factors from the available pre-test, post-test and  
78 follow-up questionnaires in a subsample of  $N=6,377$  inpatients. To select the best-fitted  
79 model, the BIC of different path models were compared. A Bayesian path analysis was  
80 performed to identify the most important factors to predict changes.

81 *Results:* Pre-post effect sizes were large for the primary outcome and moderate for various  
82 secondary outcomes. Almost all pretreatment factors exerted significant influences on  
83 different baseline factors. Several factors were found to be resistant to change during  
84 treatment: suicidality, agoraphobia, life dissatisfaction, physical disability and pain. The  
85 strongest cross-loadings were observed from suicidality on negative cognitions, from  
86 agoraphobia on anxiety, and from physical disability on perceived disability.

87 *Conclusions:* The results indicate large effects of inpatient psychotherapy for depression in  
88 routine clinical care. The strong direct influence of pretreatment factors decreased over the  
89 course of treatment. However, some factors appeared stable and difficult to treat, which  
90 might hinder treatment outcome. The results of different predictors of change are discussed.

91

### Keywords

92 Predictors of change, prediction, depression, effectiveness, routine clinical care

93

**Public Health Significance Statements**

94

- This study suggests that the naturalistic effectiveness of an inpatient treatment program for depression is large, which might serve as a benchmark of the effectiveness of depression treatment in routine inpatient care.

95

96

97

- Sociodemographic and clinical pretreatment factors influence baseline symptomatology and might be considered as risk factors for depression, but their impact decreased over the course of treatment.

98

99

100

- The main factors contributing to persistent psychopathology in depression are: suicidality; agoraphobia; life dissatisfaction; physical disability; and pain.

101

102

- When treating inpatients with depression, the strongest predictors of change are suicidality on negative cognitions; agoraphobia on anxiety; and physical disability on perceived disability.

103

104

105

- These prognostic variables should be addressed at an early treatment stage in order to improve short- and long-term treatment outcome.

106



107

**Introduction**

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

Depression is a highly prevalent and disabling mental disorder that is listed among the top ten causes of years lived with disability (Vos et al., 2015). Over the years, knowledge about the psychopathology of depression has substantially increased, and, for example, the network approach indicates that depressive symptoms might be better conceptualized within dynamic systems of causally interrelated symptoms (Robinaugh et al., 2019). Our understanding of the psychopathological mechanisms involved in the development and maintenance of depression has significantly increased: Recent developments again highlight distorted cognitive processes, but recast them using a predictive processing perspective (Kube, Schwarting, et al., 2019). Encouragingly, patients suffering from depression can be treated effectively: Many randomized controlled trials (RCTs) and meta-analyses have demonstrated the general efficacy and effectiveness of psychotherapy for depression on average (Lepping et al., 2017), although effect sizes vary greatly depending on the population under investigation, the treatment setting and the particular methodological approach. Overall, large short- and long-term effects can be observed for cognitive behavioral therapies (CBT) and interpersonal psychotherapy (IPT) in episodic depression (Lemmens et al., 2015; Lemmens et al., 2019); however, effect sizes are considerably low for the treatment of persistent depressive disorder (Cuijpers, van Straten, et al., 2010; Schramm et al., 2017, 2019), treatment-resistant depression and inpatient settings (Schefft et al., 2019). When considering all forms of psychotherapies and depressive subtypes, a new network meta-analysis suggests that the combination of psychotherapy and pharmacotherapy might be superior to monotreatment for patients with moderate depression in terms of short-term effects (Cuijpers et al., 2020) without investigating long-term effects. Although psychotherapy for depression thus seems to be effective on average, psychotherapy research on the treatment of depression has to face several challenges.

132 A critical issue in psychotherapy research is the research-practitioner gap, that is, the lack  
133 in the translation of what is known of effective depression treatments (evidence-based  
134 treatments) and what is provided to patients in routine clinical care (nonacademic  
135 community-based practice) (Kilbourne et al., 2012). Notwithstanding some controversies  
136 about the precise effect sizes (ES) of previous RCTs (Cuijpers et al., 2019; Cuijpers et al.,  
137 2019), little is known about the effectiveness of treatments for depression under routine  
138 clinical care conditions, that is, how effective current treatments in the “real world” are.  
139 Methodologically, this refers to clinical effectiveness studies under day-to-day circumstances  
140 (Phase IV-studies) with high external validity. Within the framework of practice-oriented  
141 research (Castonguay et al., 2013), an estimation of this effectiveness of depression  
142 treatments can serve as both an indicator for the implementation of evidence-based  
143 treatments for depression and a benchmark for depression treatment (Merrill et al., 2003;  
144 Minami et al., 2007, 2008). To date, such benchmark studies in this respect are rare, however  
145 (Barkham et al., 2010; Wampold & Imel, 2015). Therefore, our first objective was to  
146 determine the effectiveness of psychotherapy for depression under naturalistic conditions in  
147 a large, routinely collected dataset of German psychosomatic clinics that might serve as a  
148 benchmark.

149 Moreover, an important second challenge refers to the large heterogeneity of patients  
150 suffering from depression and their individual response to the treatment delivered. For  
151 example, despite well-known differences between patients with episodic and persistent  
152 disorders (Bird et al., 2018; Köhler et al., 2018), these differences have long been neglected  
153 in treatment delivery and have only recently been considered systematically (e.g. Boschloo  
154 et al., 2019). Most of the treatments investigated in RCTs were applied following a ‘one-  
155 size-fits-all’ approach on the ‘average patient’, with highly standardized manuals and strict  
156 inclusion and exclusion criteria, which may decrease the external validity and fail to take

157 inter-individual differences between patients into account. In fact, despite extensive research  
158 and the development of countless new treatment approaches (Goldfried, 2019), the efficacy  
159 of depression treatment seems to stagnate: One-third to half of depressed patients do not  
160 benefit from the treatment, both in terms of non-response and/or relapse (Lambert, 2017; van  
161 Bronswijk et al., 2019). For example, about 65 – 80% of the patients still report clinically  
162 significant residual symptoms at posttreatment (Lemmens et al., 2015). Thus, since the  
163 response to specific psychological treatments for depression varies greatly among  
164 individuals, a better understanding of the predictors of symptom change would be of great  
165 value enabling to optimize the depression treatments delivered (Grant et al., 2014; Simon &  
166 Perlis, 2010).

167 As a consequence, the focus of psychotherapy research concerning the treatment of  
168 depression has actually shifted to both the individual selection of treatments (Cohen &  
169 DeRubeis, 2018) and personalized treatments (Grant et al., 2014; Simon & Perlis, 2010) in  
170 recent years. As such, the ability to identify patients at risk for not benefitting from treatment  
171 yields some promising advantages (Webb et al., 2020). A treatment selection at the  
172 beginning, in which each patient can be assigned to a treatment that is most helpful to them,  
173 and a treatment tailored to important predictors of change in the short and long run might be  
174 able to reduce the potential of a chronic course of depression. At the same time, this research  
175 strategy could prevent treatment-resistance and its associated factors: reduce morbidity and  
176 mortality; shorten suffering; reduce rates of treatment dropouts; and reduce the cost burden  
177 for the patient and society (Bergfeld et al., 2018; Cepeda et al., 2018; Johnston et al., 2019;  
178 Reutfors et al., 2018), which usually allocates those patients to a difficult-to-treat population  
179 from a practitioner's perspective. In the context of personalizing treatments, the

180 identification of prognostic predictors by single-treatment studies<sup>1</sup> and prescriptive  
181 predictors by multiple-treatment studies<sup>2</sup> might enable to better select and tailor treatments to  
182 the individual patient (Cohen & DeRubeis, 2018; Webb et al., 2020). Promisingly, a few  
183 variables could be consistently identified as robust negative general predictors of treatment  
184 outcome across studies: high baseline depressive symptom severity; persistent or recurrent  
185 depressive disorder; early onset of depression; and comorbid Axis I disorders (Blom et al.,  
186 2007; Carter et al., 2011; Constantino et al., 2008; Frank et al., 2011; Jarrett et al., 1991;  
187 Mynors-Wallis & Gath, 1997). In addition to these clinical characteristics, the following  
188 cognitive factors have been more inconsistently identified as predictors in single studies:  
189 dysfunctional attitudes (Jarrett et al., 1991); perfectionism (Enns et al., 2002); beliefs about  
190 the causes of depression (Addis & Jacobson, 1996); perceived logic of therapy and  
191 childhood reasons of depression (Carter et al., 2011); cognitive functioning, particularly in  
192 terms of executive functions (Groves et al., 2018); patients' expectations of treatment  
193 outcome (Gaston et al., 1989; Ravitz et al., 2011); self-efficacy concerning coping strategies  
194 (Stiles-Shields et al., 2015); and repetitive negative thinking (Kertz et al., 2015). Those  
195 general predictors as prognostic variables have guided research so far and might be useful as  
196 indicators of patients who are difficult to treat (Frank et al., 2011). Nevertheless, the  
197 prediction of treatment outcome appears an ongoing challenge. In fact, these partially  
198 inconsistent findings were mainly traced back to methodological issues (Lee et al., 2018;  
199 Simon & Perlis, 2010). Some major limitations of previous studies result from their lack of  
200 power due to small sample sizes, the separate inclusion of sociodemographic or clinical or  
201 psychological data and statistical issues. Naturalistic studies with larger and super-regionally  
202 representative case numbers in the context of a practice-research network (Castonguay &

---

<sup>1</sup> Sometimes referred to as general predictors.

<sup>2</sup> Sometimes referred to as differential predictors.

## Predictors of change of inpatient psychotherapy for depression

10

203 Muran, 2015; Castonguay & Muran, 2016) might thus help address such statistical problems  
204 and to overcome the research-practice gap. Furthermore, some researchers call for new  
205 statistical models that use multiple resources of data by combining clinical and demographic  
206 data (Chekroud et al., 2016; Cipriani & Geddes, 2016) and the advantages of Bayesian  
207 statistics (Wagenmakers et al., 2018). For example, one recent predictor analysis with  
208 depressed patients using Bayesian statistics has found no link between the pretreatment  
209 symptom severity or the presence and severity of comorbid anxiety with treatment response  
210 to a web-based CBT; while being married or cohabitating, higher life satisfaction, lower  
211 level of dysfunctional thinking and more depressive episodes predicted a better outcome  
212 (Høifødt et al., 2015).

213 Aiming to addressing some of the aforementioned shortcomings, the current study has  
214 two objectives: (1) to determine the naturalistic short- and long-term effectiveness of  
215 specialized inpatient psychotherapy for depression, that is, benchmarking inpatient  
216 psychotherapy for depression in routine clinical psychosomatic care in Germany; and (2) to  
217 identify its predictors of change, that is, the underlying processes by which this treatment  
218 works in the short and the long run. Therefore, this study tries to shift the focus on  
219 candidates for predictors of change by combining a theory- and data-driven approach and  
220 applying Bayesian statistics (cf., Høifødt et al., 2015) with the superior goal of informing  
221 feature selection for treatment development and refinement.

222

**Methods**

223 Ethical approval was not obtained for this retrospective study because data was  
224 routinely collected as a standard clinic diagnostic procedure to ensure quality assurance and  
225 did therefore not involve prospective evaluation. This study was conducted in accordance  
226 with the ethical standards as laid out in the 1964 Declaration of Helsinki and its subsequent  
227 amendments. All patients gave informed consent to anonymous evaluations of their routinely  
228 collected data.

**Sample**

230 For the current study, we were able to analyze data routinely and consecutively  
231 collected between 2013 and 2017 from depression inpatients treated in five German acute  
232 psychosomatic clinics with specialized inpatient depression treatment as recommended in  
233 international and national guidelines for the treatment of complex and severe depression  
234 (DGPPN et al., 2017; National Institute for Health and Clinical Excellence (NICE), 2018).  
235 Inclusion criteria were: primary diagnosis of depression (F32 or F33 according to ICD-10);  
236 age of at least 18 years; and a focus of the treatment delivered mainly on depression. Taking  
237 into account the complexity and heterogeneity of patients suffering from depression, no  
238 further exclusion criteria were defined a priori (such as medication status, comorbid  
239 diagnoses, etc.). In total, 22,681 patients met these inclusion and exclusion criteria.

**Routine clinical treatment for depression**

241 The psychosomatic clinics offer multimodal modern CBT treatment programs for  
242 depression based on national and international guidelines (DGPPN et al., 2017; National  
243 Institute for Health and Clinical Excellence (NICE), 2018). The multiple sites (i.e. clinics)  
244 are part of one large clinic group. Inpatient psychotherapy of depression in German  
245 psychosomatic clinics aims to deliver a maximum of symptom-specific therapy for  
246 depression with a focus on behavioral activation and cognitive interventions based on Beck's

Predictors of change of inpatient psychotherapy for depression 12

247 CBT manual (Beck, 1979) for major depressive disorder (MDD) and – meanwhile in most of  
248 the clinics – CBASP for persistent depressive disorder (PDD) augmented with multimodal  
249 groups being based on modern approaches of depression treatment such as ACT<sup>3</sup> and other  
250 mindfulness-based interventions. Since inpatients generally suffer from more severe and  
251 more complicated illnesses than outpatients and often have no direct follow-up treatment  
252 after discharge, the following stages of the therapeutic process are of particular importance:  
253 motivational training and psychoeducation, behavioral training and cognitive interventions,  
254 transfer into daily routine and relapse prevention. The therapeutic elements of the routine  
255 treatments of the five clinics are outlined below:

256 All depressive patients receive individual psychotherapy once a week (50 minutes), in  
257 which modern CBT strategies are used (see above).

258 In addition, depressive patients take part in the following group therapies:

259 Psychoeducation group therapy (minimum 8 sessions at 2h = 16h) including individual  
260 behavior analyses of critical situations, cognitive interventions and identification of  
261 functional and dysfunctional coping behavior; group training of social skills and coping with  
262 emotions (minimum 8 sessions at 2h = 16h) (Berking, 2015; Hinsch & Pflingsten, 2015;  
263 Moses & Barlow, 2006).

264 If patients meet the criteria of a PDD diagnosis, they receive individual and group  
265 therapies tailored for PDD based on McCullough's principles of CBASP<sup>4</sup> (McCullough et  
266 al., 2015), including situation analyses with a focus on interactional behavior (Kiesler's  
267 interpersonal circumplex), disciplined personal involvement and roleplay of functional social  
268 behavior (minimum 8 sessions at 2h = 16h).

---

<sup>3</sup> Acceptance and Commitment Therapy

<sup>4</sup> Cognitive Behavioral Analysis System of Psychotherapy

269 Additional multimodal interventions include interactional ward groups (including  
270 elements of ACT) or mindfulness training groups (minimum 1h per week), activating sports  
271 and body therapy groups (minimum 3 hours per week), access to symptom-specific group  
272 therapy for comorbid disorders (e.g. anxiety disorders) or coping with job stress (8 sessions  
273 at 2h = 16h), and, according to individual indication, progressive muscle relaxation, art  
274 therapy, biofeedback and social counseling (by trained social workers) (minimum 2 hours  
275 per week). Therapy is delivered under weekly team supervision and a weekly visit by a  
276 senior physician.

277 In total, inpatients in these five psychosomatic clinics receive a minimum of 1 hour per  
278 week of individual psychotherapy, 8-12 hours per week of group therapy plus access to an  
279 average of 6-10 hours per week of multimodal interventions over an average of 6-8 weeks,  
280 leading to a total of 78-184 therapy hours according to individual indication and depending  
281 on comorbidities. In our sample, the average length of stay was  $M = 43.3$  days ( $SD = 14.8$ ).

282 For the patient-staff ratio on each depression ward, a minimum of two licensed  
283 psychotherapists are available (3-5 years of training in CBT; one medical doctor and one  
284 certified clinical psychologist); 1-2 therapists in CBT training; and at least 1,5 certified  
285 nursing staff members. These training levels and regular supervision guarantee the adherence  
286 to the CBT-based treatment program, including elements of CBASP and ACT. During the  
287 weekends, emergency care in the clinic with medical, psychological and nursing personal is  
288 warranted.

289 Psychopharmacological treatment is administered when indicated according to current  
290 national and international guidelines (DGPPN et al., 2017; National Institute for Health and  
291 Clinical Excellence (NICE), 2018), based on the informed consent of the patients. In the  
292 dataset for our analysis, data from the medication status at discharge were included. Notably,



Predictors of change of inpatient psychotherapy for depression 14

293 the major treatment focus in these psychosomatic clinics is placed on psychotherapeutic  
294 work in individual and group settings.

295 **Assessment/Measures**

296 Concerning effectiveness, change in depression was the primary outcome assessed at  
297 admission and discharge using the following measures: the Beck Depression Inventory-II  
298 (BDI-II) (Beck et al., 1996) and the Patient Health Questionnaire-9 (PHQ-9 (Kroenke et al.,  
299 2001)). Both instruments were used in order to enhance comparability with previous studies  
300 (Gyani et al., 2013; Knapstad et al., 2018). The BDI-II is an internationally used and well  
301 established 21-item scale assessing somatic, cognitive and affective symptoms of depression.  
302 The sum score ranges from 0 to 63, with higher scores reflecting more severe depressive  
303 symptoms. The BDI-II has demonstrated sound psychometric properties (Beck et al., 1996).  
304 The Patient Health Questionnaire-9 (PHQ-9) is one module of the German PHQ-D and a  
305 self-administered questionnaire assessing depressive symptoms (Gräfe et al., 2004).  
306 Extensive research supports its psychometric properties in terms of validity and reliability  
307 (Kroenke et al., 2010).

308 The secondary outcomes, which were also assessed at admission and discharge, capturing  
309 psychological distress, (health-related) and general quality of life, were assessed by the Brief  
310 Symptom Inventory (BSI), the Short-Form-36 (SF-36) and the Satisfaction with Life Scale  
311 (SWLS).

312 The BSI (Derogatis & Melisaratos, 1983) consists of 53 items and is the short version of  
313 the SCL-90-R assessing nine primary symptom dimension: Somatization (SOM); Obsessive-  
314 Compulsiveness (OBS); Interpersonal Sensitivity (INS); Depression (DEP); Anxiety (ANX);  
315 Hostility (HOS); Phobic Anxiety (PHOB); Paranoid Ideation (PAR); and Psychoticism  
316 (PSY). Additionally, global indices of distress can be calculated (General Symptom Index;  
317 GSI). Validation studies of the German version have shown a satisfactory test-retest

318 reliability (Franke, 1997). The Short Form-36 Health Survey (SF-36) (Ware et al., 1993) is a  
319 self-rated generic questionnaire assessing the perceived health-related quality of life or  
320 health status. The SF-36 contains eight domains: Physical Functioning (PF); Role - Physical  
321 (RP); Bodily Pain (BP); General Health (GH); Vitality (VT); Social Functioning (SF); Role -  
322 Emotional (RE); and Mental Health (MH). The SWLS (Diener et al., 1985) assesses  
323 satisfactions with one's life as a whole. The 5-item scale was designed to measure a person's  
324 global judgement of life satisfaction and does not allow judgements in life domains (e.g.  
325 health, finances). The SWLS has shown good psychometric properties, such as high internal  
326 consistency, high temporal stability, good convergent validity with other scales of subjective  
327 well-being and discriminant validity (as measured with scales of emotional well-being).  
328 Moreover, the SWLS has sufficient sensitivity to be useful in detecting changes in life  
329 satisfaction during the course of treatment.

330 Medication in eight classes of substances (namely, antidepressants, neuroleptics,  
331 tranquilizers, anticonvulsants, narcotics, substitution, analgesics, other medication) was  
332 assessed only at discharge. At follow-up (six months after discharge), only the BDI-II, PHQ-  
333 9 and SWLS were assessed. This data collection was part of the follow-up routine outcome  
334 management of some hospitals, which is why only a limited number of questionnaires were  
335 collected.

#### 336 **Other measures for computing path models**

337 Sixteen clinical and sociodemographic variables were asked about at intake via a self-  
338 report and therapist-rated questionnaires. Sociodemographic data included age, sex,  
339 education level and marital status, among others (see sample characteristics in the Results  
340 section). Clinical characteristics were assessed at intake, including diagnosis and previous  
341 psychotherapeutic and psychiatric treatments. In addition to the aforementioned  
342 questionnaires, a self-report questionnaire assessing social and work-related adjustment

343 (Mundt et al., 2002) was applied at intake and discharge. In addition, as two other modules  
344 of the PHQ-D, the General Anxiety Disorder-7 (GAD-7) was used as a self-administered  
345 questionnaire measuring generalized anxiety disorder, and the Patient Health Questionnaire-  
346 15 (PHQ-15) was used as a questionnaire administering somatic symptoms. Furthermore, the  
347 Global Assessment of Functioning (GAF) (Endicott et al., 1976) was used as an observer-  
348 rated index to determine psychological, social and occupational functioning levels on a  
349 continuum ranging from severe mental illness (0) to mental health (100).

### 350 **Statistics**

351 All analyses were performed using the statistical processing language R (R Core  
352 Team, 2018). To determine the effectiveness of the inpatient depression treatment, a  
353 corrected Hedge's  $g$  pre-post and, where data was available, pre-FU effect sizes (ES) were  
354 used for the primary and secondary outcomes (see above). To set an upper limit on the  
355 potential impact of treatment discontinuation on effect sizes, we also calculated last  
356 observation carried forward (LOCF) effect sizes.

357 In order to gain new insights into (Granger-) causal pathways of depression course in the  
358 short run directly after treatment and in the long run after completing treatment in a 6-month  
359 follow-up, and to assess model selection uncertainty, we decided to choose Bayesian path  
360 analysis, which is described in more detail below.

### 361 **Dimensionality reduction**

362 In total, 16 variables were available from demographic and clinical variables ('trait').  
363 Parallel analysis was performed to estimate the optimal number of components in the R  
364 package *psych* (Revelle, 2017). Afterwards, missing data were imputed using the algorithm  
365 *missForest* (Stekhoven, 2013; Stekhoven & Bühlmann, 2012) and independent components  
366 were constructed using independent component analysis (ICA) with the function *fastICA*

367 from the R package *ica* (Helwig, 2015). All hypothesized 'trait' variables were entered into  
368 one to build unrelated components, leading to a total number of 8 components.

369 For self-report data as 'state' variables, 6 questionnaires (see above) with a total of 148  
370 individual items were available at T1 and T2, with less than 80% missing data. Due to the  
371 overlap of constructs in the original 23 subscales, a Bayesian exploratory factor analysis with  
372 the function *befa* from the R package *BayesFM* (Piatek, 2017) was used to find a lower-  
373 dimensional representation of the data. Starting with the original 23 subscales factor model,  
374 the analyses comprised two steps: In the first step, there were no restrictions on the  
375 identification of factors, yielding a model with 21 factors. The second step used the results of  
376 the first step as an initialization, but here, a restriction of five items per factor was applied in  
377 order to filter for the most important factors, yielding 15 factors. Priors were based on the  
378 largest naturalistic example of another study (Conti et al., 2014). The factors were  
379 interpreted and named based on the content of their items. Confirmatory factor analysis was  
380 conducted in the R package *lavaan* (Rosseel, 2012) using the 23-factor model and the 15-  
381 factor model, and a difference in Bayesian information criterion (BIC) was computed to  
382 ensure the superiority of the smaller model.

### 383 **Model selection**

384 All available factors were clustered into 5 blocks of factors: the 8 sociodemographic-  
385 clinical factors assessed at pre-treatment; medication assessed at posttreatment; the 15 self-  
386 report factors, assessed at pre-treatment and posttreatment; as well as 4 of the 15 self-report  
387 factors at follow-up (because only for these 4 factors were data available, see above). Ten  
388 maximum likelihood path models with full maximum likelihood imputation were  
389 constructed using the R package *lavaan* (Rosseel, 2012), in which associations between  
390 factors were enabled or disabled in blocks, in order to test the basic structure of the data and  
391 to build probabilistic models of change in depression. Block-wise path coefficients were

392 estimated by computing the average of the absolutes of all path-coefficients from one block  
393 to another. Due to the missing data structure of the dataset, there were strong statistical  
394 artifacts when using the data from all clinics. These artifacts could be replicated when  
395 artificially imposing the same pattern of missing data in the data from just one clinic, where  
396 missing data was minimal. Thus, in order to avoid these artifacts, the subsequent path  
397 analysis was based on the data of only one hospital (N= 6,377). Fit measures were computed  
398 for all 10 path models, and the best model was selected.

#### 399 **Coefficient Estimation**

400 In order to gain more robust estimates for the individual path coefficients, the best path  
401 model was re-computed using a Bayesian path model with the R package *blavaan* (Merkle &  
402 Rosseel, 2015), which used the Bayesian statistical programming language *stan* in the  
403 background. A normal prior with a standard deviation of 0.05 was placed on each  
404 standardized path coefficient in order to enforce a slight ridge-like regularization and reduce  
405 false positives (Jacobucci & Grimm, 2018).

406

**Results****407 Patient characteristics**

408 The full sample consisted of 22,681 inpatients with a mean age of  $M = 47.3$  years  
409 ( $SD = 12.7$ ), 58.1% of whom were female and 78.5% in a relationship. More than half of the  
410 patients (54.5%) were married; 49.9% were fully employed, 13.3% were unemployed, and  
411 only 33.6% were able to work. For most of the patients (81.5%), the treatment received in  
412 the respective hospital was their first inpatient treatment; however, 58.1% had already been  
413 treated with outpatient psychotherapy, and 61.2% of the patients had received outpatient  
414 psychiatric treatment in the past.

415 According to the ICD-10 and regarding the depression subtype, 36.7% of the sample met  
416 the criteria of a depression with a single episode (F32), whereas the majority (63.1%) was  
417 diagnosed with recurrent depression (F33); only a few patients met the criteria of a persistent  
418 mood [affective] disorder (0.2%). On average, patients had 1.9 ( $SD = 1.1$ ) comorbid mental  
419 disorders, with neurotic, stress-related and somatoform disorders (F4) being among the most  
420 frequent comorbidities (together accounting for 39.2%). A post-hoc analysis revealed that  
421 phobic anxiety disorders (F40) with  $n = 2,406$  (10.6%) and somatoform disorders (F45) with  
422  $n = 4,669$  (20.6%) accounted for the high percentage of disorders within the F4-spectrum.  
423 The majority of the sample (83.7%) was discharged as normal.

424 The mean BDI-II sum score at baseline was  $M = 28.2$  ( $SD = 11.0$ ), and the mean PHQ-9  
425 sum score at baseline was  $M = 14.5$  ( $SD = 5.6$ ), indicating moderate to severe depressive  
426 symptoms on average. Concerning functioning, the mean GAF score at baseline was  $M =$   
427  $48.5$  ( $SD = 8.9$ ), and the mean BSI score at baseline was  $M = 1.3$  ( $SD = 0.7$ ), indicating  
428 moderate to high general psychological distress and impairment in psychosocial adaptation.  
429 In terms of life satisfaction, the SWLS was rather low ( $M = 16.8$ ;  $SD = 6.9$ ).

430 Table 1 depicts the relevant sociodemographic and clinical characteristics of the full  
431 sample (N = 22,681).

432 Insert Table 1 here.

#### 433 **Effectiveness of the inpatient treatment**

434 In the full sample, the corrected Hedge's *g* pre-post and pre-FU ES for relevant outcomes  
435 were as follows: For the primary treatment outcome, Hedge's *g* of the pre-post ES in the  
436 BDI-II was 1.32 (95% CI [1.29, 1.34]), whereas the LOCF-corrected Hedge's *g* was 1.14  
437 (95% CI [1.12, 1.16]), indicating large effects. Hedge's *g* of the pre-post ES in the PHQ-9  
438 was 1.15 (95% CI [1.13, 1.17]), and the LOCF-corrected ES was 1.04 (95% CI [1.02, 1.06]).  
439 For general psychopathological distress, Hedge's *g* of the pre-post ES in the BSI was .99  
440 (95% CI [.97, 1.02]), and the LOCF-corrected Hedge's *g* was .91 (95% CI [.89, .94]). For  
441 quality of life, Hedge's *g* of the pre-post ES in the SWLS was .49 (95% CI [.47, .51]), and its  
442 LOCF-corrected ES was .40 (95% CI [.38, .42]); Hedge's *g* for the SF-36 was .81 (95% CI  
443 [.77, .85]), whereas its LOCF-corrected ES was .77 (95% CI [.73, .80]).

444 For the primary outcome in patients, where data was available, Hedge's *g* of the pre-FU  
445 ES in the PHQ-9 was .74 (95% CI [.72, .76]), and the ES for the SWLS was .31 (95% CI  
446 [.29, .33]).

#### 447 **Testing different path models**

448 Ten path models were computed as plausible candidate models on the temporal structure  
449 of the data, based on research findings concerning risk factors of depression and predictors  
450 of treatment outcome (see Supplemental Material 1), and their comparative fit indices were  
451 computed. The comparative fit indices are displayed in Supplemental material 2, with lower  
452 fit indices indicating a better fit of the model to the data. The BIC, BIC2, AIC and CFI of  
453 Model 2 were 0, indicating the best model fit, followed by Model 8 and Model 1. Therefore,  
454 based on these criteria, Model 2 was selected for further analysis.





480 = 0.22) at baseline. Furthermore, work-related problems were strongly positively related to  
 481 negative cognitions ( $\beta = 0.13$ ), paranoia ( $\beta = 0.12$ ), concentration difficulties ( $\beta = 0.12$ ),  
 482 agoraphobia ( $\beta = 0.13$ ), perceived disability ( $\beta = 0.15$ ) and life dissatisfaction ( $\beta = 0.18$ ) at  
 483 baseline and concentration difficulties ( $\beta = 0.12$ ) at discharge. Treatment resistance was  
 484 positively correlated with anxiety ( $\beta = 0.12$ ), concentration difficulties ( $\beta = 0.12$ ) and  
 485 perceived disability ( $\beta = 0.13$ ) at baseline. Moreover, interpersonal dysfunction exerted a  
 486 strong positive influence on negative cognitions ( $\beta = 0.15$ ), aggressiveness ( $\beta = 0.12$ ),  
 487 paranoia ( $\beta = 0.14$ ) and agoraphobia ( $\beta = 0.12$ ) at baseline. Female gender was strongly  
 488 positively correlated with negative cognitions ( $\beta = 0.15$ ), loss of interest ( $\beta = 0.12$ ),  
 489 rumination ( $\beta = 0.13$ ), somatic anxiety symptoms ( $\beta = 0.14$ ), feeling of weakness ( $\beta = 0.19$ ),  
 490 physical disability ( $\beta = 0.14$ ) and pain ( $\beta = 0.22$ ). Interestingly, the wish for retirement only  
 491 had a positive influence on physical disability at baseline ( $\beta = 0.19$ ) and discharge ( $\beta =$   
 492 0.12).

#### 493 **Investigating the stability of psychopathology**

494 The autoregressive coefficients and explained variance of the self-report data from  
 495 clinical baseline data (T1) on clinical data at discharge (T2) and on clinical data at follow-up  
 496 (T3) are displayed in Table 3.

497 

Insert Table 3 here.

498 Notably, for factors with autoregressive coefficients over  $\beta = 0.5$  and at least  $R^2 = .35$ ,  
 499 suicidality ( $\beta = 0.52$ ), agoraphobia ( $\beta = 0.56$ ), life dissatisfaction ( $\beta = 0.60$ ), physical  
 500 disability ( $\beta = 0.65$ ) and pain ( $\beta = 0.52$ ) seem to be among the most stable factors from T1 to  
 501 T2, despite the treatment received, and can therefore be considered as treatment-resistant  
 502 factors. In the long run, the influence of psychopathological factors at baseline and discharge  
 503 on the same factors at follow-up diminished to some extent for life dissatisfaction (T1 on T3:  
 504  $\beta = 0.28$ , and T2 on T3:  $\beta = 0.32$ , respectively) and pain (T1 on T3:  $\beta = 0.32$ , and T2 on T3:

Predictors of change of inpatient psychotherapy for depression

23

505  $\beta = 0.25$ , respectively); and almost vanished for loss of interest (T1 on T3:  $\beta = 0.13$ , and T2  
506 on T3:  $\beta = 0.15$ , respectively) and rumination (T1 on T3:  $\beta = 0.18$ , and T2 on T3:  $\beta = 0.21$ ,  
507 respectively).

#### 508 **Unveiling underlying predictors of change**

509 The cross-loadings are delineated in Figure 2. Remarkably, the strongest influence can be  
510 observed from suicidality at T1 on negative cognitions at T2 ( $\beta = 0.14$ ); from agoraphobia at  
511 T1 on anxiety at T2 ( $\beta = 0.15$ ); and from physical disability at T1 on perceived disability at  
512 T2 ( $\beta = 0.13$ ). Interestingly, the cross-loadings of clinical T2 data on clinical follow-up data  
513 (T3) seem to be rather small overall. All coefficients range between  $-0.09 \leq \beta \leq 0.09$ .  
514 However, to gain some insights in an exploratory manner of the psychopathological state at  
515 discharge on the follow-up state beside the autocorrelations, these coefficients revealed small  
516 relationships of physical disability on pain, rumination and loss of interest. Furthermore,  
517 both perceived disability and concentration difficulties had effects on rumination and loss of  
518 interest; somatic anxiety symptoms influenced pain; and suicidality had an impact on  
519 rumination and loss of interest.

520

Insert Figure 2 here.

521

**Discussion**

522 This study aimed: (1) to examine the effectiveness of inpatient psychotherapy for  
523 depression in a naturalistic psychosomatic setting in Germany; and (2) to unravel some  
524 possible causal pathways through which treatment effects in depression are achieved, both at  
525 discharge and at a 6-month follow-up.

526 With regard to the first objective, we added to current discussions related to the  
527 absolute efficacy of psychotherapy for depression (Cuijpers et al., 2019; Cuijpers et al.,  
528 2019; Munder et al., 2019; Munder et al., 2019) by using a large inpatient sample  
529 (N=22,681) to compute effect sizes for treatment according to routine clinical care. In our  
530 sample, we found large pre-post and pre-follow-up effect sizes across several important  
531 symptom-specific and more global outcomes. This is in line with other findings, for  
532 example, the Norwegian IAPT showed pre-post-changes of  $ES = 1.1$  for depression using  
533 PHQ-9 in large multicenter cohort study (Knapstad et al., 2018). In our study, we used the  
534 BDI-II in addition to the PHQ-9 in order to enhance comparability with previous studies  
535 (Gyani et al., 2013; Knapstad et al., 2018). Notably, the scores obtained from self-report  
536 measures such as BDI-II are always somewhat lower than data obtained from clinician-rated  
537 measures such as the Hamilton Rating Scale for Depression (HRSD) (Cuijpers, Li, et al.,  
538 2010; Lambert et al., 1986). Therefore, speculatively, the effect sizes found in our analysis  
539 might be somewhat underestimated, such that the real effect sizes could actually have been  
540 even higher if observer-rated measures such as the HRSD had been used. Nevertheless, these  
541 ES estimations based on the large dataset might serve as a benchmark of short- and long-  
542 term effectiveness of specialized inpatient depression treatment in routine clinical care in  
543 Germany (cf., Schefft et al., 2019). Since the large effects found in our analyses speak to a  
544 generally high effectiveness of psychotherapy for depression in this naturalistic inpatient

545 setting, one can conclude that the so-called research-practice gap appears to be not as large  
546 as expected (Kilbourne et al., 2012).

547 To address the second objective of our study, we followed recent developments in  
548 computational psychiatry (Friston et al., 2014) to incorporate translational approaches into  
549 psychotherapy research by using Bayesian statistics. In particular, we wanted to approach –  
550 at least in an initial step – the complexity of depression as proposed in a new guidance paper  
551 of the Lancet Psychiatry Commission, especially Part 1 (“How do existing treatments  
552 work?”) and Part 8 (“Whom should we treat, for what, and with what?”) (Holmes et al.,  
553 2018). Three key findings in respect to predictors of change in depression treatment can be  
554 obtained from our analyses. First, we found that reducing the level of physical disability  
555 through inpatient psychotherapy had positive effects on perceived disability. This finding is  
556 consistent with the traditional focus of CBT on behavioral activation, which is thought to  
557 increase patients’ levels of activation and participation through encouraging them to re-  
558 engage in behaviors such as social interactions and physical activation, which patients with  
559 depression used to avoid or withdraw from (Dimaggio & Shahar, 2017). Here, we  
560 demonstrated in a large naturalistic sample that such a focus of depression treatment on  
561 physical disability indeed had positive effects on the level of perceived disability, which is  
562 arguably important for patients to participate in everyday life, work and social interactions.  
563 Second, reducing agoraphobia tendencies in depression had general positive effects on  
564 anxiety levels at posttreatment. While anxious depression is well known as a more difficult-  
565 to-treat population of depression subgroup in practice (Ionescu et al., 2013), the specific  
566 inpatient context including group therapy might have had positive effects on this  
567 relationship. In general, mindfulness-based interventions showed beneficial effects for both  
568 anxiety and depression (Hofmann et al., 2010) and therefore might also have addressed  
569 anxiety issues in our sample. Thus, also addressing anxiety issues of people with depression

570 might be a promising goal of treatment, for example, by using behavioral experiments to  
571 violate patients' dysfunctional expectations (Kube, Herzog, et al., 2019). Interestingly,  
572 anxiety and depression appear to be bidirectional risk factors for one another (Jacobson &  
573 Newman, 2017). Third, and perhaps most importantly, our results indicate an influence of  
574 suicidality on negative cognitions at posttreatment. Indeed, cognitive distortions, that is,  
575 hopelessness and helplessness (Ellis & Rutherford, 2008; Lester, 2012), for example, have  
576 been found to have a direct influence on suicidal ideation (Fazakas-Dehoog et al., 2017), and  
577 cognitive theory in general played also an important role in suicide theory (Rudd et al.,  
578 2001). Yet, our data also suggests the reciprocal influence, that is, that reducing suicidal  
579 thoughts might indeed positively influence negative cognition in a more general manner at  
580 posttreatment, highlighting once more the role of changing dysfunctional cognitions in  
581 depression treatment. For example, changing the attributional style has been shown to be  
582 significantly related to the resolution of suicidal ideation in an inpatient sample of children  
583 and adolescents (Wagner et al., 2000), which might also have been the underlying  
584 mechanism of change in our sample of depressed adults.

585         At follow-up, the current results provide some initial indications for the predictors of  
586 change in the long run. Physical and perceived disability, suicidality and concentration  
587 difficulties at discharge all predicted rumination and loss of interest at follow-up, while  
588 physical disability and somatic anxiety symptoms predicted pain at follow-up. This indicates  
589 at least that a better end-state functioning on the aforementioned factors at posttreatment  
590 predicted better outcomes at follow-up. For example, as an adjunct intervention, the  
591 Attention Training Technique (ATT) of the Metacognitive Therapy (MCT) rationale might  
592 therefore be applied to patients reporting such concentration difficulties (Wells & Fisher,  
593 2016). However, these results must be interpreted more cautiously because of the reduced  
594 data and weaker relations. Notably, a study investigating factors predicting long-term illness

## Predictors of change of inpatient psychotherapy for depression

27

595 course in a depressed inpatient sample yielded the following results: depression levels after  
596 discharge and unsatisfactory relationships within the support system predicted outcome in  
597 general; early onset, longer episode duration prior to study entry, and state orientation  
598 predicted time spent in episodes; number of previous hospitalizations and lack of  
599 psychotherapy after discharge predicted time to recurrence; and previous hospitalizations,  
600 duration of index episode, and state orientation predicted long-term psychosocial functioning  
601 (Kuehner & Huffziger, 2013). Interestingly, a current study using a large sample investigates  
602 long-term outcomes in patients with severe depression after in-hospital treatment which  
603 might yield new findings (Kirchberger et al., 2019).

604         Therefore, according to our results, the main factors contributing to persistent  
605 psychopathology in depression are: suicidality; agoraphobia; life dissatisfaction; physical  
606 disability; and pain. Importantly, our results indicate that suicidality remained fairly stable  
607 over the course of the treatment and was not significantly reduced through psychotherapy.  
608 These findings emphasize once more the need to adequately target suicidal behavior in  
609 people with depression (and other mental health problems, as well), as also claimed in Part 9  
610 of the guidance paper of the Lancet Psychiatry Commission (Holmes et al., 2018).  
611 Moreover, our results indicate that pre-treatment factors, such as sociodemographic and  
612 clinical variables, significantly influenced symptomatology at baseline, but their direct  
613 influence weakened over the course of treatment and beyond. For example, a study with  
614 severely depressed female psychiatric inpatients also showed that psychosocial factors may  
615 have less relevance to short term outcome (Andrew et al., 1993). At least for baseline  
616 symptomatology, older age, male gender and higher education were identified as  
617 “protective” factors (that is, factors that were associated with lower symptom severity); by  
618 contrast, chronicity of depression, treatment resistance and work-related problems were  
619 identified as “risk” factors (that is, factors that were associated with higher symptom

620 severity). Our results correspond to the findings of other studies: Research suggests that  
621 childhood maltreatment, post-treatment residual symptoms, and a history of recurrence were  
622 strong prognostic indicators of risk for recurrence and relapse of depression (Buckman et al.,  
623 2018). In line, lack of full remission after a previous episode, high rate of depressive  
624 recurrences, early onset of first depressive episode, nonresponse to the first antidepressant  
625 received, current suicidal risk and comorbid anxiety disorder were identified as risk factors  
626 particularly for treatment-resistant depression (Bennabi et al., 2015). Specifically, childhood  
627 maltreatment were shown to be a risk factor for severe, early-onset, treatment-resistant  
628 depression with a chronic course (Nelson et al., 2017). Furthermore, a younger age at onset,  
629 longer duration of depressive episode, and family history of mood disorders were identified  
630 as risk factors for chronic depression (Hölzel et al., 2011). Notably, job strain also seem to  
631 be a risk factor of depression among employees (Madsen et al., 2017).

632 Finally, medication at discharge was not associated with additional symptom change at a 6-  
633 month follow-up. The most commonly used long-term treatment is maintenance  
634 antidepressants despite the uncertainty of its long term efficacy (Hengartner, 2020; Pies,  
635 2012; Uher & Pavlova, 2016). In line, a meta-analysis found no significant difference in  
636 relapse after the acute phase CBT versus continuation of pharmacotherapy after remission  
637 (Cuijpers et al., 2013), i.e. discontinued CBT might be as effective as continued  
638 antidepressant treatment and more effective than discontinued antidepressants. While our  
639 results showed neither beneficial nor disadvantageous effects of medication on long term  
640 outcome, research discusses whether medication might worsen the long term course of the  
641 illness (Fava, 2003). Indeed, one study showed poorer long term outcomes among  
642 depression patients treated with medication (Vittengl, 2017). Yet, the role of medication in  
643 the long-term treatment of depression remains an important research topic.

#### 644 **Strengths and Limitations**

645 Major strengths of this work are the focus on practice-oriented research by generating  
646 benchmarks and practice-based evidence, the analysis of a large and nationwide  
647 representative dataset of inpatients with depression and the examination of inpatient  
648 treatment according to national best-practice guidelines to determine the implementation of  
649 evidence-based practice (Barkham & Mellor-Clark, 2003; Castonguay et al., 2013; Lutz et  
650 al., 2019). In addition, we adhered to recent advances in statistical procedures (Iniesta et al.,  
651 2016) and proposed a probabilistic model of predictors of change in depression treatment,  
652 which takes into account the heterogeneity and complexity of depression inpatients by  
653 incorporating multiple data sources of candidate predictors (sociodemographic, clinical and  
654 psychometric data) (Iniesta et al., 2016). This procedure increased the external validity of  
655 our results, as they are therefore compatible with the heterogeneity of depression and its  
656 treatment. Relatedly, the proposed model may help the practitioner to identify those  
657 depression inpatients at risk for non-responding or relapsing by unfolding the “blackbox” of  
658 depression treatment. Lastly, this study may be regarded as another proof-of-concept-study  
659 of using Bayesian statistics in psychotherapy research (Høifødt et al., 2015).

660 A major limitation of effect size estimation is that we exclusively used self-report  
661 data, that is, the BDI-II and PHQ-9 as the primary treatment outcome measure; accordingly,  
662 we could not include a clinician-rated measure of depressive symptoms, such the HRSD, in  
663 our analysis. Due to the naturalistic design of our practice-based study (Barkham et al.,  
664 2008), a further inherent limitation is that there is no control group, nor a randomization to  
665 different treatments, leading to a lack of internal validity. Therefore, the predictors of change  
666 as suggested by our analyses should be replicated under more controlled circumstances, that  
667 is, in prospective studies (e.g. in RCTs), and also in an outpatient setting to enhance  
668 generalizability. Furthermore, since this is a single-treatment study with routine data, we  
669 were able to identify prognostic variables in terms of general predictors, but were not able to



670 identify prescriptive variables nor to perform moderator and mediator analyses (Kraemer et  
671 al., 2002). In addition, note that with the analysis of the cross-lagged panel data available, no  
672 causal conclusions can be drawn directly from our results – only Granger causal conclusions.  
673 By taking multiple repeated measures and incorporating them in the cross-lagged model, two  
674 requisites for establishing causal relations are fulfilled, namely, establishing an association  
675 between the variables studied and taking into account the time order of the processes (e.g.  
676 the cause has to occur before the result). Such an association between variables, in which a  
677 variable  $X$  predicts future values of another variable  $Y$ , is referred to as Granger-causal:  
678 Variable  $X$  “Grangercauses” variable  $Y$  (Granger, 1969; Schuurman et al., 2016). Of note,  
679 unmeasured constructs might be related to the factors measured in this study and act as  
680 confounders and true causes of change in other factors. Other variables that have been found  
681 to contribute to treatment outcome but were not measured routinely in the clinics include, for  
682 example, depression-specific expectations (Kube et al., 2018) and childhood maltreatment  
683 (R.-Mercier et al., 2018). Furthermore, the results of our Bayesian analyses rely solely on  
684 clinician- and patient-report data. Notably, a recent study has found that models with  
685 multiple data sources (i.e. neuroimaging and genetic data) are more accurate and adequate  
686 than models with single, lower-dimension data types (i.e. exclusively phenomenological  
687 data) (Lee et al., 2018). Relatedly, some authors claim that it is unlikely that only clinical  
688 markers – even in all clinical assessment domains – can guide treatment selection or  
689 adaptation, because of biological heterogeneity, underlying pathology and its resulting  
690 phenomenologies, that is, the variability of symptom presentation of depression (Trivedi,  
691 2013).

#### 692 **Future research directions**

693       Based on our approach to identify predictors of change, some implications for future  
694 research on and treatment of depression can be derived. Our vision is to tailor psychological

695 treatments to the specific needs and problems of patients at the beginning of treatment, and  
696 to address subsequently important processes that are known to contribute to short- and long-  
697 term treatment outcomes. To personalize treatments and achieve improved outcomes, it  
698 seems promising: (1) to optimize the selection of a treatment for individual patients at the  
699 beginning of treatment (moderator analyses), that is, rapidly identifying the most appropriate  
700 treatment; and (2) to identify indicators of eventual outcomes early in treatment (mediator  
701 analyses), which facilitates adaptations to more powerful interventions (e.g. Lutz et al., 2017).  
702 To approach the first aim, future research may examine multiple treatment approaches to  
703 determine differential predictors (i.e. moderators or prescriptive variables) of treatment  
704 response in order to guide treatment selection (Cohen & DeRubeis, 2018; van Bronswijk et  
705 al., 2019), and to build clinical decision trees, which was recently applied to eating disorders  
706 (Hannöver et al., 2002). Recently, a simulation study within an artificial intelligence (AI)  
707 framework has been shown to be successful for clinical decision-making even in complex  
708 and uncertain medical settings (Bennett & Hauser, 2013). Additionally, in psychotherapy  
709 research, AI and machine learning could be very useful to guide decision-making in clinical  
710 practice (Dwyer et al., 2018).

711 For the second aim, based on our results, we believe that studies are needed that  
712 capture the complexity of the phenomenology of depressive disorders and symptom change  
713 through psychological treatments, as proposed by the Lancet Psychiatry Commission  
714 (Holmes et al., 2018). To address this complexity of mental disorders and their change  
715 through psychotherapy, we need to take into account different relevant aspects. There are  
716 clinical and mostly phenomenological characteristics with a different underlying pathology  
717 (e.g. chronicity, treatment-resistance, severity and comorbidities), patient factors (e.g.  
718 personality, attachment, motivation and other sociodemographic variables), likely therapy  
719 process factors (e.g. sudden gains) and treatment alliance (e.g. adherence, misdiagnosis and



745

**References**

- 746 Addis, M. E., & Jacobson, N. S. (1996). Reasons for depression and the process and  
747 outcome of cognitive-behavioral psychotherapies. *Journal of Consulting and Clinical*  
748 *Psychology, 64*(6), 1417–1424. <https://doi.org/10.1037//0022-006x.64.6.1417>
- 749 Andrew, B., Hawton, K., Fagg, J., & Westbrook, D. (1993). Do psychosocial factors  
750 influence outcome in severely depressed female psychiatric in-patients? *British Journal*  
751 *of Psychiatry, 163*(DEC.), 747–754. <https://doi.org/10.1192/bjp.163.6.747>
- 752 Barkham, M., Hardy, G. E., & Mellor-Clark, J. (2010). *Developing and delivering practice-*  
753 *based evidence : a guide for the psychological therapies*. Wiley-Blackwell.
- 754 Barkham, M., & Mellor-Clark, J. (2003). Bridging evidence-based practice and practice-  
755 based evidence: developing a rigorous and relevant knowledge for the psychological  
756 therapies. *Clinical Psychology & Psychotherapy, 10*(6), 319–327.  
757 <https://doi.org/10.1002/cpp.379>
- 758 Barkham, M., Stiles, W. B., Connell, J., Twigg, E., Leach, C., Lucock, M., Mellor-Clark, J.,  
759 Bower, P., King, M., Shapiro, D. A., Hardy, G. E., Greenberg, L., & Angus, L. (2008).  
760 Effects of psychological therapies in randomized trials and practice-based studies.  
761 *British Journal of Clinical Psychology, 47*(4), 397–415.  
762 <https://doi.org/10.1348/014466508X311713>
- 763 Beck, A. T. (1979). *Cognitive therapy of depression*. Guilford Press.
- 764 Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory-II*. Harcourt  
765 Assessment Inc.
- 766 Bennabi, D., Aouizerate, B., El-Hage, W., Doumy, O., Moliere, F., Courtet, P., Nieto, I.,  
767 Bellivier, F., Bubrovsky, M., Vaiva, G., Holzmann, J., Bougerol, T., Richieri, R.,  
768 Lancon, C., Camus, V., Saba, G., Haesbaert, F., d'Amato, T., Charpeaud, T., ... Haffen,  
769 E. (2015). Risk factors for treatment resistance in unipolar depression: A systematic  
770 review. *Journal of Affective Disorders, 171*, 137–141.  
771 <https://doi.org/10.1016/j.jad.2014.09.020>
- 772 Bennett, C. C., & Hauser, K. (2013). Artificial intelligence framework for simulating clinical  
773 decision-making: A Markov decision process approach. *Artificial Intelligence in*  
774 *Medicine, 57*(1), 9–19. <https://doi.org/10.1016/j.artmed.2012.12.003>

- 775 Bergfeld, I. O., Mantione, M., Figeet, M., Schuurman, P. R., Lok, A., & Denys, D. (2018).  
 776 Treatment-resistant depression and suicidality. *Journal of Affective Disorders*, *235*,  
 777 362–367. <https://doi.org/10.1016/j.jad.2018.04.016>
- 778 Berking, M. (2015). Training emotionaler Kompetenzen (Affect Regulation Training). In  
 779 *Training emotionaler Kompetenzen*. Springer Berlin Heidelberg.  
 780 <https://doi.org/10.1007/978-3-642-54017-2>
- 781 Bird, T., Tarsia, M., & Schwannauer, M. (2018). Interpersonal styles in major and chronic  
 782 depression: A systematic review and meta-analysis. *Journal of Affective Disorders*, *239*,  
 783 93–101. <https://doi.org/10.1016/j.jad.2018.05.057>
- 784 Blom, M. B. J., Spinhoven, P., Hoffman, T., Jonker, K., Hoencamp, E., Haffmans, P. M. J.,  
 785 & van Dyck, R. (2007). Severity and duration of depression, not personality factors,  
 786 predict short term outcome in the treatment of major depression. *Journal of Affective*  
 787 *Disorders*, *104*(1–3), 119–126. <https://doi.org/10.1016/J.JAD.2007.03.010>
- 788 Boschloo, L., Bekhuis, E., Weitz, E. S., Reijnders, M., DeRubeis, R. J., Dimidjian, S.,  
 789 Dunner, D. L., Dunlop, B. W., Hegerl, U., Hollon, S. D., Jarrett, R. B., Kennedy, S. H.,  
 790 Miranda, J., Mohr, D. C., Simons, A. D., Parker, G., Petrak, F., Herpertz, S., Quilty, L.  
 791 C., ... Cuijpers, P. (2019). The symptom-specific efficacy of antidepressant medication  
 792 vs. cognitive behavioral therapy in the treatment of depression: results from an  
 793 individual patient data meta-analysis. *World Psychiatry*, *18*(2), 183–191.  
 794 <https://doi.org/10.1002/wps.20630>
- 795 Buckman, J. E. J., Underwood, A., Clarke, K., Saunders, R., Hollon, S. D., Fearon, P., &  
 796 Pilling, S. (2018). Risk factors for relapse and recurrence of depression in adults and  
 797 how they operate: A four-phase systematic review and meta-synthesis. *Clinical*  
 798 *Psychology Review*, *64*, 13–38. <https://doi.org/10.1016/j.cpr.2018.07.005>
- 799 Carter, J. D., Luty, S. E., McKenzie, J. M., Mulder, R. T., Frampton, C. M., & Joyce, P. R.  
 800 (2011). Patient predictors of response to cognitive behaviour therapy and interpersonal  
 801 psychotherapy in a randomised clinical trial for depression. *Journal of Affective*  
 802 *Disorders*, *128*(3), 252–261. <https://doi.org/10.1016/j.jad.2010.07.002>
- 803 Castonguay, Louis G., & Muran, J. C. (2015). Fostering collaboration between researchers  
 804 and clinicians through building practice-oriented research: An introduction.

- 805         *Psychotherapy Research*, 25(1), 1–5. <https://doi.org/10.1080/10503307.2014.966348>
- 806     Castonguay, Louis Georges., & Muran, J. C. (2016). *Practice-oriented research in*  
807         *psychotherapy : building partnerships between clinicians and researchers*. Routledge.
- 808     Castonguay, Louis Georges, Barkham, M., Lutz, W., & McAleavey, A. A. (2013). Practice-  
809         oriented research: Approaches and application. In Michael J. Lambert (Ed.), *Bergin &*  
810         *Garfield's handbook of psychotherapy and behavior change* (6th editio). Wiley.
- 811     Cepeda, M. S., Reps, J., & Ryan, P. (2018). Finding factors that predict treatment-resistant  
812         depression: Results of a cohort study. *Depression and Anxiety*, 35(7), 668–673.  
813         <https://doi.org/10.1002/da.22774>
- 814     Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H.,  
815         Cannon, T. D., Krystal, J. H., & Corlett, P. R. (2016). Cross-trial prediction of  
816         treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry*,  
817         3(3), 243–250. [https://doi.org/10.1016/S2215-0366\(15\)00471-X](https://doi.org/10.1016/S2215-0366(15)00471-X)
- 818     Cipriani, A., & Geddes, J. (2016). Predicting treatment outcome in depression: so far, so  
819         good. *The Lancet Psychiatry*, 3(3), 192–194. [https://doi.org/10.1016/S2215-](https://doi.org/10.1016/S2215-0366(15)00542-8)  
820         0366(15)00542-8
- 821     Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment Selection in Depression. *Annual Review*  
822         *of Clinical Psychology*, 14, 15.1-15.28. <https://doi.org/10.1146/annurev-clinpsy-050817>
- 823     Constantino, M. J., Manber, R., DeGeorge, J., McBride, C., Ravitz, P., Zuroff, D. C., Klein,  
824         D. N., Markowitz, J. C., Rothbaum, B. O., Thase, M. E., & Arnow, B. A. (2008).  
825         Interpersonal styles of chronically depressed outpatients: Profiles and therapeutic  
826         change. *Psychotherapy: Theory, Research, Practice, Training*, 45(4), 491–506.  
827         <https://doi.org/10.1037/a0014335>
- 828     Conti, G., Frühwirth-Schnatter, S., Heckman, J. J., & Piatek, R. (2014). Bayesian  
829         exploratory factor analysis. *Journal of Econometrics*, 183(1), 31–57.  
830         <https://doi.org/10.1016/j.jeconom.2014.06.008>
- 831     Craske, M. G. (2019). Refining our research practices in clinical science: Challenges and  
832         steps towards solutions. *Behaviour Research and Therapy*, 116, 90–93.  
833         <https://doi.org/10.1016/j.brat.2019.03.006>

- 834 Cuijpers, P., Karyotaki, E., Reijnders, M., & Ebert, D. D. (2019). Was Eysenck right after  
 835 all? A reassessment of the effects of psychotherapy for adult depression. *Epidemiology  
 836 and Psychiatric Sciences*, 28(1), 21–30. <https://doi.org/10.1017/S2045796018000057>
- 837 Cuijpers, P., Hollon, S. D., Van Straten, A., Bockting, C., Berking, M., & Andersson, G.  
 838 (2013). Does cognitive behaviour therapy have an enduring effect that is superior to  
 839 keeping patients on continuation pharmacotherapy? A meta-analysis. *BMJ Open*, 3(4).  
 840 <https://doi.org/10.1136/bmjopen-2012-002542>
- 841 Cuijpers, P., Karyotaki, E., Reijnders, M., & Ebert, D. D. (2019). Is psychotherapy effective?  
 842 Pretending everything is fine will not help the field forward. *Epidemiology and  
 843 Psychiatric Sciences*, 28(03), 356–357. <https://doi.org/10.1017/S204579601800080X>
- 844 Cuijpers, P., Hofmann, S. G., & Andersson, G. (2010). Self-reported versus clinician-rated  
 845 symptoms of depression as outcome measures in psychotherapy research on depression:  
 846 A meta-analysis. *Clinical Psychology Review*, 30(6), 768–778.  
 847 <https://doi.org/10.1016/j.cpr.2010.06.001>
- 848 Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C. H., Cipriani, Andrea, & Furukawa, T. A.  
 849 (2020). A network meta-analysis of the effects of psychotherapies, pharmacotherapies  
 850 and their combination in the treatment of adult depression. *World Psychiatry*, 19(1),  
 851 92–107. <https://doi.org/10.1002/wps.20701>
- 852 Cuijpers, P., van Straten, A., Schuurmans, J., van Oppen, P., Hollon, S. D., & Andersson, G.  
 853 (2010). Psychotherapy for chronic major depression and dysthymia: A meta-analysis.  
 854 *Clinical Psychology Review*, 30(1), 51–62. <https://doi.org/10.1016/j.cpr.2009.09.003>
- 855 Derogatis, L. R., & Melisaratos, N. (1983). The brief symptom inventory: an introductory  
 856 report. *Psychological Medicine*, 13(03), 595–605.
- 857 DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-  
 858 Luaces, L. (2014). The Personalized Advantage Index. *PLOS ONE*, 9(1), e83875.  
 859 <https://doi.org/10.1371/journal.pone.0083875> T4 - Translating Research on Prediction  
 860 into Individualized Treatment Recommendations. A Demonstration U6 -  
 861 [http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0083875&type=pri](http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0083875&type=printable)  
 862 [ntable M4 - Citavi](http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0083875&type=printable)
- 863 DGPPN, BÄK, KBV, & AWMF. (2017). *S3-Leitlinien/Nationale VersorgungsLeitlinie:*

- 864 *Unipolare Depression* (2. Auflage). Springer.
- 865 Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life  
866 Scale. *Journal of Personality Assessment*, *49*(1), 71–75.  
867 [https://doi.org/10.1207/s15327752jpa4901\\_13](https://doi.org/10.1207/s15327752jpa4901_13)
- 868 Dimaggio, G., & Shahar, G. (2017). Behavioral activation as a common mechanism of  
869 change across different orientations and disorders. *Psychotherapy*, *54*(3), 221–224.  
870 <https://doi.org/10.1037/pst0000117>
- 871 Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for  
872 Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, *14*(1), 91–  
873 118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>
- 874 Ellis, T. E., & Rutherford, B. (2008). Cognition and suicide: Two decades of progress.  
875 *International Journal of Cognitive Therapy*, *1*(1), 47–68.  
876 <https://doi.org/10.1521/ijct.2008.1.1.47>
- 877 Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale.  
878 *Archives of General Psychiatry*, *33*(6), 766.  
879 <https://doi.org/10.1001/archpsyc.1976.01770060086012>
- 880 Enns, M. W., Cox, B. J., & Pidlubny, S. R. (2002). Group Cognitive Behaviour Therapy for  
881 Residual Depression: Effectiveness and Predictors of Response. *Cognitive Behaviour*  
882 *Therapy*, *31*(1), 31–40. <https://doi.org/10.1080/16506070252823643>
- 883 Fava, G. A. (2003). Can long-term treatment with antidepressant drugs worsen the course of  
884 depression? *Journal of Clinical Psychiatry*, *64*(2), 123–133.  
885 <https://doi.org/10.4088/JCP.v64n0204>
- 886 Fava, G. A., & Guidi, J. (2020). *The Deceptive Manifestations of Treatment Resistance in*  
887 *Depression : A New Look at the Problem*. <https://doi.org/10.1159/000507227>
- 888 Fazakas-Dehoog, L. L., Rnic, K., & Dozois, D. J. A. (2017). A cognitive distortions and  
889 deficits model of suicide ideation. *Europe's Journal of Psychology*, *13*(2), 178–193.  
890 <https://doi.org/10.5964/ejop.v13i2.1238>
- 891 Frank, E., Cassano, G. B., Rucci, P., Thompson, W. K., Kraemer, H. C., Fagiolini, A.,  
892 Maggi, L., Kupfer, D. J., Shear, M. K., Houck, P. R., Calugi, S., Grochocinski, V. J.,



- 893 Scocco, P., Bутtenfield, J., & Forgione, R. N. (2011). Predictors and moderators of time  
894 to remission of major depression with interpersonal psychotherapy and SSRI  
895 pharmacotherapy. *Psychological Medicine*, *41*(1), 151–162.  
896 <https://doi.org/10.1017/S0033291710000553>
- 897 Franke, G. H. (1997). Erste Studien zur Güte des Brief Symptom Inventory (BSI). *Zeitschrift*  
898 *Für Medizinische Psychologie*, *3*(July), 159–166.
- 899 Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational  
900 psychiatry: The brain as a phantastic organ. *The Lancet Psychiatry*, *1*(2), 148–158.  
901 [https://doi.org/10.1016/S2215-0366\(14\)70275-5](https://doi.org/10.1016/S2215-0366(14)70275-5)
- 902 Gaston, L., Marmar, C. R., Gallagher, D., & Thompson, L. W. (1989). Impact of confirming  
903 patient expectations of change processes in behavioral, cognitive, and brief dynamic  
904 psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, *26*(3), 296–302.  
905 <https://doi.org/10.1037/h0085439>
- 906 Goldfried, M. R. (2019). Obtaining consensus in psychotherapy: What holds us back?  
907 *American Psychologist*, *74*(4), 484–496. <https://doi.org/10.1037/amp0000365>
- 908 Gräfe, K., Zipfel, S., Herzog, W., & Löwe, B. (2004). Screening psychischer störungen mit  
909 dem “Gesundheitsfragebogen für Patienten (PHQ-D)”. Ergebnisse der Deutschen  
910 validierungsstudie. *Diagnostica*, *50*(4), 171–181. <https://doi.org/10.1026/0012-1924.50.4.171>
- 912 Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-  
913 spectral Methods. *Econometrica*, *37*(3), 424. <https://doi.org/10.2307/1912791>
- 914 Grant, N., Hotopf, M., Breen, G., Cleare, A., Grey, N., Hepgul, N., King, S., Moran, P.,  
915 Pariente, C. M., Wingrove, J., Young, A. H., & Tylee, A. (2014). Predicting outcome  
916 following psychological therapy in IAPT (PROMPT): a naturalistic project protocol.  
917 *BMC Psychiatry*, *14*(1), 170. <https://doi.org/10.1186/1471-244X-14-170>
- 918 Groves, S. J., Douglas, K. M., & Porter, R. J. (2018). A Systematic Review of Cognitive  
919 Predictors of Treatment Outcome in Major Depression. *Frontiers in Psychiatry*, *9*.  
920 <https://doi.org/10.3389/fpsy.2018.00382>
- 921 Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates:

- 922 Lessons from year one of IAPT. *Behaviour Research and Therapy*, 51(9), 597–606.  
 923 <https://doi.org/10.1016/j.brat.2013.06.004>
- 924 Hannöver, W., Richard, M., Hansen, N. B., Martinovich, Z., & Kordy, H. (2002). A  
 925 classification tree model for decision-making in clinical practice: An application based  
 926 on the data of the german multicenter study on eating disorders, project tr-eat.  
 927 *Psychotherapy Research*, 12(4), 445–461. <https://doi.org/10.1080/713664470>
- 928 Helwig, N. E. (2015). *Ica: independent component analysis*. R package version 1.0-1.
- 929 Hengartner, M. P. (2020). How effective are antidepressants for depression over the long  
 930 term? A critical review of relapse prevention trials and the issue of withdrawal  
 931 confounding. *Therapeutic Advances in Psychopharmacology*, 10, 2045125320921694.  
 932 <https://doi.org/10.1177/2045125320921694>
- 933 Hinsch, R., & Pflingsten, U. (2015). *Gruppentraining sozialer Kompetenzen GSK :  
 934 Grundlagen, Durchführung, Anwendungsbeispiele TT - Group training of social  
 935 competence - Foundations, practice, applications*. Beltz.  
 936 <http://search.ebscohost.com/login.aspx?direct=true&db=pdx&AN=0294917&site=ehost-live>  
 937 t-live
- 938 Hofmann, S. G., Sawyer, A. T., Witt, A. A., & Oh, D. (2010). The effect of mindfulness-  
 939 based therapy on anxiety and depression: A meta-analytic review. *Journal of  
 940 Consulting and Clinical Psychology*, 78(2), 169–183. <https://doi.org/10.1037/a0018555>
- 941 Høifødt, R. S., Mittner, M., Lillevoll, K., Katla, S. K., Kolstrup, N., Eisemann, M., Friborg,  
 942 O., & Waterloo, K. (2015). Predictors of Response to Web-Based Cognitive Behavioral  
 943 Therapy With High-Intensity Face-to-Face Therapist Guidance for Depression: A  
 944 Bayesian Analysis. *Journal of Medical Internet Research*, 17(9), e197.  
 945 <https://doi.org/10.2196/jmir.4351>
- 946 Holmes, E. A., Ghaderi, A., Harmer, C. J., Ramchandani, P. G., Cuijpers, P., Morrison, A.  
 947 P., Roiser, J. P., Bockting, C. L. H., O'Connor, R. C., Shafran, R., Moulds, M. L., &  
 948 Craske, M. G. (2018). The Lancet Psychiatry Commission on psychological treatments  
 949 research in tomorrow's science. *The Lancet Psychiatry*, 5(3), 237–286.  
 950 [https://doi.org/10.1016/S2215-0366\(17\)30513-8](https://doi.org/10.1016/S2215-0366(17)30513-8)
- 951 Hölzel, L., Härter, M., Reese, C., & Kriston, L. (2011). Risk factors for chronic depression--

## Predictors of change of inpatient psychotherapy for depression

40

- 952 a systematic review. *Journal of Affective Disorders*, 129(1–3), 1–13.  
953 <https://doi.org/10.1016/j.jad.2010.03.025>
- 954 Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., Henigsberg, N.,  
955 Dernovsek, M. Z., Souery, D., Stahl, D., Dobson, R., Aitchison, K. J., Farmer, A.,  
956 Lewis, C. M., McGuffin, P., & Uher, R. (2016). Combining clinical variables to  
957 optimize prediction of antidepressant treatment outcomes. *Journal of Psychiatric*  
958 *Research*, 78, 94–102. <https://doi.org/10.1016/j.jpsychires.2016.03.016>
- 959 Ionescu, D. F., Niciu, M. J., Henter, I. D., Zarate, C. A., & Jr. (2013). Defining anxious  
960 depression: a review of the literature. *CNS Spectrums*, 18(5), 252–260.  
961 <https://doi.org/10.1017/S1092852913000114>
- 962 Jacobson, N. C., & Newman, M. G. (2017). Anxiety and depression as bidirectional risk  
963 factors for one another: A meta-analysis of longitudinal studies. *Psychological Bulletin*,  
964 143(11), 1155–1200. <https://doi.org/10.1037/bul0000111>
- 965 Jacobucci, R., & Grimm, K. J. (2018). Comparison of Frequentist and Bayesian  
966 Regularization in Structural Equation Modeling. *Structural Equation Modeling*, 25(4),  
967 639–649. <https://doi.org/10.1080/10705511.2017.1410822>
- 968 Jarrett, R. B., Eaves, G. G., Grannemann, B. D., & Rush, A. J. (1991). Clinical, cognitive,  
969 and demographic predictors of response to cognitive therapy for depression: A  
970 preliminary report. *Psychiatry Research*, 37(3), 245–260. [https://doi.org/10.1016/0165-](https://doi.org/10.1016/0165-1781(91)90061-S)  
971 [1781\(91\)90061-S](https://doi.org/10.1016/0165-1781(91)90061-S)
- 972 Johnston, K. M., Powell, L. C., Anderson, I. M., Szabo, S., & Cline, S. (2019). The burden  
973 of treatment-resistant depression: A systematic review of the economic and quality of  
974 life literature. *Journal of Affective Disorders*, 242, 195–210.  
975 <https://doi.org/10.1016/J.JAD.2018.06.045>
- 976 Kertz, S. J., Koran, J., Stevens, K. T., & Björgvinsson, T. (2015). Repetitive negative  
977 thinking predicts depression and anxiety symptom improvement during brief cognitive  
978 behavioral therapy. *Behaviour Research and Therapy*, 68, 54–63.  
979 <https://doi.org/10.1016/j.brat.2015.03.006>
- 980 Kilbourne, A. M., Williams, M., Bauer, M. S., & Arean, P. (2012). Implementation research:  
981 Reducing the research-to-practice gap in depression treatment. In *Depression Research*

## Predictors of change of inpatient psychotherapy for depression

41

- 982        *and Treatment* (Vol. 2012). <https://doi.org/10.1155/2012/476027>
- 983    Kirchberger, I., Maleckar, B., Meisinger, C., Linseisen, J., Schmauss, M., & Baumgärtner, J.  
984        (2019). Long-term outcomes in patients with severe depression after in-hospital  
985        treatment - Study protocol of the depression long-term Augsburg (DELTA) study. *BMJ*  
986        *Open*, 9(12), e032507. <https://doi.org/10.1136/bmjopen-2019-032507>
- 987    Knapstad, M., Nordgreen, T., & Smith, O. R. F. (2018). Prompt mental health care, the  
988        Norwegian version of IAPT: clinical outcomes and predictors of change in a  
989        multicenter cohort study. *BMC Psychiatry*, 18(1), 260. [https://doi.org/10.1186/s12888-](https://doi.org/10.1186/s12888-018-1838-0)  
990        018-1838-0
- 991    Köhler, S., Chrysanthou, S., Guhn, A., & Sterzer, P. (2018). Differences between chronic  
992        and nonchronic depression: Systematic review and implications for treatment.  
993        *Depression and Anxiety*. <https://doi.org/10.1002/da.22835>
- 994    Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and  
995        moderators of treatment effects in randomized clinical trials. *Archives of General*  
996        *Psychiatry*, 59(10), 877–883. <https://doi.org/10.1001/archpsyc.59.10.877>
- 997    Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General*  
998        *Internal Medicine*, 16(9), 606–613. [https://doi.org/10.1046/j.1525-](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)  
999        1497.2001.016009606.x
- 1000    Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Löwe, B. (2010). The Patient Health  
1001        Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: A systematic  
1002        review. *General Hospital Psychiatry*, 32(4), 345–359.  
1003        <https://doi.org/10.1016/j.genhosppsy.2010.03.006>
- 1004    Kube, T., Siebers, V. H. A., Herzog, P., Glombiewski, J. A., Doering, B. K., & Rief, W.  
1005        (2018). Integrating situation-specific dysfunctional expectations and dispositional  
1006        optimism into the cognitive model of depression – A path-analytic approach. *Journal of*  
1007        *Affective Disorders*, 229. <https://doi.org/10.1016/j.jad.2017.12.082>
- 1008    Kube, T., Herzog, P., Michalak, C. M., Glombiewski, J. A., Doering, B. K., & Rief, W.  
1009        (2019). Further Specifying the Cognitive Model of Depression: Situational  
1010        Expectations and Global Cognitions as Predictors of Depressive Symptoms. *Clinical*  
1011        *Psychology in Europe*, 1(4). <https://doi.org/10.32872/cpe.v1i4.33548>

## Predictors of change of inpatient psychotherapy for depression

42

- 1012 Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2019). Distorted  
 1013 cognitive processes in major depression – A predictive processing perspective.  
 1014 *Biological Psychiatry*. <https://doi.org/10.1016/j.biopsych.2019.07.017>
- 1015 Kuehner, C., & Huffziger, S. (2013). Factors predicting the long-term illness course in a  
 1016 cohort of depressed inpatients. *European Archives of Psychiatry and Clinical  
 1017 Neuroscience*, 263(5), 413–423. <https://doi.org/10.1007/s00406-012-0379-7>
- 1018 Lambert, M J, Hatch, D. R., Kingston, M. D., & Edwards, B. C. (1986). Zung, Beck, and  
 1019 Hamilton Rating Scales as measures of treatment outcome: a meta-analytic comparison.  
 1020 *Journal of Consulting and Clinical Psychology*, 54(1), 54–59.  
 1021 <https://doi.org/10.1037//0022-006x.54.1.54>
- 1022 Lambert, Michael J. (2017). Maximizing Psychotherapy Outcome beyond Evidence-Based  
 1023 Medicine. *Psychotherapy and Psychosomatics*, 86(2), 80–89.  
 1024 <https://doi.org/10.1159/000455170>
- 1025 Lee, Y., Ragguett, R. M., Mansur, R. B., Boutilier, J. J., Rosenblat, J. D., Trevizol, A.,  
 1026 Brietzke, E., Lin, K., Pan, Z., Subramaniapillai, M., Chan, T. C. Y., Fus, D., Park, C.,  
 1027 Musial, N., Zuckerman, H., Chen, V. C. H., Ho, R., Rong, C., & McIntyre, R. S.  
 1028 (2018). Applications of machine learning algorithms to predict therapeutic outcomes in  
 1029 depression: A meta-analysis and systematic review. *Journal of Affective Disorders*, 241,  
 1030 519–532. <https://doi.org/10.1016/j.jad.2018.08.073>
- 1031 Lemmens, L. H. J. M., Arntz, A., Peeters, F., Hollon, S. D., Roefs, A., & Huibers, M. J. H.  
 1032 (2015). Clinical effectiveness of cognitive therapy v. interpersonal psychotherapy for  
 1033 depression: results of a randomized controlled trial. *Psychological Medicine*, 45(10),  
 1034 2095–2110. <https://doi.org/10.1017/S0033291715000033>
- 1035 Lemmens, L. H.J.M., van Bronswijk, S. C., Peeters, F., Arntz, A., Hollon, S. D., & Huibers,  
 1036 M. J. H. (2019). Long-term outcomes of acute treatment with cognitive therapy v.  
 1037 interpersonal psychotherapy for adult depression: follow-up of a randomized controlled  
 1038 trial. *Psychological Medicine*, 49(3), 465–473.  
 1039 <https://doi.org/10.1017/S0033291718001083>
- 1040 Lepping, P., Whittington, R., Sambhi, R. S., Lane, S., Poole, R., Leucht, S., Cuijpers, P.,  
 1041 McCabe, R., & Waheed, W. (2017). Clinical relevance of findings in trials of CBT for

## Predictors of change of inpatient psychotherapy for depression

43

- 1042 depression. *European Psychiatry*, 45, 207–211.  
 1043 <https://doi.org/10.1016/j.eurpsy.2017.07.003>
- 1044 Lester, D. (2012). The Role of Irrational Thinking in Suicidal Behavior. *Comprehensive*  
 1045 *Psychology*, 1, 12.02.07.16.CP.1.8. <https://doi.org/10.2466/12.02.07.16.cp.1.8>
- 1046 Lutz, W., Arndt, A., Rubel, J., Berger, T., Schröder, J., Späth, C., Meyer, B., Greiner, W.,  
 1047 Gräfe, V., Hautzinger, M., Fuhr, K., Rose, M., Nolte, S., Löwe, B., Hohagen, F., Klein,  
 1048 J. P., & Moritz, S. (2017). Defining and Predicting Patterns of Early Response in a  
 1049 Web-Based Intervention for Depression. *Journal of Medical Internet Research*, 19(6),  
 1050 e206. <https://doi.org/10.2196/jmir.7367>
- 1051 Lutz, W., Neu, R., & Rubel, J. (2019). *Evaluation und Effekterfassung in der*  
 1052 *Psychotherapie. Standards der Psychotherapie*. Hogrefe.
- 1053 Madsen, I. E. H., Nyberg, S. T., Magnusson Hanson, L. L., Ferrie, J. E., Ahola, K.,  
 1054 Alfredsson, L., Batty, G. D., Bjorner, J. B., Borritz, M., Burr, H., Chastang, J. F., De  
 1055 Graaf, R., Dragano, N., Hamer, M., Jokela, M., Knutsson, A., Koskenvuo, M.,  
 1056 Koskinen, A., Leineweber, C., ... Kivimäki, M. (2017). Job strain as a risk factor for  
 1057 clinical depression: Systematic review and meta-analysis with additional individual  
 1058 participant data. In *Psychological Medicine* (Vol. 47, Issue 8, pp. 1342–1356).  
 1059 Cambridge University Press. <https://doi.org/10.1017/S003329171600355X>
- 1060 McCullough, J. P., Schramm, E., & Penberthy, J. K. (2015). CBASP as a distinctive  
 1061 treatment for persistent depressive disorder. In *Routledge*. Routledge.
- 1062 Merkle, E. C., & Rosseel, Y. (2015). *blavaan: Bayesian structural equation models via*  
 1063 *parameter expansion*. <https://doi.org/10.18637/jss.v085.i04>
- 1064 Merrill, K. A., Tolbert, V. E., & Wade, W. A. (2003). Effectiveness of cognitive therapy for  
 1065 depression in a community mental health center: a benchmarking study. *Journal of*  
 1066 *Consulting and Clinical Psychology*, 71(2), 404–409. [https://doi.org/10.1037/0022-](https://doi.org/10.1037/0022-006x.71.2.404)  
 1067 [006x.71.2.404](https://doi.org/10.1037/0022-006x.71.2.404)
- 1068 Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E. G., Brown, G. S. (Jeb.), & Kircher,  
 1069 J. C. (2008). Benchmarking the Effectiveness of Psychotherapy Treatment for Adult  
 1070 Depression in a Managed Care Environment: A Preliminary Study. *Journal of*  
 1071 *Consulting and Clinical Psychology*, 76(1), 116–124. <https://doi.org/10.1037/0022->

- 1072 006X.76.1.116
- 1073 Minami, T., Wampold, B. E., Serlin, R. C., Kircher, J. C., & Brown, G. S. (2007).  
 1074 Benchmarks for Psychotherapy Efficacy in Adult Major Depression. *Journal of*  
 1075 *Consulting and Clinical Psychology*, 75(2), 232–243. [https://doi.org/10.1037/0022-](https://doi.org/10.1037/0022-006X.75.2.232)  
 1076 006X.75.2.232
- 1077 Moses, E. B., & Barlow, D. H. (2006). A New Unified Treatment Approach for Emotional  
 1078 Disorders Based on Emotion Science. *Current Directions in Psychological Science*,  
 1079 15(3), 146–150. <https://doi.org/10.1111/j.0963-7214.2006.00425.x>
- 1080 Munder, T., Flückiger, C., Leichsenring, F., Abbass, A. A., Hilsenroth, M. J., Luyten, P.,  
 1081 Rabung, S., Steinert, C., & Wampold, B. E. (2019). Is psychotherapy effective? A re-  
 1082 analysis of treatments for depression. *Epidemiology and Psychiatric Sciences*, 28(3),  
 1083 268–274. <https://doi.org/10.1017/S2045796018000355>
- 1084 Munder, T., Flückiger, C., Leichsenring, F., Abbass, A. A., Hilsenroth, M. J., Luyten, P.,  
 1085 Rabung, S., Steinert, C., & Wampold, B. E. (2019). Let's distinguish relative and  
 1086 absolute efficacy to move psychotherapy research forward. *Zeitschrift Fur*  
 1087 *Psychosomatische Medizin Und Psychotherapie*, 65(2), 178–182.  
 1088 <https://doi.org/10.13109/zptm.2019.65.2.178>
- 1089 Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social  
 1090 Adjustment Scale: A simple measure of impairment in functioning. *British Journal of*  
 1091 *Psychiatry*, 180(MAY), 461–464. <https://doi.org/10.1192/bjp.180.5.461>
- 1092 Mynors-Wallis, L., & Gath, D. (1997). Predictors of treatment outcome for major depression  
 1093 in primary care. *Psychological Medicine*, 27(3), S0033291796004126.  
 1094 <https://doi.org/10.1017/S0033291796004126>
- 1095 National Institute for Health and Clinical Excellence (NICE). (2018). *Depression in adults:*  
 1096 *recognition and management*. British Psychological Society.
- 1097 Nelson, J., Klumpp, A., Doebler, P., & Ehring, T. (2017). Childhood maltreatment and  
 1098 characteristics of adult depression: Meta-analysis. In *British Journal of Psychiatry* (Vol.  
 1099 210, Issue 2, pp. 96–104). Royal College of Psychiatrists.  
 1100 <https://doi.org/10.1192/bjp.bp.115.180752>

Predictors of change of inpatient psychotherapy for depression

45

- 1101 Piatek, R. (2017). *BayesFM: Bayesian Inference for Factor Modeling*. R package version  
 1102 0.1.2. <https://cran.r-project.org/web/packages/BayesFM/index.html>
- 1103 Pies, R. (2012). Are antidepressants effective in the acute and long-term treatment of  
 1104 depression? Sic et Non. In *Innovations in Clinical Neuroscience* (Vol. 9, Issues 5–6, pp.  
 1105 31–40). Matrix Medical Communications.
- 1106 R.-Mercier, A., Masson, M., Bussi eres, E.-L., & Cellard, C. (2018). Common  
 1107 transdiagnostic cognitive deficits among people with psychiatric disorders exposed to  
 1108 childhood maltreatment: A meta-analysis. *Cognitive Neuropsychiatry*, 23(3), 180–197.  
 1109 <https://doi.org/10.1080/13546805.2018.1461617>
- 1110 R Core Team. (2018). R: A language and environment for statistical computing. *R*  
 1111 *Foundation for Statistical Computing, Vienna*(Austria). <https://www.r-project.org/>
- 1112 Ravitz, P., McBride, C., & Maunder, R. (2011). Failures in interpersonal psychotherapy  
 1113 (IPT): factors related to treatment resistances. *Journal of Clinical Psychology*, 67(11),  
 1114 1129–1139. <https://doi.org/10.1002/jclp.20850>
- 1115 Reutfors, J., Andersson, T. M.-L., Brenner, P., Brandt, L., DiBernardo, A., Li, G., H agg, D.,  
 1116 Wing ard, L., & Bod en, R. (2018). Mortality in treatment-resistant unipolar depression:  
 1117 A register-based cohort study in Sweden. *Journal of Affective Disorders*, 238, 674–679.  
 1118 <https://doi.org/10.1016/j.jad.2018.06.030>
- 1119 Revelle, W. R. (2017). *psych: Procedures for Personality and Psychological Research*.  
 1120 [https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-](https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research)  
 1121 [personality-and-psychological-research](https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research)
- 1122 Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2019). The network  
 1123 approach to psychopathology: a review of the literature 2008–2018 and an agenda for  
 1124 future research. *Psychological Medicine*, 1–14.  
 1125 <https://doi.org/10.1017/S0033291719003404>
- 1126 Rosseel, Y. (2012). **lavaan** : An R Package for Structural Equation Modeling. *Journal of*  
 1127 *Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- 1128 Rudd, M. D., Joiner, T., & Rajab, M. H. (2001). Treating suicidal behavior: An effective,  
 1129 time-limited approach. In *Treating suicidal behavior: An effective, time-limited*



- 1130 *approach*. Guilford Press.
- 1131 Scheffl, C., Guhn, A., Brakemeier, E.-L., Sterzer, P., & Köhler, S. (2019). Efficacy of  
 1132 inpatient psychotherapy for major depressive disorder: a meta-analysis of controlled  
 1133 trials. *Acta Psychiatrica Scandinavica*. <https://doi.org/10.1111/acps.12995>
- 1134 Schramm, E., Kriston, L., Elsaesser, M., Fangmeier, T., Meister, R., Bausch, P., Zobel, I.,  
 1135 Bailer, J., Wambach, K., Backenstrass, M., Klein, J. P., Schoepf, D., Schnell, K., Gumz,  
 1136 A., Löwe, B., Walter, H., Wolf, M., Domschke, K., Berger, M., ... Härter, M. (2019).  
 1137 Two-Year Follow-Up after Treatment with the Cognitive Behavioral Analysis System  
 1138 of Psychotherapy versus Supportive Psychotherapy for Early-Onset Chronic  
 1139 Depression. *Psychotherapy and Psychosomatics*, *88*(3), 154–164.  
 1140 <https://doi.org/10.1159/000500189>
- 1141 Schramm, E., Kriston, L., Zobel, I., Bailer, J., Wambach, K., Backenstrass, M., Klein, J. P.,  
 1142 Schoepf, D., Schnell, K., Gumz, A., Bausch, P., Fangmeier, T., Meister, R., Berger, M.,  
 1143 Hautzinger, M., & Härter, M. (2017). Effect of Disorder-Specific vs Nonspecific  
 1144 Psychotherapy for Chronic Depression. *JAMA Psychiatry*, *74*(3), 233.
- 1145 Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M., & Hamaker, E. L. (2016). How to  
 1146 compare cross-lagged associations in a multilevel autoregressive model. *Psychological*  
 1147 *Methods*, *21*(2), 206–221. <https://doi.org/10.1037/met0000062>
- 1148 Simon, G. E., & Perlis, R. H. (2010). Personalized medicine for depression: can we match  
 1149 patients with treatments? *The American Journal of Psychiatry*, *167*(12), 1445–1455.  
 1150 <https://doi.org/10.1176/appi.ajp.2010.09111680>
- 1151 Stekhoven, D. J. (2013). *missForest: Nonparametric Missing Value Imputation using*  
 1152 *Random Forest*. R package version 1.4.
- 1153 Stekhoven, D. J., & Bühlmann, P. (2012). Missforest-Non-parametric missing value  
 1154 imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118.  
 1155 <https://doi.org/10.1093/bioinformatics/btr597>
- 1156 Stiles-Shields, C., Corden, M. E., Kwasny, M. J., Schueller, S. M., & Mohr, D. C. (2015).  
 1157 Predictors of outcome for telephone and face-to-face administered cognitive behavioral  
 1158 therapy for depression. *Psychological Medicine*, *45*(15), 3205–3215.  
 1159 <https://doi.org/10.1017/S0033291715001208>

## Predictors of change of inpatient psychotherapy for depression

47

- 1160 Trivedi, M. H. (2013). Modeling Predictors, Moderators and Mediators of Treatment  
 1161 Outcome and Resistance in Depression. *Biological Psychiatry*, *74*(1), 2–4.  
 1162 <https://doi.org/10.1016/J.BIOPSYCH.2013.05.009>
- 1163 Uher, R., & Pavlova, B. (2016). Long-term effects of depression treatment. In *The Lancet*  
 1164 *Psychiatry* (Vol. 3, Issue 2, pp. 95–96). Elsevier Ltd. [https://doi.org/10.1016/S2215-](https://doi.org/10.1016/S2215-0366(15)00578-7)  
 1165 [0366\(15\)00578-7](https://doi.org/10.1016/S2215-0366(15)00578-7)
- 1166 van Bronswijk, S. C., DeRubeis, R. J., Lemmens, L. H. J. M., Peeters, F. P. M. L., Keefe, J.  
 1167 R., Cohen, Z. D., & Huibers, M. J. H. (2019). Precision medicine for long-term  
 1168 depression outcomes using the Personalized Advantage Index approach: cognitive  
 1169 therapy or interpersonal psychotherapy? *Psychological Medicine*, *November*, 1–11.  
 1170 <https://doi.org/10.1017/s0033291719003192>
- 1171 van Bronswijk, S., Moopen, N., Beijers, L., Ruhe, H. G., & Peeters, F. (2019). Effectiveness  
 1172 of psychotherapy for treatment-resistant depression: a meta-analysis and meta-  
 1173 regression. *Psychological Medicine*, *49*(3), 366–379.  
 1174 <https://doi.org/10.1017/S003329171800199X>
- 1175 Vittengl, J. R. (2017). Poorer Long-Term Outcomes among Persons with Major Depressive  
 1176 Disorder Treated with Medication. *Psychotherapy and Psychosomatics*, *86*(5), 302–  
 1177 304. <https://doi.org/10.1159/000479162>
- 1178 Vos, T., Barber, R. M., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., Charlson, F.,  
 1179 Davis, A., Degenhardt, L., Dicker, D., Duan, L., Erskine, H., Feigin, V. L., Ferrari, A.  
 1180 J., Fitzmaurice, C., Fleming, T., Graetz, N., Guinovart, C., Haagsma, J., ... Murray, C.  
 1181 J. (2015). Global, regional, and national incidence, prevalence, and years lived with  
 1182 disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013:  
 1183 a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*,  
 1184 *386*(9995), 743–800. [https://doi.org/10.1016/S0140-6736\(15\)60692-4](https://doi.org/10.1016/S0140-6736(15)60692-4)
- 1185 Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R.,  
 1186 Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D.  
 1187 (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical  
 1188 ramifications. *Psychonomic Bulletin and Review*, *25*(1), 35–57.  
 1189 <https://doi.org/10.3758/s13423-017-1343-3>

## Predictors of change of inpatient psychotherapy for depression

48

- 1190 Wagner, K. D., Rouleau, M., & Joiner, T. (2000). Cognitive factors related to suicidal  
1191 ideation and resolution in psychiatrically hospitalized children and adolescents.  
1192 *American Journal of Psychiatry*, 157(12), 2017–2021.  
1193 <https://doi.org/10.1176/appi.ajp.157.12.2017>
- 1194 Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for*  
1195 *what makes psychotherapy work*. Routledge.
- 1196 Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (1993). SF-36 Health Survey Manual  
1197 and Interpretation Guide. *Boston New England Medical Centre*.  
1198 <https://doi.org/10.1097/00007632-200012150-00008>
- 1199 Webb, C. A., Cohen, Z. D., Beard, C., Forgeard, M., Peckham, A. D., & Björgvinsson, T.  
1200 (2020). Personalized prognostic prediction of treatment outcome for depressed patients  
1201 in a naturalistic psychiatric hospital setting: A comparison of machine learning  
1202 approaches. *Journal of Consulting and Clinical Psychology*, 88(1), 25–38.  
1203 <https://doi.org/10.1037/ccp0000451>
- 1204 Wells, A., & Fisher, P. (2016). *Treating depression: MCT, CBT and Third Wave Therapies*.  
1205 John Wiley & Sons, Incorporated.
- 1206

## Predictors of change of inpatient psychotherapy for depression

49

1207 Table 1

1208 *Sample characteristics of the full sample (N = 22,681)*

Characteristics	Full sample	n
Age at admission <i>M</i> (SD)	47.3 (12.7)	22,681
Sex <i>n</i> (%)		22,681
Female	13,174 (58.1)	
Male	9,507 (41.9)	
Educational score <i>M</i> (SD) <sup>a</sup>	3.2 (0.8)	21,792
Depression subtype <i>n</i> (%) <sup>b</sup>		22,681
Major depressive disorder, single episode (F32)	8,305 (36.7)	
Major depressive disorder, recurrent (F33)	14,335 (63.1)	
Persistent mood [affective] disorders (F34)	41 (0.2)	
Number of mental comorbidities <i>M</i> (SD) <sup>b</sup>	1.9 (1.1)	22,681
Most frequent mental comorbidities <i>n</i> (%) <sup>b</sup>		22,681
Mental and behavioural disorders due to psychoactive substance use (F1)	1,866 (8.2)	
Schizophrenia, schizotypal and delusional disorders (F2)	47 (0.2)	
Other mood (affective) disorders (F3)	1,640 (7.3)	
Neurotic, stress-related and somatoform disorders (F4)	8,894 (39.2)	
Behavioural syndromes associated with physiological disturbances and physical factors (F5)	1,928 (8.5)	
Disorders of adult personality and behaviour (F6)	2,384 (10.5)	
Behavioural and emotional disorders with onset usually occurring in childhood and adolescence (F9)	884 (3.9)	
Discharge type <i>n</i> (%)		11,024
Scheduled discharge	9,230 (83.7)	
Prematurely discontinued by patient	499 (4.5)	
Prematurely discharged by the team	91 (0.8)	
Prematurely by mutual agreement	967 (8.8)	
Transfer	192 (1.7)	
Marital status <i>n</i> (%)		22,658
Married	12,341 (54.5)	
Divorced	2,866 (12.6)	
Widowed	782 (3.5)	
Single	6,666 (29.4)	
In a relationship <i>n</i> (%)	17,751 (78.5)	22,626
Occupational status <i>n</i> (%)		22,650
Unemployed	3,018 (13.3)	
Retired	3,050 (13.5)	
Student, in training, home care	1,333 (5.9)	
Working full time	11,297 (49.9)	
Working half time	3,036 (13.4)	
Working occasionally	527 (2.3)	
Ability to work <i>n</i> (%)	7,593 (33.6)	22,631
First inpatient treatment <i>n</i> (%)	13,252 (81.5)	16,265
Outpatient psychotherapy <i>n</i> (%)	13,156 (58.1)	22,641
Outpatient psychiatric treatment <i>n</i> (%)	13,841 (61.2)	22,622
BDI-II score at baseline <i>M</i> (SD) <sup>c</sup>	28.2 (11.0)	20,111
PHQ-9 score at baseline <i>M</i> (SD) <sup>d</sup>	14.5 (5.6)	20,976
GAF score at baseline <i>M</i> (SD) <sup>e</sup>	48.5 (8.9)	22,038
BSI mean score at baseline <i>M</i> (SD) <sup>f</sup>	1.3 (0.7)	17,395

## Predictors of change of inpatient psychotherapy for depression

50

SF-36 mean score at baseline $M(SD)^g$	45.6 (16.4)	5,889
SWLS score at baseline $M(SD)^h$	16.8 (6.9)	20,722

1209 *Note.* M=mean, SD=Standard deviation, n=number; <sup>a</sup> Based on the German school system; scale from 0 (no  
 1210 degree) to 4 (general qualification for university entrance); <sup>b</sup> Diagnosis as given by practitioners according to  
 1211 ICD-10; <sup>c</sup> BDI-II = Becks Depression Inventory – II; 21 Items scale 0-63; <sup>d</sup> PHQ-9 = Patient Health  
 1212 Questionnaire-9; 9 items scale 0–27; <sup>e</sup> GAF = Global Assessment of Functioning; Scale 0–100; <sup>f</sup> BSI = Brief  
 1213 Symptom Inventory; 53 items, scale 0–4; <sup>g</sup> SF-36 = Short Form-36 Health Survey; <sup>h</sup> Satisfaction With Life Scale:  
 1214 5 Items, scale 5–35.

1215 Table 2

1216 *Coefficients of sociodemographic and clinical pretreatment characteristics on self-report factors at baseline/intake (T1), discharge (T2) and 6-months*  
 1217 *follow-up (T3)*

	Younger age			Lower education			Chronicity			Work related problems			Wish for retirement			Treatment resistance			Interpersonal dysfunction			Gender female		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Negative cognitions	<b>.14</b>	.05	-	.04	.03	-	<b>.08</b>	.03	-	<b>.13</b>	.12	-	.02	.04	-	.11	.08	-	<b>.15</b>	.09	-	<b>.15</b>	.04	-
Loss of interest	.04	.03	-.05	.03	.03	.07	.08	.02	.06	.10	.11	.03	.06	.06	.03	.10	.08	.01	.08	.06	-.03	<b>.12</b>	.01	-.04
Suicidality	<b>.15</b>	.06	-	.02	.02	-	<b>.13</b>	.04	-	.11	.09	-	.04	.03	-	.09	.06	-	.11	.08	-	.06	-.02	-
Rumination	.04	.06	-.05	.01	.03	.08	.06	-.01	.06	.07	.11	.05	.02	.03	.03	.11	.08	-	.09	.09	-.01	<b>.13</b>	.03	-.03
Aggressiveness	.06	<b>.12</b>	-	.04	.04	-	-	.01	-	.07	.09	-	-.02	.02	-	.06	.05	-	<b>.12</b>	.08	-	.05	-.01	-
Anxiety	.03	.03	-	.07	.05	-	.11	.01	-	.09	.1	-	.06	.03	-	<b>.12</b>	.07	-	.1	.07	-	.11	.02	-
Somatic anxiety symptoms	.01	.01	-	.08	.04	-	.10	.02	-	.06	.07	-	.08	.04	-	.06	.04	-	.05	.04	-	<b>.14</b>	.03	-
Paranoia	.11	.04	-	<b>.12</b>	.05	-	.03	.05	-	<b>.12</b>	.11	-	.02	.04	-	.07	.05	-	<b>.14</b>	.07	-	.08	-.01	-
Concentration difficulties	.08	.04	-	.04	.04	-	.06	.03	-	<b>.12</b>	<b>.12</b>	-	.02	.06	-	<b>.12</b>	.07	-	.1	.06	-	.10	-	-
Agoraphobia	.08	-	-	<b>.13</b>	.03	-	.08	.05	-	<b>.13</b>	.07	-	.06	.03	-	.09	.05	-	<b>.12</b>	.06	-	.08	.01	-
Feeling of weakness	-.03	-.01	-	.1	.05	-	.11	.02	-	.09	.08	-	.09	.04	-	.07	.03	-	.05	.03	-	<b>.19</b>	.03	-
Perceived disability	.07	.03	-	.01	.04	-	.08	.03	-	<b>.15</b>	.12	-	.06	.09	-	<b>.13</b>	.07	-	.08	.07	-	.09	.01	-
Physical disability	-.11	-.04	-	<b>.16</b>	.06	-	<b>.22</b>	.06	-	.11	.06	-	<b>.19</b>	<b>.12</b>	-	.07	.04	-	.02	.01	-	<b>.14</b>	.03	-
Life dissatisfaction	<b>.24</b>	.08	-	.09	.03	.09	.11	.01	.03	<b>.18</b>	.10	.03	.05	.03	.04	.06	.03	.02	.07	.05	-	.03	.01	-.01
Pain	-.01	.01	-.04	.09	.04	.04	.11	.02	.05	.05	.05	.01	.08	.05	.03	.05	.04	.02	.05	.03	-	<b>.22</b>	.04	.01

1218 Note. Standardized regression coefficients in a Bayesian path model with small variance prior (SD = 0.05). Bold numbers signify 95 % highest posterior density interval > 0.09 (small effect).  
 1219 T1 = clinical baseline data at intake, T2 = clinical data at discharge, T3 = clinical data at follow-up.

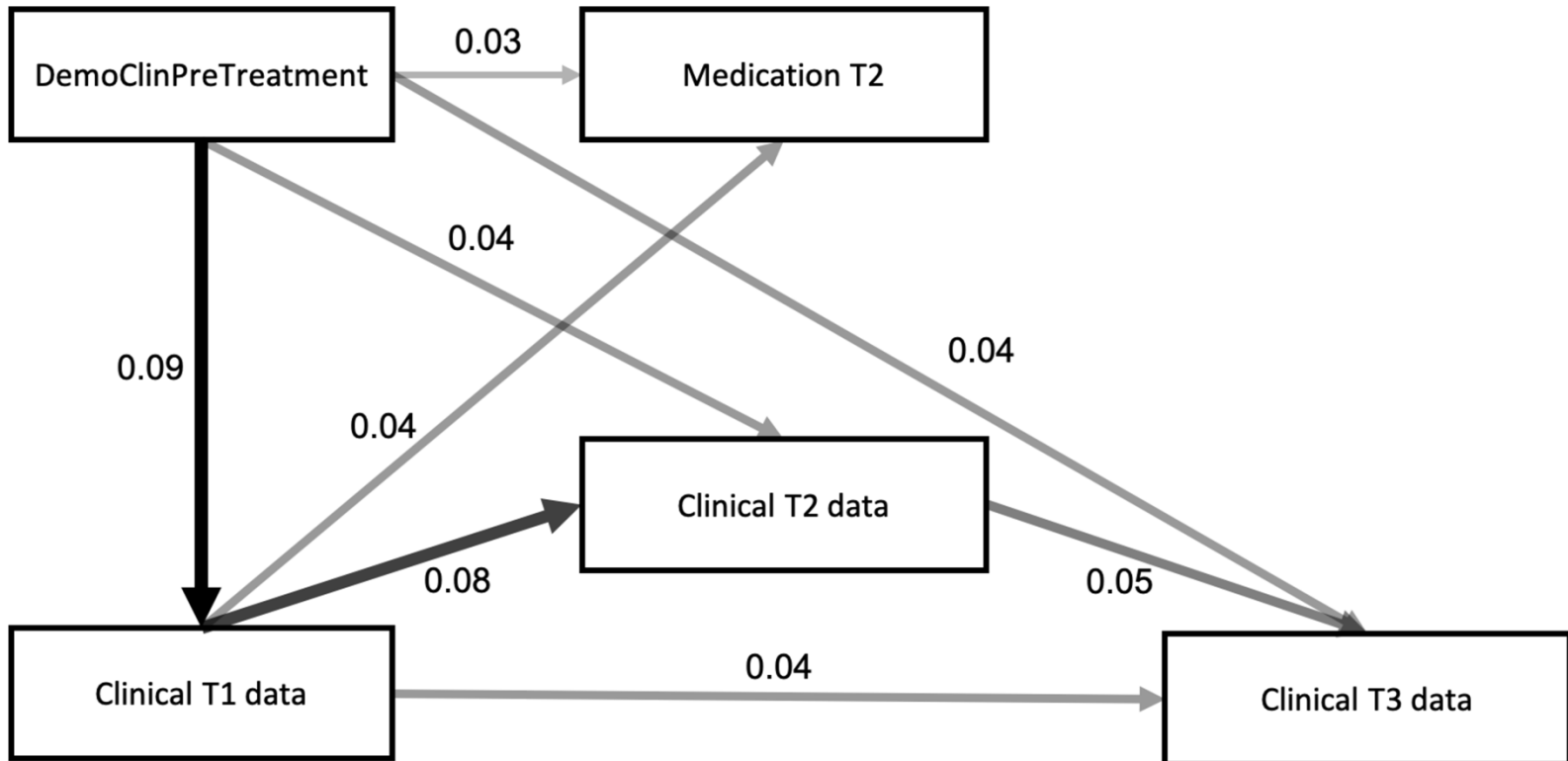
1220 Table 3  
 1221 *Autoregressive coefficients and explained variance for clinical data at baseline (intake), discharge*  
 1222 *and follow-up*

	Coefficient T1	Coefficient T2	Residual variance	R <sup>2</sup>
Negative cognitions T2	.33	-	.69	.31
Loss of interest T2	.26	-	.76	.24
Suicidality T2	.52	-	.63	.37
Rumination T2	.24	-	.77	.23
Aggressiveness T2	.32	-	.77	.23
Anxiety T2	.39	-	.69	.31
Somatic anxiety symptoms T2	.44	-	.69	.31
Paranoia T2	.42	-	.67	.33
Concentration difficulties T2	.44	-	.66	.34
Agoraphobia T2	.56	-	.58	.42
Feeling of weakness T2	.42	-	.69	.31
Perceived disability T2	.41	-	.65	.35
Physical disability T2	.65	-	.43	.57
Life dissatisfaction T2	.6	-	.5	.5
Pain T2	.52	-	.6	.4
Loss of interest T3	.13	.15	.74	.26
Rumination T3	.18	.21	.74	.26
Life dissatisfaction T3	.28	.32	.59	.41
Pain T3	.32	.25	.58	.42

1223 *Note.* Standardized autoregressive regression coefficients in a Bayesian path model with small variance prior (SD  
 1224 = 0.05); predicted variables in rows; T1 = clinical baseline data at intake, T2 = clinical data at discharge, T3 =  
 1225 clinical data at follow-up.

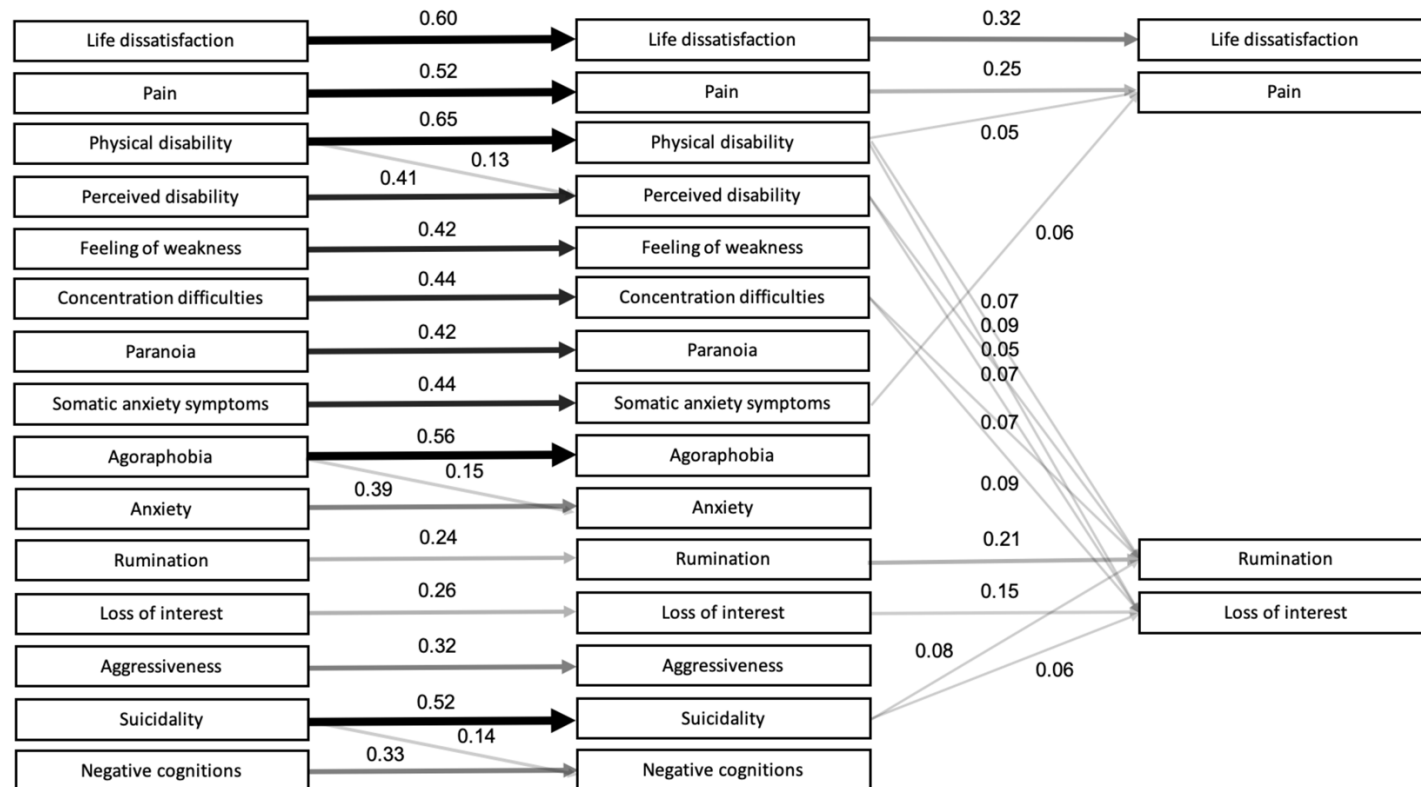
1226

**Figure 1.** Selected path model (Model 2) with averaged blockwise path coefficients.





1227 **Figure 2.** Bayesian path model using a small variance prior (SD = 0.05) with standardized regression coefficients of the relations  
 1228 between clinical T1, T2 and T3 data.



Note. Only edges from T1 to T2 with a 95 % highest posterior density interval > 0.09 (small effect) included; Only edges from T2 to T3 with a 95 % highest posterior density interval > 0 included. For detailed information regarding the underlying scales of the factors, we refer to the Supplemental Material 2.

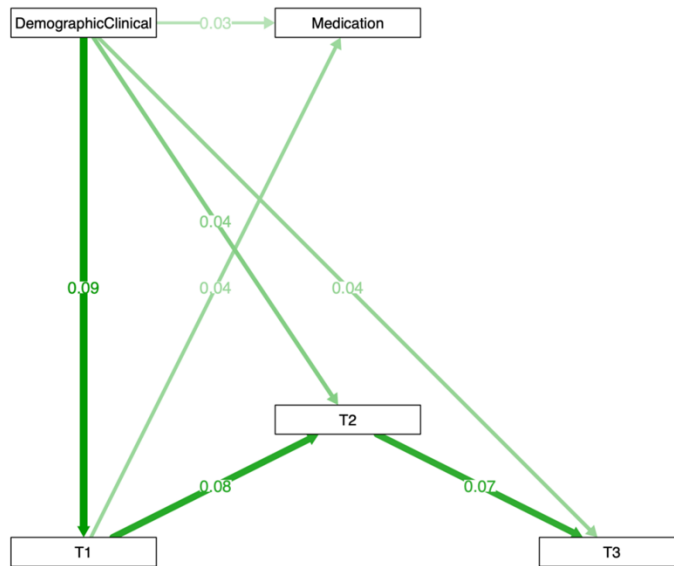
1230

Supplementary Material

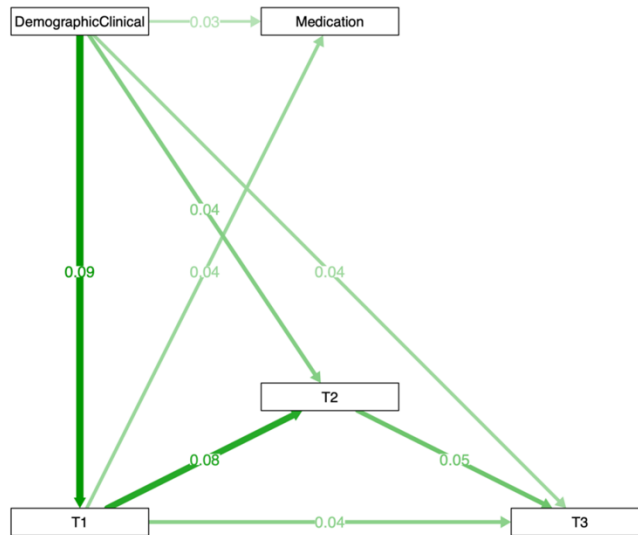
1231

Supplemental material 1. 10 path models

Model 1



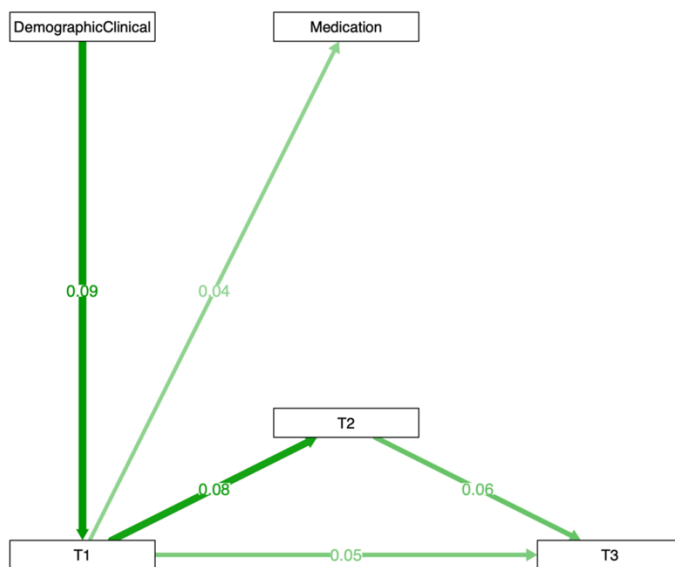
Model 2



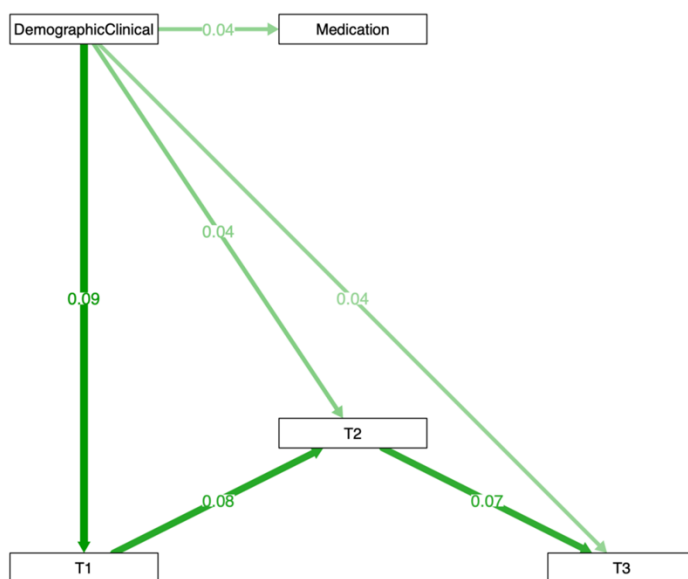
Predictors of change of inpatient psychotherapy for depression

56

Model 3



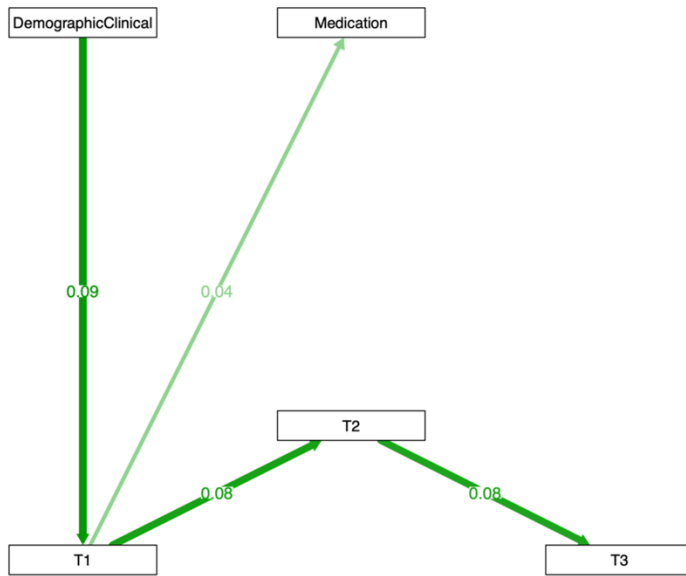
Model 4



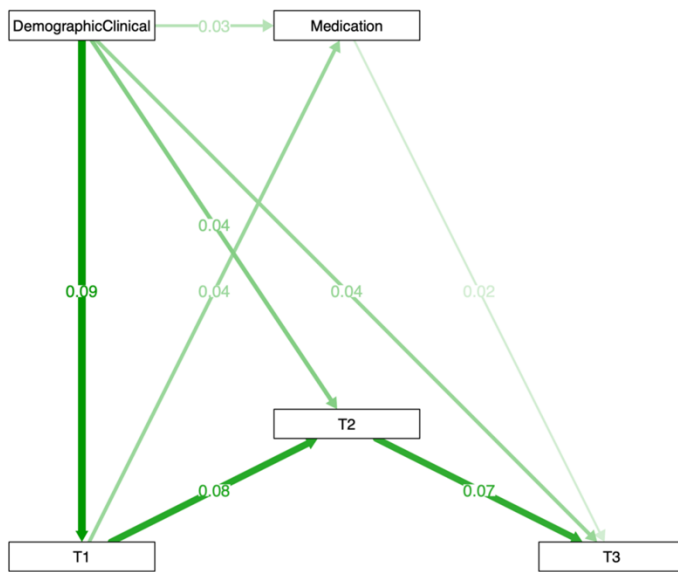
Predictors of change of inpatient psychotherapy for depression

57

Model 5



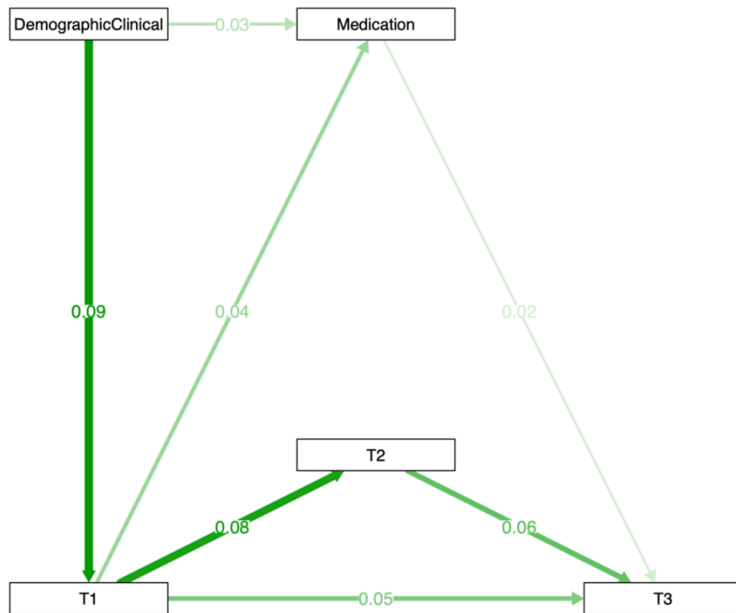
Model 6



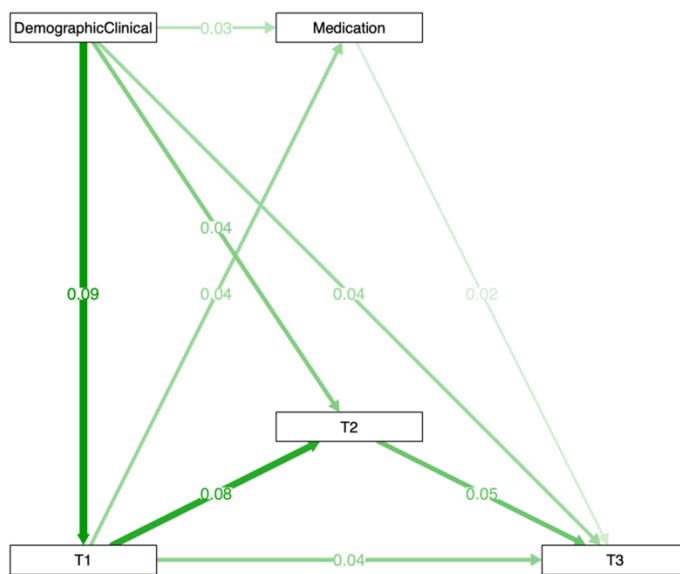
Predictors of change of inpatient psychotherapy for depression

58

Model 7



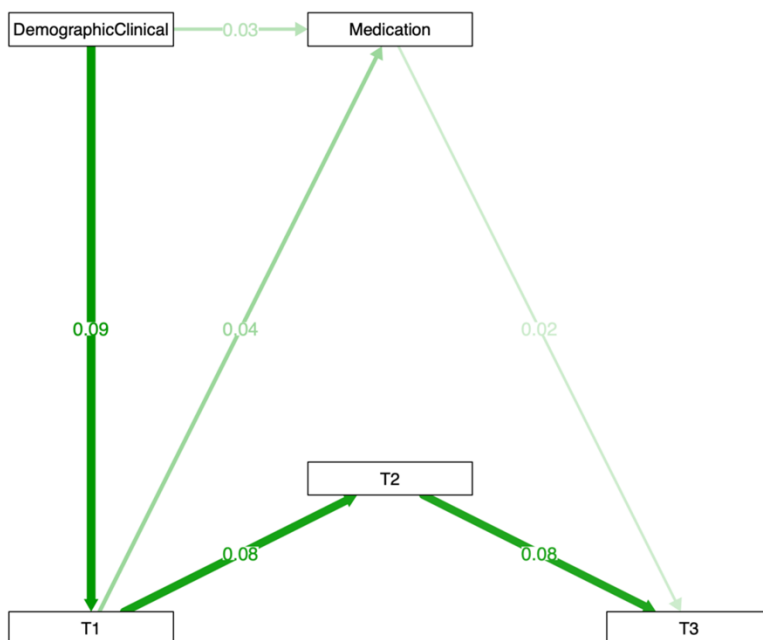
Model 8



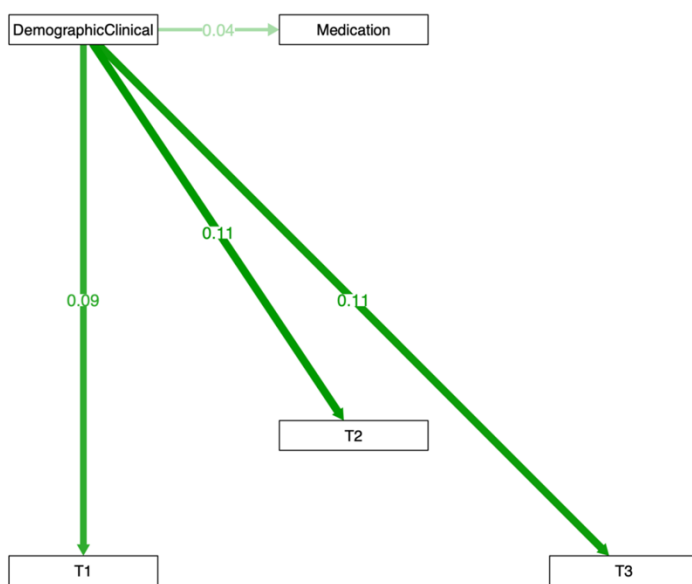
Predictors of change of inpatient psychotherapy for depression

59

Model 9



Model 10



1237 **Supplemental material 2.** Comparative fit indices1238 *Comparative fit indices of the 10 path models.*

	BIC	BIC2	AIC	CFI	GFI
Model 1	158	349	564	0.003	0.030
Model 2	0	0	0	0.000	0.029
Model 3	1192	1878	2652	0.015	0.001
Model 4	672	1244	1889	0.011	0.039
Model 5	1377	2254	3243	0.018	0.001
Model 6	387	476	576	0.003	0.030
Model 7	1487	1868	2298	0.013	0.000
Model 8	237	136	21	0.000	0.029
Model 9	1660	2232	2877	0.016	0.000
Model 10	44723	2232	2877	0.016	0.368

1239 Note. The ten path models can be found in Supplemental Material 1. BIC = Bayesian information criterion; BIC2 = Bayesian  
 1240 information criterion 2; AIC = Akaike information criterion; CFI = comparative fit index; GFI = Goodness of Fit Index.

## Anhang E: Studie 5

Received: 16 May 2019 | Revised: 4 September 2019 | Accepted: 14 September 2019

DOI: 10.1002/brb3.1447

## REVIEW

Brain and Behavior Open Access WILEY

## Assessing the unwanted: A systematic review of instruments used to assess negative effects of psychotherapy

Philipp Herzog<sup>1</sup>  | Sören Lauff<sup>1</sup> | Winfried Rief<sup>1</sup> | Eva-Lotta Brakemeier<sup>1,2</sup>

<sup>1</sup>Department of Clinical Psychology and Psychotherapy, Philipps-Universität Marburg, Marburg, Germany

<sup>2</sup>Marburg Center for Mind, Brain and Behavior (MCMBB), Marburg, Germany

## Correspondence

Philipp Herzog, Department of Clinical Psychology and Psychotherapy, Philipps-University of Marburg, Gutenbergstraße 18, D-35032 Marburg, Germany.  
Email: philipp.herzog@staff.uni-marburg.de

## Abstract

**Objective:** While the efficacy of psychotherapy in the treatment of mental disorders is well examined, systematic research into negative effects of psychotherapy seems comparatively rare. Therefore, this review evaluates instruments for assessing negative effects of psychotherapy in order to create a consensus framework and make recommendations for their assessment.

**Methods:** The study selection procedure follows current best-practice guidelines for conducting systematic reviews, with 10 included studies in three databases (PsycINFO, PubMed, and Web of Science). The nine instruments identified were each critically reviewed concerning the theoretical orientation, including the assessed domains of negative effects, psychometric properties, and diagnostic characteristics.

**Results:** Seventeen domains of negative effects of psychotherapy were identified but inconsistently assessed by the nine instruments. Most instruments provide some initial data on their psychometric properties. Regarding diagnostic characteristics, different item-response formats are used but often with reference to “attribution to therapy.”

**Conclusion:** This review indicates that the existing instruments for assessing negative effects of psychotherapy cover a wide range of relevant domains without any consensus on the most important ones and their psychometric properties are usually unsatisfactory. A framework for consensus, building on the definition and conceptualization of negative effects, is synthesized, and recommendations for improving the assessment are derived.

## KEYWORDS

assessment, instruments, negative effects, psychotherapy, side effects

## SUMMATIONS

- Through an extensive database search, nine relevant instruments measuring negative effects of psychotherapy were identified.
- These instruments cover a wide range of relevant domains and provide to some extent psychometric data, but these efforts are still in its infancy and need further examination.
- A framework toward a consensus definition and conceptualization was synthesized, and future directions were given

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Brain and Behavior* published by Wiley Periodicals, Inc.

*Brain and Behavior*. 2019;00:e01447.

<https://doi.org/10.1002/brb3.1447>

[wileyonlinelibrary.com/journal/brb3](http://wileyonlinelibrary.com/journal/brb3) | 1 of 13



in order to improve the assessment of negative effects of psychotherapy.

## LIMITATIONS

- The main limitation is that this review comprises a relatively small number of eligible studies and instruments that investigate and assess negative effects.
- A further limitation is that this review focuses on psychometric properties of each instrument and therefore has not considered their clinimetric properties.

## 1 | INTRODUCTION

The efficacy of psychotherapy for treating mental disorders has been well examined over several decades (Huhn et al., 2014; Scheffl, Guhn, Brakemeier, Sterzer, & Köhler, 2019). In particular, the evidence base of cognitive behavioral therapy (CBT) is considered to be robust and strong (Butler, Chapman, Forman, & Beck, 2006; David, Cristea, & Hofmann, 2018; Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012). However, in comparison with research on positive effects supporting the efficacy of psychotherapy, research on negative effects is still rare. At the beginning of the 21st century, attention increased to the negative effects of psychotherapeutic interventions (Crawford et al., 2016; Scott, 2017). Psychotherapists as well as researchers highlight that negative effects are common in face-to-face care, for example, in group psychotherapy (Schneibel et al., 2017), as well as in Internet-based interventions (Boettcher, Rozental, Andersson, & Carlbring, 2014). Linden et al. (2018) define negative effects<sup>1</sup> as adverse events (AEs) related to treatment comprising side effects (SE), malpractice (MP), and unethical conduct (UC).<sup>2</sup> In contrast to MP and UC, SE are AEs caused by a correctly performed psychotherapy, that is, lege artis delivered treatment, and comprises different life domains (such as transient symptom deterioration, conflicts in interpersonal relationships, and stigmatization concerns). According to this definition, SE may be not only unexpected, but also expected and sometimes even intended effects. Accordingly, research suggests that approximately 58.7% of all patients in psychiatric hospitals, 45.2% in psychosomatic hospitals, and 93.8% in a convenience sample of former psychotherapy patients report at least one negative effect during psychotherapy (Ladwig, Rief, & Nestoriuc, 2014; Rheker, Beisel, Kräling, & Rief, 2017). This high prevalence of negative effects emphasizes the importance of evaluating negative effects not only once at the end of treatment, but also in the course of treatment and after its completion. However, recent reviews have shown that instruments assessing negative effects are heterogeneous and not systematically reported in randomized controlled trials (Jonsson, Alaia, Parling, & Arnberg, 2014), for example,

in studies on persistent depressive disorder (Meister et al., 2016). In line with this result, the Consolidated Standards of Reported Trials (CONSORT) group claims that monitoring of negative effects in clinical studies on behavioral health is limited (Ioannidis et al., 2004).

Thus, despite their high prevalence, there are comparatively few systematic research studies on negative effects of psychotherapeutic interventions. Systematic research of their occurrence is hindered by a confusion of different definitions of negative effects (Parry, Crawford, & Duggan, 2016) as well as the diversity of terms and their inconsistent use (Linden, 2013), which leads to difficulties in developing adequate instruments for assessing negative effects. In this context, researchers use different terms such as "deterioration effects," "side effects," "negative effects," "negative outcome," "unwanted/undesirable effects," "adverse events/effects," "harm," "mistakes," and "treatment-emergent reactions" synonymously, fostering confusion among researchers and psychotherapists. There are thus no instruments that are accepted worldwide as a "gold standard" and used consistently in studies. In conclusion, there is little systematic research on negative effects of psychotherapy, leading to an increasing need to improve current research methods on a sound theoretical basis.

In order to overcome current problems and the gap between their relevance and evaluation, the most recent methodological recommendations for trials of psychological interventions explicitly emphasize that the assessment of negative effects of psychotherapy should be performed using suitable methods of evaluation (Guidi et al., 2018). Thus, the main objective of this systematic review is to summarize and examine the available instruments for assessing negative effects in psychotherapy. To date, no review has focused on assessment tools, their theoretical foundation, or psychometric quality, underscoring their unique contribution to an often-neglected research field. Moreover, the secondary objectives of the present review are (a) to create a framework of negative effects on an empirical basis, (b) to give recommendations for improving the assessment instruments, and (c) to provide an outlook on the development of future instruments, including theoretical considerations on the framework.

## 2 | MATERIALS AND METHODS

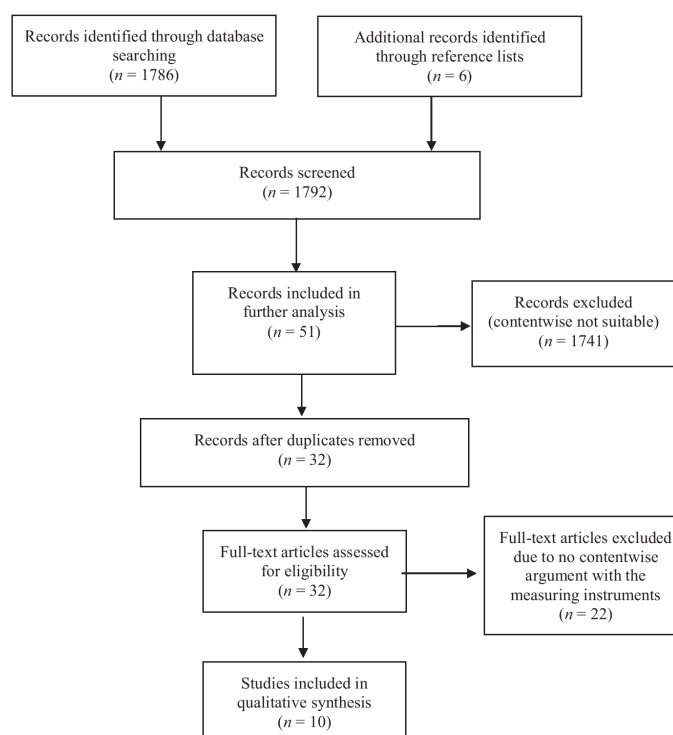
### 2.1 | Study selection

The entire study selection process followed the current guidelines of meta-analyses and systematic reviews (Cuijpers, 2016). In August 2018, the first two authors conducted a computer database search of PsycINFO, PubMed, and Web of Science. The literature search was limited to the databases mentioned above, as redundancies already occurred there. The following search terms and logic were used in the database search: (Deterioration effects OR negative effects OR adverse effects OR negative treatment effects OR negative outcome OR side effects OR unwanted effects) AND (psychotherapy\*) AND (instrument OR inventory OR questionnaire OR assessment OR scale OR survey). The database search was limited to "Title," "Abstract," and "Keywords."

<sup>1</sup>Sometimes referred to as treatment-emergent reactions.

<sup>2</sup>Formerly designated as malpractice and unethical behavior (MUB) (Ladwig et al., 2018).

**FIGURE 1** Flow diagram of included studies according to PRISMA (Moher et al., 2009)



The database search with the defined parameters yielded 1,786 hits in PsycINFO, PubMed, and Web of Science (349 in PsycINFO, 233 in PubMed, and 1,204 in Web of Science). In addition, six articles were identified by reference list screening, resulting in a total of 1,792 articles. These records were carefully screened for title and abstract. Thereafter, 1,741 articles were excluded because their contents were considered unsuitable for this review. Of the remaining 51 matching hits, 19 duplicates were identified and removed for further analysis. The number of articles in the full-text analysis was thus reduced to 32. These 32 articles were read by the first two authors, and any that did not include instruments for assessing negative effects of psychotherapy were excluded from further analysis. In total, the analysis process selected 10 studies in which nine instruments were described. All articles have been included in the qualitative synthesis of this review.

All search criteria were limited to full-text articles published from 1986 to 2018, since the Vanderbilt Negative Indicators Scale (VNIS; Suh, Strupp, & O'Malley, 1986) was the first published structured assessment scale of negative effects of psychotherapy.

The objective of this study was to conduct a comprehensive review of current practice in the evaluation of all published qualitative and quantitative research on negative effects. Thus, studies were not excluded because of their psychometric properties, such as no data on reliability and validity or theoretical foundation. In addition,

no restrictions were imposed on the place of origin of the studies, the year of publication, and the type of mental disorders presented in the sample. Nevertheless, the database search was limited to the availability of full texts in English and German.

Figure 1 displays a flowchart diagram of the selection process according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Moher, Liberati, Tetzlaff, & Altman, 2009).

## 2.2 | Instrument assessment

This systematic review is based on review frameworks for the evaluation of instruments developed through the integration of standards and guidelines for psychometric evaluation, including the Journal Article Reporting Standards (APA Publications & Communications Board Working Group on Journal Article Reporting Standards, 2008) and guidelines for the evaluation of test instruments (Cicchetti, 1994). In addition, criteria for the evaluation of instruments according to Groth-Marnat (2009) were used.

The main objective of the review framework was to identify the theoretical orientation and psychometric characteristics of the relevant instruments. Due to their theoretical orientation, the underlying theoretical construct was identified and clustered in relation to

the assessed domains. To evaluate the psychometric properties of the instruments, the following components were identified as relevant within this emerging field of research: validity (content-related, construct, predictive/criterion) and reliability (internal consistency, test-retest, inter-rater). Diagnostic characteristics were based on number of items, item sensitivity, as well as country of origin and language versions.

All included studies were independently coded by the first two authors to allow complete extraction of relevant characteristics of each instrument and to ensure cross-checking. The following method was used for coding: Each reviewer read the identified studies and encoded all information related to the above review framework; the extracted information from the individual studies was then discussed and systematically included in the review framework. When discrepancies and misunderstandings in coding occurred, they were resolved by discussing the information, aimed at reaching a consensus between the first two authors.

### 3 | RESULTS

Table 1 presents a summary of the key characteristics of the nine identified assessment tools. The following instruments were identified: (a) Vanderbilt Negative Indicators Scale (VNIS; Suh et al., 1986); (b) Unwanted Effects-Adverse Treatment Reaction checklist (UE-ATR; Linden, 2013); (c) Inventory for the Assessment of Negative Effects of Psychotherapy (INEP; Ladwig et al., 2014); (d) Experiences of Therapy Questionnaire (ETQ; Parker, Fletcher, Berk, & Paterson, 2013); (e) Negative Effects Questionnaire (NEQ; Rozental, Kottorp, Boettcher, Andersson, & Carlbring, 2016); (f) Unwanted Events-Adverse Treatment Reactions in the context of group psychotherapy (UE-G; Linden, Walter, Fritz, & Muschalla, 2015); (g) Side Effects of Psychotherapy Scale (SEPS; Moritz et al., 2015); (h) Exploitation Index (EI; Epstein & Simon, 1990); and (i) Positive and Negative Effects of Psychotherapy Scale (PANEPS; Moritz et al., 2018). The VNIS, UE-ATR, and EI are therapist-rated instruments, while the INEP, ETQ, NEQ, SEPS, and PANEPS are patient-rated instruments; the UE-G is a patient-rated instrument in the context of group psychotherapy. The EI (Epstein & Simon, 1990) was excluded for further analyses because the authors had no access to the full text in the common database and were not successful with the full-text request. Furthermore, the UE-ATR (Linden, 2013) was excluded for the evaluation of psychometric properties because the authors describe the checklist only as a useful, informative, and attention-grabbing tool for recognizing negative effects, in contrast to a scale with solid psychometric properties.

#### 3.1 | Theoretical orientation

The choice of an adequate instrument for measuring negative effects of psychotherapy is often determined by its theoretical orientation with the intended use. A total of 17 domains were identified to evaluate the theoretical constructs of the individual instruments.

Table 2 presents a summary of the eight reviewed instruments for measuring common domains of negative effects. Three observations can be recorded:

1. No instruments are identical with regard to the recording of different domains;
2. No domain was covered by all instruments; and
3. Three domains were assessed by all but one instrument.

"Therapeutic misconduct" was assessed by all instruments except the UE-ATR, "deterioration/emergence of symptoms" by all instruments except the ETQ, and "quality of therapy" by all except the INEP. Furthermore, "stigma" was recorded by five out of eight instruments and "therapeutic relationship (e.g., dependency, idealization)" by four out of eight, which indicates that they are relevant domains. Other different domains were assessed only by some of the instruments reviewed. "Treatment response" was assessed by NEQ, SEPS, and PANEPS, "changes and strains in life areas (e.g., work, family, relationship)" by UE-ATR and INEP, and "wanted effects" by ETQ, SEPS, and PANEPS. Another visual analysis revealed that some domains were assessed by only one instrument. "Expectation towards therapy" was assessed only by VNIS; "intrapersonal changes" only by INEP; "therapy setting (e.g., room size)," "relationship to other patients," "global experience," and "hopelessness" only by UE-G; and "Well-being of the patient," "noncompliance to treatment," and "Prolongation of the treatment" only by UE-ATR. The UE-ATR and PANEPS have the largest overlap of negative effect indicators, with 8 out of 17 indicators.

#### 3.2 | Evaluation of psychometric properties

Table 3 summarizes the psychometric properties of seven instruments (without UE-ATR). Three types of validity aspects were identified as relevant (content-related, construct, and criterion). In addition, three types of reliability aspects were considered (internal consistency, test-retest, and inter-rater).

Content-related validity was defined as the representativeness and relevance of the assessment tool for the underlying construct assessed with this tool (Groth-Marnat, 2009). Two instruments, INEP and NEQ, showed strong content-related validity through qualitative analysis of patients' experiences, a pilot study, a comprehensive literature review, or the advice of experts in the respective research area. In addition, PANEPS was theoretically based on INEP, UE-ATR, and SEPS, integrating all the advantages of the individual instruments and indicating good content-related validity. One instrument, UE-G, also demonstrated good content-related validity by aligning its item pool with the concepts of other researchers and their clinical experiences.

Construct validity was defined by the extent to which the instrument measures a theoretical basic construct (Groth-Marnat, 2009). Four of the seven reviewed instruments provided relevant research data on their construct validity by means of factor analysis or correlational analysis. Three of them, INEP, NEQ, and PANEPS,

**TABLE 1** Summary of assessment instruments for negative effects in psychotherapy

Instrument	Author	Items	Domains	Item sensitivity	Country of origin	Language
Vanderbilt Negative Indicators Scale (VNIS)	Suh et al. (1986)	42	Unrealistic expectations Deficiencies in therapeutic commitment Inflexible use of therapeutic techniques Poor therapeutic relationship Poor match	6 points (different subscales)	USA	English
Unwanted Effects-Adverse Treatment Reaction checklist (UE-ATR)	Linden (2013)	16	UE classes as follows: Lack of clear treatment results Prolongation of treatment Noncompliance of the patient Emergence of new symptoms Deterioration of symptoms Negative well-being of the patient Strains in the patient-therapist relationship Very good patient-therapist relationship Strains in family relations Changes in family relations Strains in work relations Changes in the work situation Sick leave of the patient Problems in the extended social net Any change in the life circumstances of the patient Stigmatization	5 points (severity) 6 points (relation to treatment) 8 points (context of development)	Germany	German, English
Inventory for the Assessment of Negative Effects of Psychotherapy (INEP)	Ladwig et al. (2014)	21	Intrapersonal changes Therapeutic misconduct Relationship Family and friends Work Stigma	7 points (bipolar) 4 points (unipolar)	Germany	German, English
Experiences of Therapy Questionnaire (ETQ)	Parker et al. (2013)	63	Negative therapist Preoccupying therapy Beneficial therapy Idealization of therapist	5 points (strongly disagree-strongly agree)	Ireland	English
Negative Effects Questionnaire (NEQ)	Rozental et al. (2016)	32	Symptoms Quality Dependency Stigma Hopelessness Failure	2 points (yes-no) 5 points (perceived burden) 2 points (relation to treatment)	Sweden	Danish, Dutch, English, Finnish, French, German, Italian, Japanese, Norwegian, Spanish, Swedish
Unwanted Events and Adverse Treatment Reactions in the context of group psychotherapy (UE-G)	Linden et al. (2015)	46	Group size or room Content Other group members Therapist Repercussions Global experience	5 points (extent of perceived burden)	Germany	German
Side Effects of Psychotherapy Scale (SEPS)	Moritz et al. (2015)	147	Wanted effects related to treatment Adverse treatment reactions Malpractice Unethical conduct Deterioration of illness related to treatment Treatment nonresponse related to treatment Other treatment-emergent reaction	4 points (true-not true)	Germany	German
Exploitation Index (EI)	Epstein and Simon (1990) <sup>a</sup>	32	NA	4 points (never-often)	USA	English
Positive and Negative Effects of Psychotherapy Scale (PANEPS) <sup>b</sup>	Moritz et al. (2018)	43	Positive effects Side effects Malpractice Unethical conduct	4 points (true-not true)	Germany	German, English

<sup>a</sup>No access to full article in the common database available. In the end, the authors sent a full-article request without success.

<sup>b</sup>PANEPS is a revised and shortened version of the SEPS.

provided information on the factor structure of the questionnaires, while one instrument, ETQ, showed significant correlations with related constructs.

Criterion validity was defined as comparison of the scores on the instrument with performance on another external tool (Groth-Marnat, 2009). Only one instrument, INEP, published data on the criterion validity through regression analysis on an external criterion, in fact "satisfaction with therapy."

Reliability of an instrument has been defined as the extent to which a score is stable, consistent, predictable, and accurate over time (Groth-Marnat, 2009). Relevant indices of reliability are the internal consistency assessed by Cronbach's  $\alpha$ , the test-retest reliability, and the inter-rater reliability. With the exception of UE-G, all instruments provided some reliability data, indicating a moderate to high reliability for these instruments. Alpha coefficients of .70–.79 are considered "fair," .80–.89 as "good," and .90 or higher as "excellent" (Cicchetti, 1994), while reliability should be at least .90 for clinical decisions and .70 for research (Groth-Marnat, 2009).

Five out of seven validation studies used the internal consistency assessed by Cronbach's  $\alpha$ , which ranged from .55 to .97, indicating acceptable to excellent reliability. The test-retest reliability was only reported by ETQ, leading to values of .76–.96. The inter-rater reliability was only given for one instrument, VNIS, resulting in values of .89–.96.

### 3.3 | Diagnostic characteristics

In terms of their practical use in diagnostics, the choice of instrument for clinicians and researchers is often guided by various practical considerations such as the time of administration and user-friendliness. The most economic self-administered instruments in administration are the UE-ATR and INEP, with only 16 and 21 items, respectively. With regard to item sensitivity, the results indicate the wide variety of item-response categories used in the studies, ranging from multiple dichotomous and continuous scales used within an instrument to only one continuous item-response

**TABLE 2** Diagnostic domains of assessment instruments

	VNIS <sup>a</sup>	UE-ATR <sup>b</sup>	INEP <sup>c</sup>	ETQ <sup>d</sup>	NEQ <sup>e</sup>	UE-G <sup>f</sup>	SEPS <sup>g</sup>	PANEPS <sup>h</sup>
Stigma	○	●	●	●	●	○	○	●
Therapeutic misconduct	●	○	●	●	●	●	●	●
Deterioration/emergence of symptoms	●	●	●	○	●	●	●	●
Quality of therapy	●	●	○	●	●	●	●	●
Therapeutic relationship (e.g., dependency, idealization)	●	●	○	●	○	○	○	●
Expectations towards therapy	●	○	○	○	○	○	○	○
Treatment response	○	○	○	○	●	○	●	●
Intrapersonal changes	○	○	●	○	○	○	○	●
Changes and strains in life areas (e.g., work, family, relationship)	○	●	●	○	○	○	○	○
Wanted effects (e.g., benefit)	○	○	○	●	○	○	●	●
Therapy setting (e.g., room size)	○	○	○	○	○	●	○	○
Relationship to other patients	○	○	○	○	○	●	○	○
Global experience	○	○	○	○	○	●	○	○
Hopelessness	○	○	○	○	○	●	○	○
Well-being of the patient	○	●	○	○	○	○	○	○
Noncompliance to treatment	○	●	○	○	○	○	○	○
Prolongation of the treatment	○	●	○	○	○	○	○	○

Note: ● Assessed; ○ not assessed. This overview does not include the Exploitation Index (Epstein & Simon, 1990) due to no access to full-text publication. Although the UE-G is an instrument for measuring negative effects of group therapy, it was included in this review to identify its underlying theoretical foundation.

<sup>a</sup>Vanderbilt Negative Indicators Scale (Suh et al., 1986).

<sup>b</sup>Unwanted Effects-Adverse Treatment Reaction checklist (Linden, 2013).

<sup>c</sup>Inventory for the Assessment of Negative Effects of Psychotherapy (Ladwig et al., 2014).

<sup>d</sup>Experiences of Therapy Questionnaire (Parker et al., 2013).

<sup>e</sup>Negative Effects Questionnaire (Rozental et al., 2016).

<sup>f</sup>Unwanted Events-Adverse Treatment Reactions in the context of group psychotherapy (Linden et al., 2015).

<sup>g</sup>Side Effects of Psychotherapy Scale (Moritz et al., 2015).

<sup>h</sup>Positive and Negative Effects of Psychotherapy Scale (Moritz et al., 2018).

**TABLE 3** Evaluation of psychometric properties of instruments for measuring negative effects in psychotherapy

Instrument <sup>a</sup>	Studies	Total N	Validity			Reliability			
			Content-related	Construct	Criterion	Internal consistency	Test-re-test	Inter-rater	Population
Vanderbilt Negative Indicators Scale (VNIS)	1. Strauß, Strupp, Burgmeier-Lohse, Wille, and Storm (1992)	18	NA	NA	NA	NA	NA	0.89–0.96	Clinical
Inventory for the Assessment of Negative Effects of Psychotherapy (INEP)	1. Ladwig et al. (2014)	200 (195)	Literature review Consulting researchers	Factor analysis: 7-factor solution explains 55.8% of total variance	Significant regression on satisfaction with therapy	0.86	NA	NA	Clinical
Experiences of Therapy Questionnaire (ETQ)	1. Parker et al. (2013) 2. Parker, Paterson, Fletcher, McClure, and Berk (2014)	707 46	NA	Significant correlations with related constructs of the therapist satisfaction scale	NA	0.90–0.96	0.76–0.96	NA	Clinical
Negative Effects Questionnaire (NEQ)	1. Rozentel et al. (2016)	653	Results of consensus statement Pilot study Qualitative analysis of patients' experiences	Factor analysis: 6-factor solution explains 57.64% of total variance	NA	0.72–0.93	NA	NA	Clinical
Unwanted Events—Adverse Treatment Reactions in the context of group psychotherapy (UE-G)	1. Linden et al. (2015)	71	Clinical experiences Concepts of Roback (2000) and Burlingame, MacKenzie, and Strauß (2004)	NA	NA	NA	NA	NA	Clinical
Side Effects of Psychotherapy Scale (SEPS)	1. Moritz et al. (2015)	173 (85)	NA	NA	NA	0.55–0.97	NA	NA	Clinical
Positive and Negative Effects of Psychotherapy Scale (PANEPS)	1. Moritz et al. (2018) 2. Peth, Jelinek, Nestorluc, & Moritz, (2018)	135	Theoretical based on the SEPS, INEP, and UE-ATR	Factor analysis: 4-factor solution	NA	0.72–0.92	NA	NA	Clinical

Abbreviation: NA, not available.

<sup>a</sup>The summary does not include the UE-ATR checklist (Linden, 2013) as the authors did not aim to develop a scale with psychometric properties, or the Exploitation Index (Epstein & Simon, 1990) due to no access to the full-article publication.

category. What they all have in common is that attribution to therapy is important when recording negative effects of psychotherapy (cf. Linden et al., 2018). In line with this, the INEP, NEQ, and UE-ATR query the relationship between negative effect and treatment. Most instruments have been developed in English- and German-speaking countries and therefore are only available in English and/or German, but one instrument, the NEQ, has already been translated into several languages.

## 4 | DISCUSSION

The main objective of this study was to conduct a systematic review of the current instruments for assessing the negative effects of psychotherapy by evaluating their theoretical orientation and psychometric properties, including diagnostic characteristics. This will help researchers and practitioners to select the appropriate tools for evaluating negative effects for their respective purposes, as proposed by Guidi et al. (2018). A secondary objective was to derive a bottom-up framework of negative effects from the available data in order to refine the conception and definition of negative effects in psychotherapy and to give recommendations for improving assessment.

Overall, the results of this systematic review indicate that the available instruments for negative effects and their empirical evidence can largely be classified as insufficient, despite promising study approaches in recent years. Since 1986, only 32 studies have been published using standardized survey methods. Overall, nine instruments were identified to assess some sort of negative effects: Eight of them were studied in light of their theoretical orientation and seven of them with regard to the psychometric properties reported. The underlying theoretical constructs were clustered into 17 different domains across all instruments. Negative effects seem to be composed of several different and to some extent distinct factors. The largest coverage of the evaluated domains was achieved by the UE-ATR. The dimension "expectations towards psychotherapy" is considered as a negative effect by only the VNIS which has been developed from a psychodynamic perspective. This dimension must therefore be interpreted under the light of its therapeutic orientation. In terms of psychometric properties, most data were available for the INEP, NEQ, and PANEPS, which all reported promising results supporting different aspects of validity and internal consistency. Since side effects are often single events rather than dimensional phenomena, side effect instruments should primarily be instruments for event monitoring. In the context of psychometric properties as empirical criteria in the assessment of side effects, therefore, the content validity seems to be most important, in the sense of assessing whether all possible side effects as events are covered by the respective scale. In addition, the PANEPS assesses both negative and positive effects of psychotherapy. With regard to practical issues, the INEP is identified as the most economic self-report questionnaire (with 21 items) and the UE-ATR as the most economic assessment tool for practitioners (with only 16 items).

### 4.1 | Defining negative effects of psychotherapy: toward a consensus framework

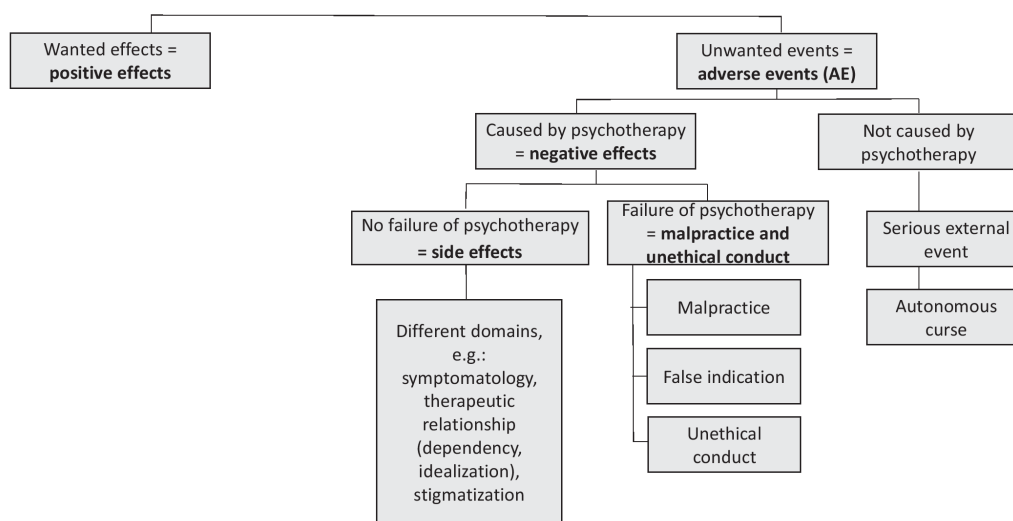
This review investigated the theoretical orientation of negative effect instruments. No instrument is able to record all derived domains, and most studies lack clear definitions of negative effects. When definitions are given, they vary between studies. As there is no consensus on a model that covers all positive and negative effects of psychotherapy, assessing the diagnostic domains of all existing instruments could be useful in determining the importance of domains for building a framework. The majority of the reviewed instruments assess six core characteristics: (a) stigma, (b) therapeutic misconduct, (c) deterioration/emergence of symptoms, (d) quality of therapy, (e) therapeutic relationship (e.g., dependency and idealization), and (f) treatment response. Therapeutic misconduct is an important and very sensitive topic. Therapeutic misconduct is recorded by several instruments and can be regarded as a negative effect of an incorrectly performed therapy, whereby side effects represent negative effects of a correctly performed therapy. Of note, therapeutic misconduct can therefore never be accepted and should always lead to professional or legal consequences. Some authors argue that the wanted effects (i.e., positive effects of psychotherapy) should also be evaluated in order to minimize negative priming (Moritz et al., 2018). Negative priming may cause negative expectations about the occurrence of side effects of a particular treatment, even in psychological interventions (Bootzin & Bailey, 2005), and can therefore be associated with reported side effects—a phenomenon called the "nocebo effect," which so far has been used mainly in psychopharmacological trials (Colloca & Miller, 2011). In recent years, more and more researchers have considered negative expectations as a key feature in mental disorders (Rief et al., 2015). By assessing the side effects of psychological interventions, these side effects might be at least partially triggered by the nocebo effect. Several authors gave initial indications on how to deal with the nocebo effect (Webster, Weinman, & Rubin, 2016), for example, by reducing expectations of symptoms or limiting symptom suggestions. In this context, the informed consent could be adapted (Cohen, 2014). It should be noted, however, that this hypothesis has not (yet) been supported by independent studies and need further empirical data in the context of psychological treatments. Patients might be "nocebo-susceptible" to side effects, which may be interpreted as one of many patient criteria that increase the risk of side effects. Future research should pay more attention to risk factors of side effects.

In addition, researchers discuss whether positive side effects and by-products should also be included in the framework of negative effects (Hoyer, 2016). The authors argue that the classic model of side effects in psychological interventions was derived from pharmacological models of side effects and their focus on symptom deterioration, and therefore cannot cover the complexity of the biopsychosocial model of medicine and psychological interventions. The spectrum of potential negative effects in psychological interventions is greater than in pharmacological treatments, as it also

includes negative events in social interactions (Szapocznik & Prado, 2007). However, research on this concept is not yet well established. For example, the improvement of quality of life was considered as one of these positive side effects, whereas other authors argue that this should always be addressed as a goal of therapy and therefore considered as a (secondary) outcome (Caspar & Jacobi, 2007). In addition, the concept of positive side effects may be misleading as most instruments also covered areas other than symptom deterioration. In light of the proposed model, positive side effects might be covered by the assessment of several domains and an indication of their valence (i.e., side effects do not necessarily have to be negative according to this definition). Therefore, future studies should consider evaluating the valence of side effects to determine their effects on treatment and outcome.

The terms "side effects" and "negative effects" are sometimes used interchangeably in the relevant literature, leading to an inaccurate use of the technical terms. The authors therefore argue for a clear distinction between them in order to clarify the underlying constructs and to counter the confusion of different terminologies for the same construct. This should encourage the use of consistent and uniform terms in future research. It seems helpful to classify wanted effects (positive effects) and unwanted events (often referred to as adverse events) within a framework based on the previous findings of diagnostic features assessed by these instruments. In other terminology, adverse events (AEs) can be divided into treatment-emergent reactions (AE related to treatment) and those unrelated to treatment (Linden, 2013; Moritz et al., 2018). This classification is displayed in Figure 2. On this basis, the authors try to create a consensus definition that is consistent with a

recently published article by Linden et al. (2018). By integrating and synthesizing these findings within one framework, negative effects can be defined as unwanted events caused by psychotherapy. In addition, an attempt is made to distinguish between side effects and malpractice/unethical conduct. While side effects are unwanted events caused by lege artis psychotherapy (i.e., psychotherapy performed correctly), malpractice/unethical behavior can be classified as unwanted events caused by failures in psychotherapy. Side effects should include several domains, for example, not exclusively but most importantly: (a) stigma, (b) changes in symptoms (e.g., deterioration or emergence of symptoms), (c) changes and strains in life areas (e.g., work, family, relationship), and (d) therapeutic relationship (dependency and idealization). Thus, negative effects include side effects as well as malpractice and/or unethical behavior. On the other hand, wanted events caused by psychotherapy can be classified as positive effects. Furthermore, unwanted events unrelated to psychotherapy may also occur due to serious external events or the autonomic course of the disorder. In this context, it is important to ask about the relationship between unwanted events and treatment as a crucial criterion. The various self-report scales indicate the causal relationship to treatment by asking the patient whether the unwanted event experienced and reported was likely caused by (a) "the treatment I received" or (b) "other circumstances" (e.g., INEP and NEQ). While the UE-ATR as a therapist rating indicates the relation to treatment on a 5-point scale ranging from 1 = unrelated, 2 = probably unrelated, 3 = possibly related, 4 = probably related, to 5 = related. However, establishing causal relationships to the psychological interventions received is still difficult and has been a topic of discussion for decades (e.g., May, 1971).



**FIGURE 2** Framework for the classification of negative effects



Recommendations for the assessment of negative effects in psychotherapy
<b>Use of an instrument with sound underlying conceptual model oriented towards the following criteria:</b>
Incorporating different domains of side effects (e.g., symptomatology, stigmatization, dependence or idealization to therapeutic relationship)
Distinguishing side effects and malpractice/therapeutic misconduct
Measuring degree of burden and evaluating its relevance to treatment outcome (e.g., by using quality of life questionnaires)
Measuring the attribution to psychotherapy
Considering different settings (individual vs. group treatment, outpatient vs. inpatient, face-to-face vs. internet or mobile-based, etc.)
Considering different perspectives (patient, therapist, relative)
Considering different therapeutic orientation (cognitive-behavioral treatment, psychodynamic treatment, etc.)
Inclusion of items measuring positive effects
<b>Further empirical examination of existing measures through different types of studies:</b>
Qualitative studies (interview) to determine the main criteria improving the content validity
Quantitative studies to determine psychometric properties due to different validity aspects (construct, predictive, criterion) and reliability of self- and therapist-rated instruments

**FIGURE 3** Recommendations for the assessment of negative effects

#### 4.2 | Improving the assessment of negative effects

Our analysis has shown several ways to improve the assessment of negative effects. Recommendations are delineated in Figure 3. In summary, the use and development of instruments for assessing negative effects must be based on a strong theoretical background and a sound underlying conceptual model that includes a clear definition and classification of positive and negative effects, and the above-synthesized framework might be a useful tool to comply with this. In particular, on the basis of the results of this review, the following recommendations for evaluating negative effects in psychotherapy can be derived. First, instruments need to take into account different domains of side effects (in particular stigma, symptom change, changes and strains in life areas, dependence, or idealization of the therapeutic relationship). Second, the results highlight the distinction between side effects, malpractice, and unethical conduct. The recommendation, which can be implemented by various instruments, is to use one instrument to assess side effects and another to assess malpractice and/or unethical conduct. It should be noted that the correct assessment of unethical behavior and misconduct, through both self-report and the practitioner's report, is difficult. Thirdly, the instruments need to assess the level of burden to evaluate the relevance (and therefore impact) of side effects and also assess the attribution to psychotherapy; therefore, future studies should consider evaluating the relevance of side effects to determine their effects on treatment and outcome. With the exception of INEP, no other reviewed instrument reported data on important criterion validity aspects. Criterion validity can be measured by scales that assess treatment outcome in terms of specific and general symptom reduction or the patients' quality of life, for example. Various instruments are available to assess symptom reduction and the quality of life (e.g., SF-36: Zwingmann, Metzger, and Jäckel (1998), WHOQOL-BREF: The WHOQOL Group (1998)), which should be used in future studies. Fourthly, when assessing side effects, researchers must consider the setting (individual and/or group treatment, outpatient or inpatient, face-to-face, Internet or mobile-based, etc.), the perspective (patient, therapist, relative), and the therapeutic orientation

(cognitive behavioral treatment, psychodynamic treatment, etc.). Depending on the therapeutic orientation, it can vary whether events are regarded as side effects or part of an effective therapy. For example, CBT therapists may describe a deterioration of symptoms or dependence on the therapist as a side effect rather than a psychoanalyst, who may regard it as a component of the effective therapy. At best, instruments should be developed by scientists and practitioners of different orientations, which has not been done so far with the instruments of this review. Most instruments were developed and validated in the context of CBT, while only the VNIS was developed with a psychodynamic focus. Therefore, some important constructs of psychotherapeutic effects in general might not have been covered by the strong focus on CBT, such as the therapeutic relationship. Future research should therefore conduct studies with a broader therapeutical background by incorporating also mental health services with a focus on, for example, psychoanalytical and psychodynamic therapy as well as so-called "Third Wave" therapies. While it is important for the recording of negative effects that appropriate instruments are used in prospective (descriptive) studies, it also seems to be very useful to develop strategies on how negative effects can be prevented. For example, it could be investigated whether the involvement of significant others in treatment at an early stage and homework to strengthen autonomy can counteract dependence on the therapist. In addition, psychotherapists should be trained in the detection, monitoring, and minimization of possible side effects. Fifth, the positive effects of psychotherapy—in contrast to the concept of positive side effects—should also be taken into account when assessing side effects in order to minimize the potential of the aforementioned nocebo effects.

To realize this recommendation, a suitable assessment tool for measuring side effects of psychotherapy may be embedded in the broader context of the so-called Routine Outcome Monitoring (ROM). ROM has been shown to be effective at reducing treatment failure and likewise enhancing the positive effects of psychotherapy (Lambert & Harmon, 2018; Lambert, Whipple, & Kleinstäuber, 2018). ROM therefore yields some considerable merit for the implementation of evidence-based practice in routine care, and the

assessment of side effects may broaden and enrich current ROM strategies. Moreover, further empirical examination of existing measures through different types of studies is needed, that is, qualitative studies (interview) to determine the main criteria improving the content validity and quantitative studies to determine psychometric and clinimetric properties due to different validity aspects (construct, predictive, and criterion) and reliability of both self- and therapist/observer-rated instruments.

### 4.3 | Limitations of the review

The first limitation of this systematic review concerns the relatively small number of eligible studies and instruments that investigate and assess negative effects. Secondly, this review included all available assessment tools of negative effects (e.g., UE-G is an instrument that only measures the negative effects of group psychotherapy). Since there has been no consensus on negative effects so far, the heterogeneity of the examined instruments may be considered as one limitation of this review. However, the authors adhere to this approach to do an exhaustive search and examine all relevant underlying theoretical foundations in order to extract the diagnostic features and synthesize a comprehensive model. Thirdly, this review did not examine the clinimetric properties of each instrument (Bech, 2016), which might be especially important in terms of research on psychological interventions (Fava, Rafanelli, & Tomba, 2012). Fourthly, in psychotherapy outcome research, the reliable change index (RCI) is used extensively for defining deterioration using standardized rating scales (Jacobson, Follette, & Revenstorf, 1984). The RCI has not been considered in this review because the scope of this review was to study assessment tools of negative effects during the course of therapy or after completion. In general, negative effects were considered more as a process variable than an outcome variable. Within this framework, deterioration as one potential side effect might not be seen as an outcome, more as a transient and short-term effect that may occur during the therapy process. Finally, the lack of variability of the patients participating in these studies could be another limiting factor narrowing to some extent the use of such instruments across highly heterogeneous mental health issues. For example, patients with more severe psychiatric disorders (such as personality disorder or schizophrenia) may experience more serious side effects than patients with less severe disorders (such as mild depression and no comorbidities). In this context, first studies suggest that inpatients who are usually more severely ill report more side effects than outpatients (see Brakemeier et al., 2018; Rheker et al., 2017). Future studies should therefore specifically include severely ill patient groups in order to identify specific negative effects and compare them between different patient groups.

### 4.4 | Future research directions

There are several future directions for improving the assessment of negative effects of psychotherapy. First, existing instruments need to be evaluated with regard to their psychometric properties (see

Figure 3). Of note, psychometric research was mainly developed outside the clinical field and although psychometrics has been used successfully in clinical psychology research and has led to some advances in evaluation, it has guided research to rely strongly on its advantages and to neglect its disadvantages. Thus, when developing new assessment tools in future research the clinimetric properties should be considered (Bech, 2016; Fava et al., 2012; Feinstein, 1987): those features of an instrument that identify clinically relevant changes in mental health over time (discrimination properties such as responsiveness/sensitivity; Fava, Tomba, & Bech, 2017) and predict long-term incremental validity within the clinical decision-making process (Fava et al., 2012). In the case of negative effects, besides the evaluation of psychometric properties, they should be linked to treatment outcome in order to determine the impact (relevance) of treatment on the individual patient's life. For example, within process-outcome research, future studies could link the occurrence of negative effects to treatment outcome, for example, by using the RCI (Jacobson & Truax, 1991). Initial attempts have been made to address the relevance of negative effects on treatment in an inpatient cognitive behavioral analysis system of psychotherapy (CBASP) sample (Brakemeier et al., 2018). In line with this, the current methodological recommendations for trials of psychological interventions support the usefulness of clinimetrics (Guidi et al., 2018). Second, there is considerable need to develop new instruments for assessing negative effects in specific populations (e.g., children and adolescents) and for different settings (e.g., short forms and specific items for group therapy and inpatient use). Third, most instruments are self-rated; thus, validated clinician-rated instruments would be valuable to provide therapists with a standardized tool to monitor negative effects during treatment. A promising approach is the UE-ATR, which should be validated in future studies. Fourth, longitudinal research designs could provide insights into the predictive validity of instruments (including clinimetric, discriminant, and incremental validity) as well as to improve our understanding of the influence of negative effects on treatment outcome (i.e., response, remission, relapse, and dropout) in order to determine the relevance of the negative effects (cf. Brakemeier et al., 2018). The prevalence of negative effects seems to vary widely from study to study, depending on the selection of instrument (Ladwig et al., 2014; Moritz et al., 2015, 2018; Rheker et al., 2017); therefore, an instrument that is recognized worldwide as the "gold standard" is desirable for use in most studies in order to make study results comparable. In addition, current methodological guidelines for trials plead for the assessment of negative effects of psychotherapy using suitable evaluation methods (Guidi et al., 2018). In order to monitor and counteract negative effects, in particular side effects, further studies need to develop a process scale that assesses negative effects during therapy and with a clear time frame. This would strengthen the clinimetric properties and thus clinical usefulness of an instrument in clinical practice. Further, such a process scale that regularly assesses side effects of psychological interventions may be a useful extension for ROM (Lambert & Harmon, 2018). The aim should always be to carry out effective psychotherapies with as few side effects as possible.

## ACKNOWLEDGMENTS

None.

## CONFLICT OF INTEREST

The authors declared no conflicts of interest.

## AUTHORS' CONTRIBUTION

All authors have been significantly involved in the research and/or article preparation. All authors have approved the final article.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

Philipp Herzog  <https://orcid.org/0000-0002-8267-2083>

## REFERENCES

- APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63(9), 839–851. <https://doi.org/10.1037/0003-066X.63.9.839>
- Bech, P. (2016). Clinimetric dilemmas in outcome scales for mental disorders. *Psychotherapy and Psychosomatics*, 85(6), 323–326. <https://doi.org/10.1159/000448810>
- Boettcher, J., Rozentel, A., Andersson, G., & Carlbring, P. (2014). Side effects in Internet-based interventions for social anxiety disorder. *Internet Interventions*, 1(1), 3–11. <https://doi.org/10.1016/j.invent.2014.02.002>
- Bootzin, R. R., & Bailey, E. T. (2005). Understanding placebo, nocebo, and iatrogenic treatment effects. *Journal of Clinical Psychology*, 61(7), 871–880. <https://doi.org/10.1002/jclp.20131>
- Brakemeier, E.-L., Herzog, P., Radtke, M., Schneibel, R., Breger, V., Becker, M., ... Normann, C. (2018). CBASP als stationäres Behandlungskonzept der therapieresistenten chronischen depression: Eine Pilotstudie zum Zusammenhang von Nebenwirkungen und Therapieerfolg. *PPmP - Psychotherapie Psychosomatik Medizinische Psychologie*, 68(09/10), 399–407. <https://doi.org/10.1055/a-0629-7802>
- Burlingame, G. M., MacKenzie, K. R., & Strauß, B. (2004). Small group treatment: Evidence for effectiveness and mechanisms of change. In M. J. Lambert (Ed.), *Bergin & Garfield's Handbook of psychotherapy and behavior change* (pp. 647–696). New York: Wiley & Sons.
- Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review*, 26(1), 17–31. <https://doi.org/10.1016/j.cpr.2005.07.003>
- Caspar, F., & Jacobi, F. (2007). Psychotherapieforschung. In W. Hiller, E. Leibing, F. Leichsenring, & S. K. D. Sulz (Eds.), *Lehrbuch der Psychotherapie für die Ausbildung zur/zum Psychologischen Psychotherapeuten und für ärztliche Weiterbildung, Band 1: Wissenschaftliche Grundlagen der Psychotherapie* (pp. 395–410). München, Germany: CIP-Medien.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Cohen, S. (2014). The nocebo effect of informed consent. *Bioethics*, 28(3), 147–154. <https://doi.org/10.1111/j.1467-8519.2012.01983.x>
- Colloca, L., & Miller, F. G. (2011). The nocebo effect and its relevance for clinical practice. *Psychosomatic Medicine*, 73(7), 598–603. <https://doi.org/10.1097/PSY.0b013e3182294a50>
- Crawford, M. J., Thana, L., Farquharson, L., Palmer, L., Hancock, E., Bassett, P., ... Parry, G. D. (2016). Patient experience of negative effects of psychological treatment: Results of a national survey. *The British Journal of Psychiatry*, 208(3), 260–265. <https://doi.org/10.1192/bjp.bp.114.162628>
- Cuijpers, P. (2016). *Meta-analyses in mental health research. A practical guide*. Amsterdam, the Netherlands: Vrije Universiteit.
- David, D., Cristea, I., & Hofmann, S. G. (2018). Why cognitive behavioral therapy is the current gold standard of psychotherapy. *Frontiers in Psychiatry*, 9, 4. <https://doi.org/10.3389/fpsy.2018.00004>
- Epstein, R. S., & Simon, R. I. (1990). The exploitation index: An early warning indicator of boundary violations in psychotherapy. *Bulletin of the Menninger Clinic*, 54(4), 450–465.
- Fava, G. A., Rafanelli, C., & Tomba, E. (2012). The clinical process in psychiatry. *The Journal of Clinical Psychiatry*, 73(02), 177–184. <https://doi.org/10.4088/JCP.10r06444>
- Fava, G. A., Tomba, E., & Bech, P. (2017). Clinical pharmacopsychology: Conceptual foundations and emerging tasks. *Psychotherapy and Psychosomatics*, 86(3), 134–140. <https://doi.org/10.1159/000458458>
- Feinstein, A. R. (1987). *Clinimetrics*. Retrieved from [https://books.google.de/books/about/Clinimetrics.html?xml:id=P6iDQgAACAJ&redir\\_esc=y](https://books.google.de/books/about/Clinimetrics.html?xml:id=P6iDQgAACAJ&redir_esc=y)
- Groth-Marnat, G. (2009). *Handbook of psychological assessment*. Hoboken, NJ: John Wiley & Sons Inc.
- Guidi, J., Brakemeier, E.-L., Bockting, C. L. H., Cosci, F., Cuijpers, P., Jarrett, R. B., ... Fava, G. A. (2018). Methodological recommendations for trials of psychological interventions. *Psychotherapy and Psychosomatics*, 87(5), 276–284. <https://doi.org/10.1159/000490574>
- Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy Research*, 36(5), 427–440. <https://doi.org/10.1007/s10608-012-9476-1>
- Hoyer, J. (2016). "Positive Nebenwirkungen" von Psychotherapie. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 45(3), 163–173. <https://doi.org/10.1026/1616-3443/a000370>
- Huhn, M., Tardy, M., Spineli, L. M., Kissling, W., Förstl, H., Pitschel-Walz, G., ... Leucht, S. (2014). Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders. *JAMA Psychiatry*, 71(6), 706. <https://doi.org/10.1001/jamapsychiatry.2014.112>
- Ioannidis, J. P. A., Evans, S. J. W., Gotszke, P. C., O'Neill, R. T., Altman, D. G., Schulz, K., & Moher, D. (2004). Better reporting of harms in randomized trials: An extension of the CONSORT statement. *Annals of Internal Medicine*, 141(10), 781–788. <https://doi.org/10.7326/0003-4819-141-10-200411160-00009>
- Jacobson, N. S., Follette, W. C., & Revenstorff, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15(4), 336–352. [https://doi.org/10.1016/S0005-7894\(84\)80002-7](https://doi.org/10.1016/S0005-7894(84)80002-7)
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12–19. <https://doi.org/10.1037/0022-006X.59.1.12>
- Jonsson, U., Alaie, I., Parling, T., & Arnborg, F. K. (2014). Reporting of harms in randomized controlled trials of psychological interventions for mental and behavioral disorders: A review of current practice. *Contemporary Clinical Trials*, 38(1), 1–8. <https://doi.org/10.1016/j.cct.2014.02.005>

- Ladwig, I., Rief, W., & Nestoriuc, Y. (2014). Welche Risiken und Nebenwirkungen hat Psychotherapie? – Entwicklung des Inventars zur Erfassung Negativer Effekte von Psychotherapie (INEP). *Verhaltenstherapie*, 24, 252–263. <https://doi.org/10.1159/000367928>
- Lambert, M. J., & Harmon, K. L. (2018). The merits of implementing routine outcome monitoring in clinical practice. *Clinical Psychology: Science and Practice*, 25(4), e12268. <https://doi.org/10.1111/cpsp.12268>
- Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*, 55(4), 520–537. <https://doi.org/10.1037/pst0000167>
- Linden, M. (2013). How to define, find and classify side effects in psychotherapy: From unwanted events to adverse treatment reactions. *Clinical Psychology and Psychotherapy*, 20(4), 286–296. <https://doi.org/10.1002/cpp.1765>
- Linden, M., Strauß, B., Scholten, S., Nestoriuc, Y., Brakemeier, E.-L., & Wasilewski, J. (2018). Definition and decision-making in the determination and detection of side effects of psychotherapy. *PPmP - Psychotherapie Psychosomatik Medizinische Psychologie*, 68(09/10), 377–382.
- Linden, M., Walter, M., Fritz, K., & Muschalla, B. (2015). Unerwünschte Therapiewirkungen bei verhaltenstherapeutischer Gruppentherapie: Häufigkeit und Spektrum. *Nervenarzt*, 86(11), 1371–1382. <https://doi.org/10.1007/s00115-015-4297-6>
- May, P. R. (1971). For better or for worse? Psychotherapy and variance change: A critical review of the literature. *The Journal of Nervous and Mental Disease*, 152(3), 184–192. <https://doi.org/10.1097/00005053-197103000-00004>
- Meister, R., von Wolff, A., Mohr, H., Nestoriuc, Y., Härter, M., Hölzel, L., & Kriston, L. (2016). Adverse event methods were heterogeneous and insufficiently reported in randomized trials on persistent depressive disorder. *Journal of Clinical Epidemiology*, 71, 97–108. <https://doi.org/10.1016/j.jclinepi.2015.10.007>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. & PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Moritz, S., Fieker, M., Hottenrott, B., Seeralan, T., Cludius, B., Kolbeck, K., ... Nestoriuc, Y. (2015). No pain, no gain? Adverse effects of psychotherapy in obsessive-compulsive disorder and its relationship to treatment gains. *Journal of Obsessive-Compulsive and Related Disorders*, 5, 61–66. <https://doi.org/10.1016/j.jocrd.2015.02.002>
- Moritz, S., Nestoriuc, Y., Rief, W., Klein, J. P., Jelinek, L., & Peth, J. (2018). It can't hurt, right? Adverse effects of psychotherapy in patients with depression. *European Archives of Psychiatry and Clinical Neuroscience*, 269(5), 577–586. <https://doi.org/10.1007/s00406-018-0931-1>
- Parker, G., Fletcher, K., Berk, M., & Paterson, A. (2013). Development of a measure quantifying adverse psychotherapeutic ingredients: The experiences of therapy questionnaire (ETQ). *Psychiatry Research*, 206(2–3), 293–301. <https://doi.org/10.1016/j.psychres.2012.11.026>
- Parker, G., Paterson, A., Fletcher, K., McClure, G., & Berk, M. (2014). Construct validity of the Experiences of Therapy Questionnaire (ETQ). *BMC psychiatry*, 14(1), 369.
- Parry, G. D., Crawford, M. J., & Duggan, C. (2016). Iatrogenic harm from psychological therapies-time to move on. *British Journal of Psychiatry*, 208(3), 210–212.
- Peth, J., Jelinek, L., Nestoriuc, Y., & Moritz, S. (2018). Unerwünschte Effekte von Psychotherapie bei depressiven Patienten-Erste Anwendung der Positive and Negative Effects of Psychotherapy Scale (PANEPS). *PPmP-Psychotherapie- Psychosomatik- Medizinische Psychologie*, 68(9-10), 391–398.
- Rheker, J., Beisel, S., Kräling, S., & Rief, W. (2017). Rate and predictors of negative effects of psychotherapy in psychiatric and psychosomatic inpatients. *Psychiatry Research*, 254, 143–150. <https://doi.org/10.1016/j.psychres.2017.04.042>
- Rief, W., Glombiewski, J. A., Gollwitzer, M., Schubö, A., Schwarting, R., & Thorwart, A. (2015). Expectancies as core features of mental disorders. *Current Opinion in Psychiatry*, 28(5), 378–385. <https://doi.org/10.1097/YCO.0000000000000184>
- Roback, H. B. (2000). Adverse outcomes in group psychotherapy: Risk factors, prevention, and research directions. *The Journal of psychotherapy practice and research*, 9(3), 113.
- Rozental, A., Kottorp, A., Boettcher, J., Andersson, G., & Carlbring, P. (2016). Negative effects of psychological treatments: An exploratory factor analysis of the negative effects questionnaire for monitoring and reporting adverse and unwanted events. *PLoS ONE*, 11(6), 1–22. <https://doi.org/10.1371/journal.pone.0157503>
- Schefflt, C., Guhn, A., Brakemeier, E.-L., Sterzer, P., & Köhler, S. (2019). Efficacy of inpatient psychotherapy for major depressive disorder: A meta-analysis of controlled trials. *Acta Psychiatrica Scandinavica*, 139(4), 322–335. <https://doi.org/10.1111/acps.12995>
- Schneibel, R., Wilbertz, G., Scholz, C., Becker, M., Brakemeier, E.-L., Bschor, T., ... Schmol, D. (2017). Adverse events of group psychotherapy in the in-patient setting - Results of a naturalistic trial. *Acta Psychiatrica Scandinavica*, 136(3), 247–258. <https://doi.org/10.1111/acps.12747>
- Scott, J. (2017). The science and practice of psychotherapy: The critical need for quality assurance and vigilance to monitor the ratio of benefits to risks of therapies. *Acta Psychiatrica Scandinavica*, 136(3), 233–235. <https://doi.org/10.1111/acps.12780>
- Suh, C. S., Strupp, H. H., & O'Malley, S. S. (1986). The Vanderbilt process measures: The psychotherapy process scale (VPPS) and the negative indicators scale (VNIS). In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 285–323). New York, NY and London, UK: Guilford Press.
- Strauß, B., Strupp, H. H., Burgmeier-Lohse, M., Wille, H., & Storm, S. (1992). Deutschsprachige Version der Vanderbilt-Psychotherapie-Skalen: Beschreibung und Anwendung in zwei Kurztherapien. *Zeitschrift für Klinische Psychologie, Psychopathologie und Psychotherapie*, 40(4), 411–430.
- Szapocznik, J., & Prado, G. (2007). Negative effects on family functioning from psychosocial treatments: A recommendation for expanded safety monitoring. *Journal of Family Psychology*, 21(3), 468–478. <https://doi.org/10.1037/0893-3200.21.3.468>
- The WHOQOL Group (1998). Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychological Medicine*, 28(3), 551–558.
- Webster, R. K., Weinman, J., & Rubin, G. J. (2016). A systematic review of factors that contribute to nocebo effects. *Health Psychology*, 35(12), 1334–1355. <https://doi.org/10.1037/hea0000416>
- Zwingmann, C., Metzger, D., & Jäckel, W. H. (1998). Short form-36 health survey (SF-36): Psychometrische analysen der deutschen version bei rehabilitanden mit chronischen rückenenschmerzen. *Diagnostica*, 44(4), 209–219.

**How to cite this article:** Herzog P, Lauff S, Rief W, Brakemeier E-L. Assessing the unwanted: A systematic review of instruments used to assess negative effects of psychotherapy. *Brain Behav*. 2019;00:e01447. <https://doi.org/10.1002/brb3.1447>

## Anhang F: Studie 6



1                            **Negative effects of an inpatient CBASP program:**  
 2                            **Frequencies and their impact on treatment outcome in chronic**  
 3                            **depression**

4

5 **Philipp Herzog<sup>1</sup>, Sophia Häusler<sup>1</sup>, Claus Normann<sup>2</sup>, Eva-Lotta Brakemeier<sup>1,3,4\*</sup>**

6

7 <sup>1</sup> Department of Clinical Psychology and Psychotherapy, Philipps-Universität Marburg, Marburg,  
8 Germany9 <sup>2</sup> Department of Psychiatry and Psychotherapy, Universitätsklinikum Freiburg, Freiburg, Germany10 <sup>3</sup> Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig University  
11 Giessen, Marburg, Germany12 <sup>4</sup> Psychosomatic Clinic, Schön Klinik Bad Arolsen, Bad Arolsen, Germany

13

14

15 **\* Correspondence:**

16

17 Prof. Dr. Eva-Lotta Brakemeier

18 Email: [eva-lotta.brakemeier@uni-marburg.de](mailto:eva-lotta.brakemeier@uni-marburg.de)

19

20

21

**Short title:**

22

Negative effects of CBASP

23

24 **Keywords:** Negative effects, inpatient psychotherapy, chronic depression, cognitive behavioral  
25 analysis system of psychotherapy, CBASP, treatment outcome, dependence

26

27 **Number of words:** 4000 (excluding the abstract, section titles, figure and table captions, funding  
28 statement, acknowledgments, and references in the bibliography)29 **Number of figures:** 130 **Number of tables:** 3

**Negative effects of CBASP**

31

**Abstract**

32 **Background:** A growing number of studies indicate that the Cognitive Behavioral Analysis System  
33 of Psychotherapy (CBASP) is effective in treating chronic depression. However, there is no  
34 systematic research into possible negative effects. Therefore, the objectives of the study were to  
35 investigate the frequency of negative effects of an inpatient CBASP program and their impact on  
36 treatment response.

37 **Methods:** 52 patients with chronic depression and treatment-resistance who completed the structured  
38 12-weeks inpatient CBASP program in an open trial retrospectively completed the Inventory for the  
39 Assessment of Negative Effects of Psychotherapy during follow-up data-collection. Severity of  
40 depressive symptoms was assessed self- and observer-rated at admission, discharge and 6-months  
41 follow-up. Frequencies of negative effects were calculated and binary logistic regression analyses  
42 were conducted to determine the relationship to treatment outcome.

43 **Results:** 92.3% of patients reported having experienced at least one negative effect and 45.2%  
44 indicated dependence on their therapist. Stigmatization and financial concerns as well as  
45 intrapersonal changes were reported by about one third. Only dependence on the therapist negatively  
46 impacted treatment outcome in both outcome measures.

47 **Conclusions:** Most of the reported negative effects appear to be benign. However, since dependence  
48 on the therapist seems to have a negative impact on treatment outcome, therapists should consciously  
49 address this during therapy.

**Negative effects of CBASP**

50

**Introduction**

51 Systematic research to assess and report negative effects such as side effects and other unwanted  
52 effects of psychotherapy is lacking (1). By definition, negative effects of psychotherapy can be  
53 divided into unwanted effects caused by malpractice or unethical behaviour and unwanted effects  
54 caused by correct treatment (referred to as side or adverse effects) (2). The relevance of negative  
55 effects depends on the severity and duration and should therefore be considered in relation to the  
56 short- and long-term treatment outcome: If the negative effect is negatively related to treatment  
57 outcome, it can be described as relevant; if there is no or a positive association to treatment outcome,  
58 it can be labelled as benign (2).

59 Over the past ten years, several instruments for assessing negative effects of psychotherapy have  
60 been developed and partially validated, notably the Inventory for the Assessment of Negative Effects  
61 of Psychotherapy (INEP; 3) and the Negative Effect Questionnaire (NEQ; 4). According to the INEP,  
62 a recent study reveals that 58.7% of patients from a psychiatric hospital and 45.2% of patients from a  
63 psychosomatic hospital reported to have experienced at least one negative effect during therapy (5).  
64 Another INEP study indicates that 93.8% of former psychotherapy patients reported having  
65 experienced at least one negative effect during or after psychotherapy, with the highest rates  
66 concerning intrapersonal changes, stigmatization and relationships (3). In a recent inpatient study,  
67 which did not use INEP to measure negative effects, 60-65% of psychiatric inpatients reported  
68 deterioration of mood state and unwanted treatment reactions; unwanted treatment reactions  
69 decreased in the course of treatment but were negatively associated with the treatment outcome (6).  
70 In addition, first research data indicates that negative effects have a negative impact on the outcome  
71 of treatment for obsessive-compulsive disorder (7). Overall, knowledge about the occurrence of  
72 specific negative effects in different treatment settings and their effects on treatment outcomes is too  
73 limited to determine the relevance of the negative effects. However, these reported high frequencies  
74 of negative effects of psychotherapy in different treatment settings and mental disorders underline the  
75 importance of further investigations of negative effects, especially in seriously burdened patients like  
76 those suffering from treatment-resistant chronic depression (CD).

77 The Cognitive Behavioural Analysis System of Psychotherapy (CBASP; 8, 9), is a disorder-specific  
78 treatment for patients with CD. Since chronically depressed patients have often experienced  
79 childhood maltreatment (10), the main goal of therapy is to enable patients to experience healing  
80 relationships. Through disciplined personal involvement, the therapist discloses her/his positive and  
81 negative personal feelings and reactions that the patient triggers in her/him, to teach the patient that  
82 people today respond to him differently than she/he expected or feared, supported by interpersonal  
83 discrimination exercises (11). CBASP can therefore be described as an interpersonal learning  
84 therapy. Several research studies indicate the efficacy of CBASP as an outpatient treatment for CD  
85 (12, 13) and CBASP as an inpatient treatment program (14). However, in addition to these positive  
86 effects, it appears important to investigate also negative effects of CBASP. Preliminary results of a  
87 self-constructed, non-validated questionnaire to assess side effects of an inpatient CBASP program  
88 provided some interesting findings; however, the interpretation and generalizability of these results is  
89 hampered by methodological limitations of the questionnaire (15). Thus, research data with validated  
90 questionnaires (such as INEP) for specific negative effects during CBASP in CD and their relation to  
91 treatment response are lacking. In addition, it is of high clinical interest to further investigate negative  
92 effects in inpatient treatment, as inpatient treatment might trigger specific negative effects due to its  
93 short but intensive treatment (6).

**Negative effects of CBASP**

94 Therefore, the objectives of the current study were to exploratively investigate (1) the frequencies of  
95 negative effects of an inpatient CBASP treatment program and (2) the impact of specific negative  
96 effects on the clinician- and self-rated treatment response in order to determine their relevance.



**Negative effects of CBASP**97 **Materials and Methods**

98 The study was conducted at the Affective Research Unit of the Department of Psychiatry and  
 99 Psychotherapy, University of Freiburg Medical School, and approved by the Ethics Committee of the  
 100 University of Freiburg. It has been performed in accordance with the ethical standards laid down in  
 101 the 1964 Declaration of Helsinki and its later amendments. All patients gave their written informed  
 102 consent prior to their inclusion in the study.

103 The present study is part of a larger research project of which the feasibility and outcome data have  
 104 already been published (14, 16).

105 **Patients**

106 70 consecutive patients who met the inclusion of suffering from CD according to DSM-IV, age 18-70  
 107 years, speaking fluently German, and being resistant to outpatient treatment, were enrolled in the  
 108 CBASP inpatient program. Treatment resistance was defined as fulfilling the criterion for either  
 109 medication resistance (no response to two or more adequate trials of antidepressants, according to  
 110 (17)) and/or psychotherapy resistance (no response to at least two health-insurance-reimbursed  
 111 psychotherapies with at least 22 sessions each). Exclusion criteria were defined as a history of bipolar  
 112 I disorder, comorbid substance dependency with less than three months of abstinence, antisocial  
 113 personality disorder, severe forms of autism, and mental disorders due to organic factors according to  
 114 the DSM-IV criteria. 65 of the 70 patients completed the study (drop-out rate 7.1%; reasons for  
 115 dropout: serious conflicts with other patients (3 patients), severe psychosocial problems impossible to  
 116 handle due to the distance to the patient's hometown (1 patient), and diagnosis of mild cognitive  
 117 impairment after 3 weeks of treatment (1 patient) (comp. 14). 52 patients of these completers  
 118 (retention rate 80 %) filled in the INEP (3) between 6 and 12 months after discharge. No significant  
 119 differences were found between the patients who filled in the INEP and those who did not ( $p > .05$ ).  
 120 The drop-out patients refused to fill-out follow-up questionnaires. The relatively long period between  
 121 6 and 12 months is due to the fact that the study idea on this research question arose during the  
 122 conduct of the study (see limitations). In this manuscript, we analyse the data of the 52 patients who  
 123 completed INEP.

124 **Study treatment: CBASP inpatient program**

125 Based on the CBASP treatment developed by McCullough (8, 9), the CBASP concept was modified  
 126 and manualized for inpatient use (18). The structured inpatient treatment program was limited to 12  
 127 weeks and included individual CBASP sessions, group CBASP sessions, physiotherapy with Kiesler  
 128 Circle Training, group occupational therapy, nurse-patient encounters, and social worker contact for  
 129 each patient. In addition to this program, patients received algorithm-based pharmacotherapy  
 130 according to current national and international guidelines for the treatment of depression (19) and  
 131 according to clinical expert supervision. Further details on the CBASP inpatient program are  
 132 provided in (14, 18).

133 **Measures**134 **Inventory for the Assessment of Negative effects of Psychotherapy (INEP)**

135 INEP is a self-report questionnaire assessing the negative effects of psychotherapy after completion  
 136 of psychotherapy (3). The 21-item scale covers seven domains of negative effects: Intrapersonal  
 137 changes, dependence on therapist, family, friends, partnership, stigmatization and financial concerns,  
 138 as well as malpractice. Patients were asked to indicate their agreement or disagreement with these

### Negative effects of CBASP

139 statements on a Likert-scale and to indicate whether they attribute this change to psychotherapy or  
 140 other life circumstances. For example, an item that measures dependence on therapist is formulated  
 141 as follows: “During therapy and/or after its completion, it is harder for me to make important  
 142 decisions on my own.”.  
 143 INEP has demonstrated good internal consistency ( $\alpha = .86$ ), while the subscale “malpractice” showed  
 144 only satisfactory internal consistency ( $\alpha = .73$ ). Initial results of factor analysis showed a 7-factor  
 145 solution that supports its construct validity (3). During follow-up data collection, INEP was assessed  
 146 between 6 and 12 months after discharge.

#### 147 **24-item Version of the Hamilton Rating Depression Scale (HRDS-24)**

148 HRDS-24 is the 24-item version of the well-established clinician-rated Hamilton Rating Depression  
 149 Scale assessing the symptom severity of depression and served as primary outcome measure (20).  
 150 Each item is rated from 0 to 2 or 0 to 4, total score is reported as a sum score and ranges from 0 to  
 151 76, while higher sum scores indicate more severe depressive symptoms. HRDS-24 was assessed at  
 152 intake, discharge, and at a 6 months follow-up by blinded and trained raters. A priori, treatment  
 153 response was defined as a decrease in symptom severity of at least 50% in the HRDS-24.

#### 154 **Beck Depression Inventory-II (BDI-II)**

155 BDI-II is an internationally widely used 21-item self-report questionnaire measuring somatic,  
 156 cognitive, and affective symptoms of depression (21). It serves as secondary outcome measure in the  
 157 present study. Scores are ranging from 0 to 63, with higher values indicating more severe depressive  
 158 symptoms. Like the HRDS-24, BDI-II was assessed at intake, discharge, and at a 6-month follow-up.  
 159 Treatment response was a priori defined in the same way as for HRDS-24, i.e. decrease in symptom  
 160 severity of at least 50% of the BDI-II sum score.

#### 161 **Other baseline measures**

162 At the beginning of the study (baseline), sociodemographic questions were asked by a self-report  
 163 questionnaire including age, gender, education level, and marital status. Clinical characteristics were  
 164 also assessed, including diagnosis of CD according to DSM-IV, early onset of depression, age at  
 165 onset, inpatient treatment and psychotherapy in the past, medication and psychotherapy resistance  
 166 and suicide attempts in the past. Finally, axes I and axis II comorbidity were assessed with SCID I  
 167 (22) and SCID II (23).

#### 168 **Statistics/Statistical Analyses**

169 A data screening according to the suggestions of (24) and a test of the assumptions of logistic  
 170 regression were carried out. The data screening showed that between 6.2% and 36.9% of the  
 171 variables used to measure treatment outcome were missing. In BDI-II, 6.2% of the data were missing  
 172 at baseline. After completion of the treatment, 7.7% and six months later 35.4% of the data were  
 173 missing. Regarding HRDS-24, all data were available for measurements at baseline and discharge.  
 174 Six months after discharge, 6.2% of the HRDS-24 data was missing. The Little MCAR test (25) was  
 175 performed to analyze missing values. The results were not significant implicating that missing values  
 176 appeared random. According to (24), missing values were estimated using expectation maximization  
 177 (EM) procedure. To investigate the relationship between negative effects of psychotherapy and  
 178 individual treatment response, a binary logistic regression was calculated using the backward  
 179 stepwise method (the Backward:LR method). The two important assumptions for logistic regression  
 180 (linearity in logistic regression and the absence of multicollinearity) were fulfilled. In order to  
 181 evaluate the contribution of a single predictor to a model, the Wald test was calculated. The

**Negative effects of CBASP**

182 efficiency coefficient  $\text{Exp}(B)$ , also called Odds-Ratio (OR), and its confidence intervals were  
183 calculated to evaluate the effect of the predictor variables. All analyses were conducted using SPSS,  
184 version 21 (IBM Corp., 2012).

## Negative effects of CBASP

### 185 Results

#### 186 Patient characteristics

187 The mean age at baseline of the 52 patients was 48.1 years (SD = 10.1 years), 61.5% were female.  
 188 The mean patient sum score in the HRDS-24 at baseline was 31.3 (SD = 6.4), while the mean BDI-II  
 189 sum score at baseline was 33.6 (SD = 10.5), both indicating severe depression. Further relevant  
 190 sociodemographic and clinical characteristics are depicted in Table 1.

191 Insert Table 1 here.

#### 192 Frequencies of reported negative effects

193 Figure 1 presents the frequencies of the seven INEP domains of negative effects caused by therapy.  
 194 According to INEP, 92.3% reported having experienced at least one negative effect. Regarding the  
 195 different domains, 45.2% reported having experienced dependence on their therapist. Experiences of  
 196 stigmatization and financial concerns were reported by 35.9%, while intrapersonal changes in terms  
 197 of symptom deterioration were experienced by 33.0% of patients. Furthermore, the lowest rates of  
 198 negative effects were reported concerning family (13.5%), friends (13.5%), and partnership (17.2%).  
 199 Some patients (6.1%) reported malpractice, with this comparatively high figure resulting from items  
 200 stating that their therapist forced them to do things they did not want to do (such as role-playing) (2  
 201 patients who partly agreed, 5 patients who agreed somewhat), and that patients felt hurt by therapists  
 202 statements (1 patient who totally agreed, 1 who partly agreed, 8 who agreed somewhat). No patient  
 203 reported sexual abuse, physical assault or other misconduct.

204 Insert Figure 1 here.

#### 205 Prediction of treatment response measured by HRDS-24

206 According to the HRDS-24 criterion for treatment response, 46 of the 52 patients (88.5 %) responded  
 207 to the 12-week treatment program. Six months after discharge, 32 patients (61.5%) still reached the  
 208 response criterion. The results of the last step of the binary logistic regression using the backward  
 209 stepwise method for treatment response regarding HRDS-24 are depicted in Table 2. The factor  
 210 “Intrapersonal changes” appears negatively related to treatment response at post-treatment ( $b = -.36$ ,  
 211  $\text{Wald } \chi^2(1) = 5.05$ ,  $p = .03$ ). The Exp(B) value indicates that when “Intrapersonal changes” are  
 212 increased by one unit, the odds ratio is .70 times as large and therefore patients are 30% less likely to  
 213 respond to treatment. However, after 6-months of follow-up, the factor “Dependence” is significantly  
 214 associated with treatment response ( $b = -1.02$ ,  $\text{Wald } \chi^2(1) = 5.08$ ,  $p = .02$ ). The Exp(B) value  
 215 indicates that when “Dependence” is increased by one unit, the odds ratio is .36 times as large and  
 216 therefore patients are 64% less likely to respond to treatment. Other factors were not significantly  
 217 associated with treatment response ( $p > .05$ ).

218 Insert Table 2 here.

#### 219 Secondary analysis: Prediction of treatment response measured by BDI-II

220 According to the BDI-II criterion for treatment response, 27 out of the 52 patients (51.9%) met the  
 221 response criterion at discharge. Six months after discharge, 18 (34.6%) patients still met the response  
 222 criterion. Table 3 displays the results of the last step of the binary logistic regression using the  
 223 backward stepwise method for treatment response regarding BDI-II. The factor “Dependence” ( $b = -$   
 224  $.81$ ,  $\text{Wald } \chi^2(1) = 3.91$ ,  $p < .05$ ) appears to be negatively associated with treatment response at post-  
 225 treatment, while the Exp(B) value indicates that when “Dependence” is increased by one unit, the

**Negative effects of CBASP**

226 odds ratio is .51 times as large and the patients are 49% less likely to respond to treatment. After 6-  
227 months of follow-up assessment, the factor “Friends” appears negatively related to treatment  
228 response ( $b = -.69$ , Wald  $\chi^2(1) = 4.87$ ,  $p = .03$ ). Others factors were not significantly associated with  
229 treatment response ( $p > .05$ ).

230 Insert Table 3 here.

231

## Negative effects of CBASP

232

## Discussion

233 To date, there are only few published studies investigating specific negative effects and their impact  
234 on the outcome of different specific psychotherapies (exception e.g. 7). A better understanding of the  
235 frequency and relevance of negative effects is also relevant to adequately inform the patient about  
236 possible risks of the treatment. This study therefore aimed at 1) assessing the negative effects of a  
237 CBASP inpatient program, as measured by the established and validated instrument INEP (3), and 2)  
238 evaluating the impact on treatment response in order to assess the relevance of negative effects. To  
239 achieve the first objective, the reported frequencies of negative effects of the inpatient CBASP  
240 treatment were examined. Notably, over 90 percent of patients reported retrospectively to have  
241 experienced at least one negative effect during treatment. This finding is consistent with previous  
242 studies investigating negative effects in patient populations being treated in outpatient settings (3),  
243 but exceeds reported frequencies in inpatient routine clinical care (5, 6). This percentage is also  
244 higher than in a recent study that also focuses on depressive patients who, however, filled out a  
245 different questionnaire than INEP via the Internet and previously underwent outpatient  
246 psychotherapy (26). Our comparatively high percentage may be explained by the specific  
247 characteristics of patients with treatment-resistant CD, i.e., severe symptoms, early onset (age < 21  
248 years), suicidality, and high percentage of reported childhood maltreatment (10, 27, 28), as well as  
249 the high dosaged short-term inpatient CBASP program with a strong focus on negative relationship  
250 experiences during childhood and the therapists-patient relationship (29). Most frequently in this  
251 study, patients reported having developed a dependence on their therapist (almost half of the  
252 patients). Stigmatization, financial concerns, and intrapersonal changes due to transient symptom  
253 deterioration were reported by one third of all patients (second most frequent). At first glance, the  
254 result that 6.1% of patients reported malpractice appears alarmingly high. A precise analysis of the  
255 items that form this scale, however, shows that this comparatively high percentage is due to two  
256 items stating that patients felt forced by the therapist to do things they did not want to do, and that  
257 patients felt hurt by therapists' statements. In the case of the first item, the patients probably thought  
258 mainly of the interpersonal role-plays, which are intended in the CBASP strategy situational analysis  
259 in group and individual therapies. Chronically depressed patients usually have difficulties performing  
260 the role plays at the beginning of treatment due to their pronounced interpersonal problems (30). In  
261 addition, some patients may initially find statements from therapists as painful, which they make in  
262 the context of disciplined personal involvement. Therapists explain to patients the consequences of  
263 their interpersonally difficult behaviour, which may initially be confrontational. The goal, however,  
264 is long-term healing experiences in the relationship. Accordingly, patients should notice in the course  
265 of treatment how helpful these interpersonal strategies are, which is supported by studies that show  
266 that after CBASP therapies the interpersonal problems have actually decreased (31, 32). Since we  
267 could not find any negative correlation to the treatment outcome, such specific malpractice aspects  
268 appear to be benign. It should be stressed that 0% reported sexual abuse, physical assaults or other  
269 misconduct. Of note, the subscale malpractice of the INEP showed questionable psychometric  
270 properties, e.g. only satisfactory internal consistency (3).

271 Concerning the second objective, the results of the regression analyses suggest that, in particular,  
272 dependence on the therapist, as the most frequent dimension of negative effects, seems to play a  
273 significant role for treatment response on a self- and clinician-rated instrument. While dependence on  
274 the therapist is negatively associated with self-rated treatment response defined by BDI-II at  
275 discharge, the same factor is negatively linked to clinician-rated treatment response by HRDS-24 also  
276 in the long run. Given the limitations of this study (see below), our results cautiously suggest that the  
277 more a patient reports dependence on her/his therapist, the less likely he might benefit from  
278 treatment. Yet, there are many possible explanations for this preliminary finding:

10

### Negative effects of CBASP

- 279 • It appears plausible that the high number of personality disorders (61.4% overall, of which  
280 5.7% were diagnosed with a dependent personality disorder) and personality accentuations  
281 (67.1% overall, of which 32.9% were diagnosed with a dependent personality accentuation)  
282 of our sample may explain the relatively high percentage of patients reporting dependence on  
283 their therapist. Since studies show that personality disorders per se are a negative predictor of  
284 the outcome of psychotherapy in depressed patients (e.g. 33), they might function as the  
285 underlying factor being responsible for the finding that the reported dependence on the  
286 therapist is negatively related to treatment outcome.
- 287 • Notably, it could also be argued that the dependence factor is not a side effect, but simply a  
288 consequence of a poor therapeutic alliance during treatment. Since psychotherapy research  
289 has often confirmed that a positive therapeutic alliance is associated with a positive outcome  
290 (e.g. 34), dependence as an indicator of a negative alliance could explain the worse response.  
291 However, it has recently been reported that care dependency seems to be associated with a  
292 better therapeutic alliance (35).
- 293 • In addition, the level of severity of the personality dimension ‘dependency’ may have a  
294 differential influence on the treatment outcome. Interestingly, a recent study having  
295 investigated the impact on treatment outcome of the personality dimension dependency  
296 (according to 36) in treatment-resistant chronically depressed patients found that patients with  
297 more maladaptive dependent features did not benefit from a long-term psychoanalytic  
298 psychotherapy (LTPP) or treatment as usual (TAU), while those with less maladaptive  
299 dependent features showed considerable gains from LTPP but not from TAU (37).
- 300 • The specific strategies of CBASP might trigger dependence on the therapist. Notably, many  
301 patients suffering from CD reported to have experienced both childhood maltreatment (e.g.  
302 10, 28) and current interpersonal problems such as submissive or hostile behaviour (30, 31) or  
303 emotional and behavioural avoidance (38, 39). The association between childhood  
304 maltreatment and interpersonal problems has recently been reported (40). CBASP-specific  
305 techniques (in particular the disciplined personal involvement and the interpersonal  
306 discrimination exercise) may allow those patients to experience new healing and corrective  
307 relationships – sometimes for the first times in their lives, characterized by predictability,  
308 interpersonal closeness and warmth. Conversely, this new experience could also initially  
309 promote dependence on their therapists, especially when isolated patients have no other  
310 positive significant others in their life.
- 311 • Finally, the applied intensive inpatient CBASP program was limited to 12 weeks. Patients  
312 who have experienced dependence on their therapist may not feel sufficiently prepared yet for  
313 the demands of daily life at the end of this comparatively short treatment period, which may  
314 lead to an unfavourable treatment outcome.

315 Moreover, our results showed that “intrapersonal changes” appear negatively related to treatment  
316 response defined by HRDS-24 at discharge. This result cautiously indicates that the more a patient  
317 has suffered from intrapersonal changes (like transient deterioration of symptoms) during therapy, the  
318 less likely he/she might benefit from the therapy in the short-term, but not in the long-term. Lastly,  
319 the result that negative effects on “friends” is negatively related to long-term treatment response  
320 defined by BDI-II after 6 months, may indicate that the more a patient reports negative effects on  
321 friends caused by treatment the less he/she may improve in terms of treatment response. This result  
322 may be interpreted against the background of theories and approaches that consider CD primarily as a  
323 relationship disorder (8).

**Negative effects of CBASP****324 Limitations**

325 The interpretability of the results of this study is reduced by some limitations. First, the INEP data  
326 were collected retrospectively during a follow-up period and the period between discharge and INEP  
327 survey varied between 6 and 12 months. Thus, memory bias, forgetfulness, retrieval errors, or  
328 important experiences occurring after treatment has been completed may have distorted the reported  
329 negative effects differently. Further, the hindsight bias must be considered when interpreting the  
330 results of this study, i.e., that patients who did not respond see their treatment in a less positive light  
331 and report more negative effects. This retrospective evaluation of negative effects also means that the  
332 short-term outcome was recorded before the evaluation of the negative effects. Therefore, the  
333 analyses should also be interpreted with caution, as the chronological sequence of the recording of  
334 statistical predictors before the variable to be predicted (here: outcome) could not be fulfilled in this  
335 way. It is essential that future studies should record negative effects regularly in the therapy process  
336 and at uniform measurement times. Secondly, INEP does not include any specific negative effects of  
337 an inpatient setting, such as group therapy sessions or conflicts with other patients or staff, nor does it  
338 consider the influence of pharmacotherapy which should be directly addressed in future research.  
339 Moreover, INEP does not simultaneously capture positive effects as the Positive and negative Effects  
340 of Psychotherapy Scale (PANEPS) instrument does (7, 26), which is why this study could not  
341 examine the relationship between negative and positive effects (apart from the outcome). Thirdly,  
342 although there are promising findings underlining the 7-factor structure of INEP (3), these seven  
343 factors still lack some psychometric evaluations. Finally, the lack of a control group, additional  
344 algorithm-based pharmacotherapy and a relatively small sample size generally complicate the ability  
345 to interpret the results. Due to the lack of a control group, we could not rule out that negative effects  
346 could also be due to psychotherapy per se, and not specifically due to the inpatient CBASP treatment.

**347 Conclusions**

348 In this study, inpatient CBASP treatment seems to trigger negative effects, which may be explained  
349 by the specific characteristics of patients with treatment-resistant CD and the focus of CBASP  
350 techniques on the patient-therapist relationship. Interestingly, most reported negative effects do not  
351 appear to have an impact on treatment outcome. However, dependence on the therapist, as the most  
352 frequent dimension of negative effects, seems to be negatively linked to both observer- and self-rated  
353 treatment response. If large randomized controlled trials find that CBASP is more likely to trigger  
354 dependence on the therapist than other psychotherapy concepts, and that this perceived dependence  
355 actually has a negative impact on outcomes, then clinical implications such as prolonging treatment  
356 and focusing more on self-help and autonomy of the patient should be considered.



## Negative effects of CBASP

357

## References

- 358 1. Jonsson, U., Alaie, I., Parling, T., Arnberg, F.K. Reporting of harms in randomized controlled  
359 trials of psychological interventions for mental and behavioral disorders: A review of current  
360 practice. *Contemp. Clin. Trials* **38**(1), 1–8 (2014).  
361 <https://doi.org/10.1016/j.cct.2014.02.005>
- 362 2. Herzog, P., Lauff, S., Rief, W., & Brakemeier, E.-L. (2019). Assessing the unwanted: A  
363 systematic review of instruments used to assess negative effects of psychotherapy. *Brain and*  
364 *Behavior*.  
365 <https://doi.org/10.1002/brb3.1447>
- 366 3. Ladwig, I., Rief, W., Nestoriuc, Y. Welche Risiken und Nebenwirkungen hat Psychotherapie?  
367 – Entwicklung des Inventars zur Erfassung Negativer Effekte von Psychotherapie (INEP).  
368 *Verhaltenstherapie* **24**, 252–263 (2014).  
369 <https://doi.org/10.1159/000367928>
- 370 4. Rozental, A., Kottorp, A., Boettcher, J., Andersson, G., Carlbring, P. The Negative Effects  
371 Questionnaire: psychometric properties of an instrument for assessing negative effects in  
372 psychological treatments. *Behav. Cogn. Psychoth.* **15**, 1–14 (2019).  
373 <https://doi.org/10.1017/S1352465819000018>
- 374 5. Rheker, J., Beisel, S., Kräling, S., Rief, W. Rate and predictors of negative effects of  
375 psychotherapy in psychiatric and psychosomatic inpatients. *Psychiatry Res.* **254**, 143–150  
376 (2017).  
377 <https://doi.org/10.1016/j.psychres.2017.04.042>
- 378 6. Schneibel, R. et al. Adverse events of group psychotherapy in the in-patient setting - results of  
379 a naturalistic trial. *Acta Psychiatr. Scand.* **136**(3), 247–258 (2017).  
380 <https://doi.org/10.1111/acps.12747>
- 381 7. Moritz, S. et al. No pain, no gain? Adverse effects of psychotherapy in obsessive-compulsive  
382 disorder and its relationship to treatment gains. *J. Obsess.-Compuls. Rel.* **5**, 61–66 (2015).  
383 <https://doi.org/10.1016/j.jocrd.2015.02.002>
- 384 8. McCullough, J.P. *Treatment for Chronic Depression: Cognitive Behavioral Analysis System*  
385 *of Psychotherapy (CBASP)*. New York: Guilford Press (2000).
- 386 9. McCullough, J.P., Schramm, E., Penberthy, J.K. *CBASP - Cognitive Behavioral Analysis*  
387 *System of Psychotherapy: Chronische Depressionen effektiv behandeln*. Junfermann Verlag,  
388 Paderborn (2015).
- 389 10. Nelson, J., Klumpp, A., Doebler, P., Ehring, T. Childhood maltreatment and  
390 characteristics of adult depression: Meta-analysis. *Br J Psychiatry* **210**(2), 96–104 (2017).  
391 <http://doi.org/10.1192/bjp.bp.115.180752>
- 392 11. Köhler, S., Guhn, A., Betzler, F., Stiglmayr, C., Brakemeier, E.L., Sterzer, P. Therapeutic  
393 Self-Disclosure within DBT, Schema Therapy, and CBASP: Opportunities and Challenges.  
394 *Front. Psychol.* **8**, 2073 (2017).  
395 <http://doi.org/10.3389/fpsyg.2017.02073>

## Negative effects of CBASP

- 396 12. Negt, P., Brakemeier, E.L., Michalak, J., Winter, L., Bleich, S., Kahl, K.G. The treatment of  
397 chronic depression with cognitive behavioral analysis system of psychotherapy: a systematic  
398 review and meta-analysis of randomized-controlled clinical trials. *Brain. Behav.* **6**(8), e00486  
399 (2016).  
400 <https://doi.org/10.1002/brb3.486>
- 401 13. Schramm, E. et al. Effect of disorder-specific vs nonspecific psychotherapy for chronic  
402 depression. *JAMA Psychiatry* **74**(3), 233 (2017).  
403 <https://doi.org/10.1001/jamapsychiatry.2016.3880>
- 404 14. Brakemeier, E.L. et al. Overcoming treatment resistance in chronic depression: A pilot study  
405 on outcome and feasibility of the Cognitive Behavioral Analysis System of Psychotherapy as  
406 an inpatient treatment program. *Psychother. Psychosom.* **84**(1), 51–56 (2015).  
407 <https://doi.org/10.1159/000369586>
- 408 15. Brakemeier, E.L. et al. CBASP als stationäres Behandlungskonzept der therapieresistenten  
409 chronischen Depression: Eine Pilotstudie zum Zusammenhang von Nebenwirkungen und  
410 Therapieerfolg. *Psychoth. Psychosom. Med. Psychol.* **68**(9–10), 399–407 (2018).  
411 <http://doi.org/10.1055/a-0629-7802>
- 412 16. Brakemeier, E.L. et al. Feasibility and outcome of cognitive behavioral analysis system of  
413 psychotherapy (CBASP) for chronically depressed inpatients: a pilot study. *Psychother.*  
414 *Psychosom.* **80**(3), 191–4 (2011).  
415 <https://doi.org/10.1159/000320779>
- 416 17. Thase, M.E., Rush, A.J. When at first you don't succeed: sequential strategies for  
417 antidepressant nonresponders. *J. Clin. Psychiatr.* **58**(13), 23–9 (1997).
- 418 18. Brakemeier, E.L., Normann, C. *Praxisbuch CBASP. Behandlung chronischer Depression.*  
419 Weinheim, Basel: Beltz (2012).
- 420 19. Middleton, H., Shaw, I., Hull, S., Feder, G. NICE guidelines for the management of  
421 depression. *BMJ*, **330**, 7486 (2005).  
422 <https://doi.org/10.1136/bmj.330.7486.267>
- 423 20. Hamilton, M. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* **23**(1), 56–62  
424 (1960).  
425 <https://doi.org/10.1136/jnnp.23.1.56>
- 426 21. Beck, A.T., Steer, R.A., Brown, G.K. *Beck Depression Inventory-II*. San Antonio, TX:  
427 Harcourt Assessment Inc. (1996).
- 428 22. First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W. Structured clinical interview for  
429 DSM-IV axis I disorders SCID-I: clinician version. *American Psychiatric Publishing* (1997).
- 430 23. First, M.B., Gibbon, M., Spitzer, R.L., Benjamin, L.S. User's Guide for the Structured  
431 Clinical Interview for DSM-IV Axis II Personality Disorders: SCID-II. *American Psychiatric*  
432 *Press* (1997).
- 433 24. Tabachnick, B.G., Fidell, L.S. *Using multivariate statistics*. Boston: Pearson Education  
434 (2013).

## Negative effects of CBASP

- 435 25. Little, R.J.A. A test of missing completely at random for multivariate data with missing  
436 values. *J. Am. Stat. Assoc.* **83**(404), 1198 (1988).
- 437 26. Moritz, S., Nestoriuc, Y., Rief, W., Klein, J.P., Jelinek, L., Peth, J. It can't hurt, right?  
438 Adverse effects of psychotherapy in patients with depression. *Eur. Arch. Psychiatry. Clin.*  
439 *Neurosci.* **269**(5), 577-586 (2018).  
440 <https://doi.org/10.1007/s00406-018-0931-1>.
- 441 27. Sabaß, L. et al. Cognitive Behavioral Analysis System of Psychotherapy as group  
442 psychotherapy for chronically depressed inpatients: a naturalistic multicenter feasibility trial.  
443 *Eur. Arch. Psychiatry Clin. Neurosci.* **268**(8), 783-796 (2018).  
444 <https://doi.org/10.1007/s00406-017-0843-5>
- 445 28. Wiersma, J.E. et al. The importance of childhood trauma and childhood life events for  
446 chronicity of depression in adults. *J. Clin. Psychiatry*, **70**(7), 983–989 (2009).
- 447 29. Köhler, S., Sterzer, P., Normann, C., Berger, M., Brakemeier, E.L. Überwindung der  
448 Therapieresistenz bei chronischer Depression [Overcoming treatment-resistant in chronic  
449 depression: The role of inpatient psychotherapy]. *Nervenarzt*, **87**(7), 701–707 (2016).  
450 <http://doi.org/10.1007/s00115-015-0034-4>
- 451 30. Bird, T., Tarsia, M., Schwannauer, M. Interpersonal styles in major and chronic depression: A  
452 systematic review and meta-analysis. *J. Affect. Disord.* **239**, 93–101 (2018).  
453 <http://doi.org/10.1016/j.jad.2018.05.057>
- 454 31. Constantino, M.J. et al. Interpersonal styles of chronically depressed outpatients: Profiles and  
455 therapeutic change. *Psychotherapy (Chic.)*, **45**(4), 491–506 (2008).  
456 <https://doi.org/10.1037/a0014335>
- 457 32. Constantino, M.J., Adams, M.L., Pazzaglia, A.M., Bernecker, S.L., Ravitz, P., McBride, C.  
458 Baseline patient characteristics as predictors of remission in interpersonal psychotherapy for  
459 depression. *Psychother. Res.* **23**(2), 190–200 (2013).  
460 <http://doi.org/10.1080/10503307.2013.765997>
- 461 33. Markowitz, J.C. et al. Longitudinal effects of personality disorders on psychosocial  
462 functioning of patients with major depressive disorder. *J. Clin. Psychiatry*, **68**(2), 186–193  
463 (2007).
- 464 34. Flückiger, C., Del Re, A.C., Wampold, B.E., Horvath, A.O. The alliance in adult  
465 psychotherapy: A meta-analytic synthesis. *Psychotherapy (Chic.)* **55**(4), 316-340 (2018).  
466 <https://doi.org/10.1037/pst0000172>
- 467 35. Geurtzen, N., Keijsers, G. P. J., Karremans, J. C., & Hutschemaekers, G. J. M. Patients' care  
468 dependency in mental health care: Development of a self-report questionnaire and preliminary  
469 correlates. *J. Clin. Psychol.* **74**(7), 1189–1206 (2018).  
470 <https://doi.org/10.1002/jclp.22574>
- 471 36. Blatt, S.J. *Experiences of depression: Theoretical, clinical, and research perspectives.*  
472 Washington, DC: American Psychological Association (2004).  
473 <http://dx.doi.org/10.1037/10749-000>

## Negative effects of CBASP

- 474 37. Rost, F., Luyten, P., Fearon, P., Fonagy, P. Personality and outcome in individuals with  
475 treatment-resistant depression - Exploring differential treatment effects in the Tavistock Adult  
476 Depression Study (TADS). *J Consult Clin Psychol* **87**(5), 433–445 (2019).  
477 <https://doi.org/10.1037/ccp0000391>
- 478 38. Brockmeyer, T., Kulesa, D., Hautzinger, M., Bents, H., Backenstrass, M. Differentiating  
479 early-onset chronic depression from episodic depression in terms of cognitive-behavioral and  
480 emotional avoidance. *J. Affect. Disord.* **175**, 418–423 (2015).  
481 <http://doi.org/10.1016/j.jad.2015.01.045>
- 482 39. Ley, P., Helbig-Lang, S., Czilwik, S., Lang, T., Worlitz, A., Brücher, K., Petermann, F.  
483 Phenomenological differences between acute and chronic forms of major depression in  
484 inpatients. *Nordic J. Psychiatry* **65**(5), 330–337 (2011).  
485 <http://doi.org/10.3109/08039488.2011.55212>
- 486 40. Klein, J. P. et al. Do interpersonal fears mediate the association between childhood  
487 maltreatment and interpersonal skills deficits? A matched cross-sectional analysis. *Psychoth.*  
488 *Res.* 1–12 (2018) [Epub ahead of print].  
489 <https://doi.org/10.1080/10503307.2018.1532125>

490

491

492

493

494

495

496

497 **Author Contributions:** PH originally drafted and edited the manuscript, ELB supervised, reviewed  
498 and edited the manuscript. SH was significantly involved in the computation of the statistical  
499 analyses and visualization of the results. CN was responsible in the acquisition of this data for the  
500 present work.

501 **Acknowledgements:** We thank all patients who participated in this study and all CBASP therapists  
502 and co-therapists delivering the inpatient treatment.

503 **Conflict of Interest:** The authors declare that the research was conducted in the absence of any  
504 commercial or financial relationships that could be construed as a potential conflict of interest.

505 **Funding:** This research did not receive any specific grant from funding agencies in the public,  
506 commercial, or not-for-profit sectors.

507 **Data Availability Statement:** Data sharing is not applicable to this article as only secondary  
508 analyses were performed.

## Negative effects of CBASP

509 Table 1 Sociodemographic characteristics of the sample (N=52)

Characteristics	
Age at entry <i>M</i> (SD)	48.1 (10.1)
Sex <i>n</i> (%)	
male	20 (38.5)
female	32 (61.5)
Education level <i>n</i> (%)	
No educational degree	4 (7.7)
Primary education	27 (51.9)
Secondary education	3 (5.8)
Higher education	18 (34.6)
Marital status <i>n</i> (%)	
Single	11 (21.2)
Married/couples relationship	29 (55.8)
Divorced/in separation	12 (23.1)
Diagnosis of Chronic Depression (DSM-IV <sup>a</sup> ) <i>n</i> (%)	
Double Depression	24 (46.2)
Recurrent Major Depression	17 (32.7)
Chronic Major Depression	11 (21.2)
Early onset of depression <sup>b</sup> <i>n</i> (%)	42 (80.8)
Age at onset <i>M</i> (SD)	15.0 (10.5)
Comorbid Axis I disorder <sup>c</sup>	43 (61.4)
Comorbid Axis II disorder <sup>d</sup>	47 (67.1)
Inpatient treatment in the past <sup>e</sup>	43 (82.7)
Psychotherapy in the past <sup>f</sup>	50 (96.2)
Medication resistance <sup>g</sup>	46 (88.5)
Psychotherapy resistance <sup>h</sup>	46 (88.5)
Suicide attempt in the past	19 (36.5)
HRDS-24 score at baseline <i>M</i> (SD) <sup>i</sup>	31.3 (6.4)
BDI-II score at baseline <i>M</i> (SD) <sup>j</sup>	33.6 (10.5)

510 *Note.* *M* = mean, *SD* = standard deviation, *n* = number511 <sup>a</sup> DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, 4th Edition512 <sup>b</sup> Before the age of 21513 <sup>c</sup> Assessed with Structured Clinical Interview (SCID) I (22)514 <sup>d</sup> Assessed with Structured Clinical Interview (SCID) II (23)515 <sup>e</sup> Inpatient treatment in a psychiatric or psychosomatic hospital516 <sup>f</sup> Treatments with minimum 22 sessions517 <sup>g</sup> No response to two or more adequate trials of antidepressants518 <sup>h</sup> No response to two or more health insurance-reimbursed psychotherapies with each

519 minimum 22 sessions in Germany

520 <sup>i</sup> HRDS-24 = Hamilton Rating Depression Scale, 24 Items, scale 0-75 (20)521 <sup>j</sup> BDI-II = Beck Depression Inventory-II, 21 Items, scale 0-63 (21)

## Negative effects of CBASP

522 Table 2

523 Results of the logistic regression of negative effects on treatment response at discharge (T2) and  
 524 follow-up 6 months after discharge (T3) measured by Hamilton Rating Depression Scale, 24 item  
 525 version (HRDS-24)

Variables	B	Wald	p	Exp(B)	95.0% C.I. for EXP(B)	
					Lower	Upper
Criterion: Response on HRDS-24 T2						
Family	-1.90	1.93	.17	.15	.01	2.19
Intrapersonal changes	-.36	5.05	.03*	.70	.51	.96
Criterion: Response on HRDS-24 T3						
Dependence	-1.02	5.08	.02*	.36	.15	.88
Malpractice	.56	1.61	.20	1.75	.74	4.17

526 Note. T2 = Assessment at discharge; T3 = Follow up-assessment 6 months after discharge;  
 527 HRDS-24 = 24-item Version of the Hamilton Rating Depression Scale (20)

528 \*p<.05

## Negative effects of CBASP

529 Table 3

530 *Results of the logistic regression of negative effects on treatment response at discharge (T2)*  
 531 *and follow-up 6 months after discharge (T3) measured by Beck Depression Inventory (BDI-II)*

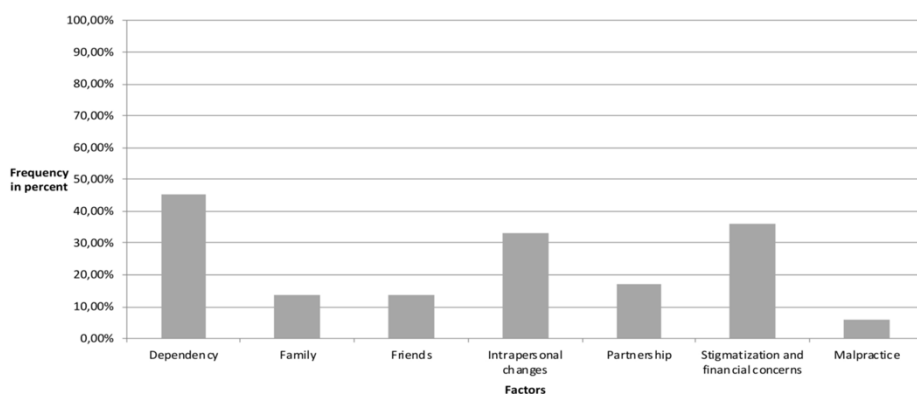
Variables	<i>B</i>	<i>Wald</i>	<i>p</i>	<i>Exp(B)</i>	95.0% C.I. for <i>EXP(B)</i>	
					<i>Lower</i>	<i>Upper</i>
Criterion: Response on <i>BDI-II</i> T2						
Dependence	-.81	3.91	.05*	.45	.20	.99
Criterion: Response on <i>BDI-II</i> T3						
Friends	-.68	4.87	.03*	.51	.28	.93
Malpractice	.62	2.28	.13	1.86	.83	4.15

532 *Note.* T2 = Assessment at discharge; T3 = Follow-up-assessment 6 months after discharge;  
 533 *BDI-II* = Beck Depression Inventory-II (21).

534 \**p*<.05

### Negative effects of CBASP

535 **Figure 1.** Frequencies of reported negative effects of psychotherapy according to the seven factors of  
536 the INEP (3).



### 537 Contribution to the Field Statement

538 Systematic research on assessing and reporting negative effects of psychotherapy in different samples  
539 and settings is lacking. The link of negative effects to treatment outcome seems to be a promising  
540 approach to indicate their relevance. Frequencies of negative effects of an inpatient psychotherapy  
541 for chronic depression (i.e. CBASP program) are high, but in line with findings in other samples and  
542 settings. While almost all patients reported having experienced at least one negative effect, only  
543 dependence on the therapist appears to be relevant for treatment response. When replicated in  
544 randomized controlled trials, therapists should consider a potential dependence and address this issue  
545 during therapy to improve outcome of CBASP.



## Anhang G: Publikationsverzeichnis

## Zeitschriftenartikel (peer-reviewed)

**Herzog, P.** & Jelinek, L. (under review). Die Rolle von Verhaltensexperimenten in der Behandlung von Zwangsstörungen: Ein Update zur effektiven Gestaltung und Durchführung in der Praxis. *Verhaltenstherapie*.

**Herzog, P.**, Feldmann, M., Kube, T., Langs, G., Gärtner, T., Rauh, E., Doerr, R., Hillert, A., Voderholzer, U., Rief, W., Endres, D., & Brakemeier, E.-L. (under review). Inpatient psychotherapy for depression in a large routine psychosomatic care sample: A Bayesian approach to examining effectiveness and predictors of change, *Journal of Consulting and Clinical Psychology*.

**Herzog, P.**, Osen, B., Stierle, C., Middendorf, T., Voderholzer, U., Koch, S., Feldmann, M., Rief, W., & Brakemeier, E.-L. (under review). Determining prognostic variables of treatment outcome in obsessive-compulsive disorder: Effectiveness and its predictors in routine clinical care, *Behavior Therapy*.

**Herzog, P.**, Häusler, S., Normann, C., & Brakemeier, E.-L. (under review). Negative effects of an inpatient CBASP program: Frequencies and their impact on treatment outcome in chronic depression. *Frontiers in Psychiatry (Psychological Therapies)*.

Brakemeier, E.-L., Linden, M., Nestoriuc, Y., Scholten, S., **Herzog, P.\***, & Strauß, B.\* (in revision). Side Effects of Psychotherapy - About the Importance of Considering Side Effects in Psychotherapy (Research). *Psychotherapy and Psychosomatics*.

**Herzog, P.**, Voderholzer, U., Gärtner, T., Osen, B., Svitak, M., Doerr, R., Rolving-Dijkstra, M., Feldmann, M., Rief, W., & Brakemeier, E.-L. (in revision). Predictors of outcome during inpatient psychotherapy for posttraumatic stress disorder: A single-treatment, multi-site, practice-based study, *Psychotherapy Research*.

**Herzog, P.**, Feldmann, M., Voderholzer, U., Gärtner, T., Armbrust, M., Rauh, E., Doerr, R., Rief, W., & Brakemeier, E.-L. (accepted). Drawing the Borderline: Predicting treatment outcomes in patients with borderline personality disorder, *Behavior Research and Therapy*.

- Kube, T., Berg, M., Kleim, B., & **Herzog, P.** (2020). Rethinking post-traumatic stress disorder – A predictive processing perspective. *Neuroscience & Biobehavioral Reviews*, 113(April), 448–460. <https://doi.org/10.1016/j.neubiorev.2020.04.014>
- Herzog, P.**, Lauff, S., Rief, W., & Brakemeier, E.-L. (2019). Assessing the unwanted: A systematic review of instruments used to assess negative effects of psychotherapy. *Brain and Behavior*, 9(12). <https://doi.org/10.1002/brb3.1447>
- Kube, T., **Herzog, P.**, Michalak, C. M., Glombiewski, J. A., Doering, B. K., & Rief, W. (2019). Further Specifying the Cognitive Model of Depression: Situational Expectations and Global Cognitions as Predictors of Depressive Symptoms. *Clinical Psychology in Europe*, 1(4), e33548. <https://doi.org/10.32872/cpe.v1i4.33548>
- Brakemeier, E.-L., **Herzog, P.**, Radtke, M., Schneibel, R., Breger, V., Becker, M., ... Normann, C. (2018). CBASP als stationäres Behandlungskonzept der therapieresistenten chronischen Depression: Eine Pilotstudie zum Zusammenhang von Nebenwirkungen und Therapieerfolg. *PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie*, 68(09/10), 399–407. <https://doi.org/10.1055/a-0629-7802>
- Kube, T., Siebers, V. H. A., **Herzog, P.**, Glombiewski, J. A., Doering, B. K., & Rief, W. (2018). Integrating situation-specific dysfunctional expectations and dispositional optimism into the cognitive model of depression – A path-analytic approach. *Journal of Affective Disorders*, 229, 199-205. <https://doi.org/10.1016/j.jad.2017.12.082>

## Kongressbeiträge

### Symposia

- Herzog, P.**, & Brakemeier, E.-L. (2019). Predicting changes in patients suffering from depression in routine clinical care: A Bayesian approach. Lecture presented at the 9<sup>th</sup> World Congress of Behavioural and Cognitive Therapies 2019, Berlin, Germany.
- Herzog, P.** (2019). Assessing the unwanted: Detecting and monitoring side effects of psychological interventions. Lecture presented at the 9<sup>th</sup> World Congress of Behavioural and Cognitive Therapies 2019, Berlin, Germany.

Brakemeier, E.-L., **Herzog, P.**, Klein, J. P. (2019). Pain or Gain? First results on type and frequency of side effects triggered by (inpatient) CBASP treatments and their impact on outcome. Lecture presented at the 7<sup>th</sup> CBASP Network Meeting of the CBASP-Network, Göttingen, Germany.

**Herzog, P.** (2019). Die Kunst der Vorhersage von individuellen Therapieverläufen bei Zwangsstörungen: Wirksamkeit und deren Prädiktoren in der Routineversorgung. Lecture presented at the 37<sup>th</sup> congress of the German Psychological Association (Clinical Psychology and Psychotherapy Section), Erlangen, Germany.

**Herzog, P.** (2019). Assessing the unwanted: Die Erfassung von Nebenwirkungen in der Psychotherapie. Lecture presented at the 37<sup>th</sup> congress of the German Psychological Association (Clinical Psychology and Psychotherapy Section), Erlangen, Germany.

**Herzog, P.** (2018). Assessing the unwanted: Die Erfassung von Nebenwirkungen in der Psychotherapie. Lecture presented at the Congress 2018 of the German Association for Psychiatry, Psychotherapy and Psychosomatics e. V. (DGPPN), Berlin, Germany.

### Open paper-Vorträge

**Herzog, P.**, Voderholzer, U., Feldmann, M., Rief, W., & Brakemeier, E.-L. (2018). From explanation to prediction: Identifying predictors of treatment outcome in PTSD. Open paper presented at the 48<sup>th</sup> Congress of the European Association of Behavioral and Cognitive Therapies (EABCT), Sofia, Bulgaria.

### Poster

**Herzog, P.**, Voderholzer, U., Feldmann, M., Rief, W., & Brakemeier, E.-L. (2019). The art of predicting treatment outcome in obsessive-compulsive disorder: Effectiveness and its predictors in routine clinical care. Poster presented at the 25<sup>th</sup> World Congress of the International College of Psychosomatic Medicine 2019, Florence, Italy.

Feldmann, M., **Herzog, P.**, Voderholzer, U., Armbrust, M., Rief, W., & Brakemeier, E.-L. (2019). The art of prophecy remains difficult: Predicting treatment outcomes in inpatients with Borderline Personality Disorder. Poster presented at the 9<sup>th</sup> World Congress of Behavioural and Cognitive Therapies 2019, Berlin, Germany.

Feldmann, M., **Herzog, P.**, Voderholzer, U., Armbrust, M., Rief, W., & Brakemeier, E.-L. (2019). Prognosen bleiben schwierig: Zur Vorhersage von Therapieerfolg und Therapieabbruch bei Patienten mit Borderline-Persönlichkeitsstörung in der stationären Routineversorgung. Poster presented at the 37<sup>th</sup> congress of the German Psychological Association (Clinical Psychology and Psychotherapy Section), Erlangen, Germany.

**Herzog, P.**, Voderholzer, U., Feldmann, M., Rief, W., & Brakemeier, E.-L. (2018). Von der Beschreibung zur Vorhersage: Eine Untersuchung von Prädiktoren des Therapieoutcomes bei Patienten mit Posttraumatischer Belastungsstörung in der stationären Routineversorgung. Poster presented at the 36<sup>th</sup> official Congress of the German Psychological Society (Clinical Psychology and Psychotherapy Section), Landau, Germany.

**Herzog, P.**, Kube, T., Siebers, V., Glombiewski, J. A., Brakemeier, E.-L., & Rief, W. (2017). Integrating situation-specific dysfunctional expectations and dispositional optimism into the cognitive model of depression – A path-analytic approach. Poster presented at the 47<sup>th</sup> Congress of the European Association of Behavioral and Cognitive Therapies (EABCT), Ljubljana, Slovenia.

**Herzog, P.**, Nestoriuc, Y., Linden, M., Strauß, B., & Brakemeier, E.-L. (2017). Assessing negative effects of psychotherapy in different settings – Development of a process scale useful to administer during CBASP treatment. Poster presented at the 6<sup>th</sup> CBASP Network meeting of the CBASP-Network, Munich, Germany.

Schamong, I., Bollmann, S., Struck, N., **Herzog, P.**, Künemund, A., Lauff, S., Ehmer, T., Döring, B., Rief, W., Zimmermann, J., Gärtner, T., Brakemeier, E.-L. (2017). Introduction of the new concept and of an RCT trial examining the efficacy of CBASP personalized for inpatients with early trauma and psychotherapy-resistance. Poster presented at the 6<sup>th</sup> CBASP Network meeting of the CBASP-Network, Munich, Germany.

## Anhang H: Eidesstattliche Erklärung

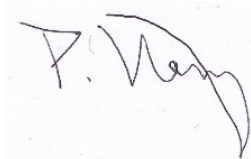
Hiermit versichere ich, meine Dissertation

*“Optimierung der Effekte von Psychotherapie: Wirksamkeit, Prädiktoren und Nebenwirkungen  
in der stationären Routineversorgung”*

selbst und ohne fremde Hilfe verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel genutzt. Alle vollständig oder sinngemäß übernommenen Zitate sind als solche gekennzeichnet.

Die Dissertation wurde weder in der vorliegenden noch in einer ähnlichen Form bei einer anderen in- oder ausländischen Hochschule anlässlich eines Promotionsgesuchs oder zu anderen Prüfungszwecken eingereicht.

Marburg an der Lahn, Juni 2020



---

Dipl.-Psych. Philipp Herzog