

Der Einfluss genereller Veränderungsmechanismen auf das Therapieergebnis in der kognitiven Verhaltenstherapie

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

„Doctor rerum naturalium“

der Georg-August-Universität Göttingen

im Promotionsprogramm Biologie (Basisprogramm)

der Georg-August University School of Science (GAUSS)

vorgelegt von

Sebastian Gmeinwieser, M. Sc.

aus Fritslar

Göttingen, 04.11.2019

Betreuungsausschuss

Prof. Dr. York Hagmayer, Abteilung für Kognitionswissenschaften und Entscheidungspsychologie, Georg-Elias-Müller Institut für Psychologie, Georg-August-Universität Göttingen

Prof. Dr. Timo Brockmeyer, Abteilung für Klinische Psychologie und Psychotherapie, Georg-Elias-Müller Institut für Psychologie, Georg-August-Universität Göttingen

Prof. Dr. Hannes Rakoczy, Abteilung für Kognitive Entwicklungspsychologie, Georg-Elias-Müller Institut für Psychologie, Georg-August-Universität Göttingen

Mitglieder der Prüfungskommission

Referent:

Prof. Dr. York Hagmayer, Abteilung für Kognitionswissenschaften und Entscheidungspsychologie, Georg-Elias-Müller Institut für Psychologie, Georg-August-Universität Göttingen

Korreferent:

Prof. Dr. Timo Brockmeyer, Abteilung für Klinische Psychologie und Psychotherapie, Georg-Elias-Müller Institut für Psychologie, Georg-August-Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Hannes Rakoczy, Abteilung für Kognitive Entwicklungspsychologie, Georg-Elias-Müller Institut für Psychologie, Georg-August-Universität Göttingen

Prof. Dr. Eric Leibing, Klinik für Psychosomatik und Psychotherapie, Georg-August-Universität Göttingen

Prof. Dr. Michael Waldmann, Abteilung für Kognitionswissenschaften und Entscheidungspsychologie, Georg-Elias-Müller Institut für Psychologie, Georg-August-Universität Göttingen

Prof. Dr. Margarethe Boos, Abteilung für Sozial- und Kommunikationspsychologie, Georg-Elias-Müller Institut für Psychologie, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 10.12.2019

Selbstständigkeitserklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die von mir angegebenen Quellen und Hilfsmittel verwendet habe. Wörtlich und sinngemäß aus anderen Werken entnommene Stellen habe ich unter Angabe der Quellen kenntlich gemacht. Die Richtlinien zur Sicherung der guten wissenschaftlichen Praxis an der Universität Göttingen wurden von mir beachtet.

Danksagung

Prof. Dr. York Hagmayer möchte ich besonders danken, der mich bei meiner Dissertation von Anfang an und auch im Verlauf dieser Arbeit am stärksten unterstützt hat. Er stand mir immer verlässlich, mit Rat und Kritik, zur Seite, sodass ich vor allem von seiner wissenschaftlichen und professionellen Arbeitsweise profitieren konnte. Des Weiteren möchte ich Prof. Dr. Timo Brockmeyer danken, der sich dieser Dissertation angenommen hat und zu jeder Zeit eine Bereicherung für meinen Betreuungsausschuss war. Prof. Dr. Thomas Probst möchte ich danken, der mich sehr bei meinen Publikationen unterstützt hat, gerade in der frühen Phase, als sich mein Vorhaben noch entwickelt hat und in der Rohfassung war, hat er mir richtungsweisende Perspektiven aufgezeigt.

Darüber hinaus möchte ich den Kolleginnen und Kollegen der Abteilung für Klinische Psychologie und Psychotherapie sowie dem Therapie- und Beratungszentrum danken. Dr. Ruhl, der mir mein Promotionsvorhaben ermöglicht hat, Cornelia Bernardi und Annika Rudolph die mich beim Aufbereiten der Daten unterstützten.

Mein Dank gebührt außerdem meiner Familie, die stets hinter mir stand und meinen Freunden, die mich auch in stressigen Zeiten auf andere Gedanken bringen konnten.

Abstract - Deutsch

Es gibt zahlreiche Belege dafür, dass Psychotherapie, vor allem die kognitive Verhaltenstherapie, wirksam ist. Welche Bestandteile einer Psychotherapie dagegen genau wirksam sind, wird dagegen kontrovers diskutiert. Forschung zu spezifischen Wirkfaktoren einzelner Therapieformen steht dabei Forschung zu übergreifenden, sogenannten generellen Veränderungsmechanismen gegenüber. Die in dieser Dissertation enthaltenen Publikationen beschäftigen sich mit drei ausgewählten generellen Veränderungsmechanismen: interpersonale Erfahrungen (Therapiebeziehung, Offenheit/Wohlbefinden des Patienten), intrapersonale Erfahrungen (Einsicht und Bewältigung von Problemen) und Problemaktualisierung (emotionale Aktivierung in der Therapie). Untersucht wurden dabei Zusammenhänge dieser generellen Veränderungsmechanismen und Bezug zu verschiedenen Therapieergebnismaßen mit unterschiedlichen statistischen Methoden. Als Stichprobe dienten dabei Patienten aus der Routineversorgung, die eine ambulante kognitive Verhaltenstherapie erhielten (524-911 Patienten je nach Studie).

Die erste Studie untersuchte, ob gerade die frühen Veränderungen der generellen Veränderungsmechanismen mit dem Therapieergebnis assoziiert sind. Das Therapieergebnis wurde dabei als Symptombelastung und als Lebenszufriedenheit operationalisiert und jeweils kontinuierlich und dichotom, anhand des reliable change index betrachtet. Die Ergebnisse zeigen signifikante Zusammenhänge der generellen Veränderungsmechanismen mit dem Therapieergebnis. Dabei ist es wichtig, ob das Therapieergebnis als Symptombelastung oder Lebenszufriedenheit, kontinuierlich oder dichotom oder ob die generellen Veränderungsmechanismen aus Therapeuten- oder aus Patientenperspektive eingeschätzt wurden. In den statistischen Modellen klärten die generellen Veränderungsmechanismen zwischen 2-4 % zusätzliche Varianz auf.

In der zweiten Studie wurde die motivationale Inkongruenz als möglicher Moderator zwischen den frühen generellen Veränderungsmechanismen und dem Therapieergebnis be-

trachtet. Vor Allem wurde dabei auf die intrapersonalen Erfahrungen fokussiert; interpersonale Erfahrungen und Problemaktualisierung wurden dagegen als exploratorische Hypothesen zusätzlich untersucht. Motivationale Inkongruenz entsteht vor allem, wenn bestimmte Grundbedürfnisse eines Menschen, wie etwa das Bedürfnis nach Bindung, im Alltag nicht umgesetzt werden kann. Vergangene Studien konnten zeigen, dass gerade Patienten mit psychischen Belastungen unter einer hohen Inkongruenz leiden. Die Ergebnisse zeigen eine signifikante Interaktion der frühen patienteneingeschätzten intrapersonalen Erfahrungen mit der Inkongruenz in Bezug zum Therapieergebnis. Patienten mit einer erhöhten Inkongruenz profitierten dabei stärker von stärkeren frühen intrapersonalen Erfahrungen. Die interpersonalen Erfahrungen aus Patientenperspektive waren dagegen unabhängig von der Inkongruenz mit dem Therapieergebnis assoziiert.

Die dritte Studie legte den Fokus auf Patienten, die die Therapie abgebrochen hatten. Therapieabbruch oder drop-out ist ein häufiges Problem in der Psychotherapie, da Patienten sich unter Umständen nicht hinreichend verbessern können oder sich sogar verschlechtern. Als statistische Methode diente dabei die Survivalanalyse, da in diese zeitlich stabile (z.B. initiale Behandlungsdiagnose) und zeitliche variable (z.B. interpersonale Erfahrungen) Prädiktoren aufgenommen werden können. Zudem konnten so die individuellen Therapieverläufe jedes einzelnen Patienten berücksichtigt werden. Die Ergebnisse zeigen, dass die therapeuten- und patienteneingeschätzten interpersonalen Erfahrungen mit dem Therapieergebnis assoziiert waren.

Insgesamt zeigen die Studien, dass die patienteneingeschätzten interpersonalen Erfahrungen am konsistentesten mit verschiedenen Therapieergebnismaßen assoziiert sind. Die intrapersonalen Erfahrungen, sowie die Therapeutenperspektive der generellen Veränderungsmechanismen zeigen dagegen weniger konsistente Ergebnisse. Die Problemaktualisierung ist dagegen in keiner der Studien und aus keiner der betrachteten Perspektive ein signifikanter Prädiktor. Die regelmäßige Kontrolle der interpersonalen Erfahrungen über den Thera-

pieverlauf kann daher für die therapeutische Praxis empfohlen werden. Die zukünftige Forschung sollte sich weiterhin an generellen Veränderungsmechanismen orientieren, da diese einen Einblick in den therapeutischen Prozess ermöglichen und das Therapieergebnis verbessern können.

Abstract – English

Numerous studies have shown the efficiency and effectiveness of psychotherapy, especially cognitive behavioral therapy. Which components elicit psychotherapeutic change is still discussed. Studies focused on specific factors of certain treatments but some studies also focused on general change mechanisms. The studies of this dissertation considered three of these general change mechanisms: interpersonal experiences (i.e. therapeutic alliance, openness to/contentment with treatment), intrapersonal experiences (i.e. insight, mastery of problems) and problem actuation (i.e. emotional activation during treatment). Main focuses were the association between these general change mechanisms and different treatment outcomes. The sample consisted of patients from routine practice of an outpatient clinic who received cognitive behavioral therapy (524 – 911 patients depending on the study).

The first study was concerned with the association of early developments in general change mechanisms and different treatment outcomes: symptom severity and life satisfaction each as continuous and dichotomous outcome, based on the reliable change index. Results showed that early developments in general change mechanisms were associated with treatment outcome depending on the outcome (symptom severity or life satisfaction; continuous or dichotomous) and on the rater's perspective (therapist's or patient's perspective). General change mechanisms explained 2-4 % of additional variance.

The second study considered motivational incongruence as a possible moderator of the relationship between early general change mechanisms and treatment outcome. The main hypotheses focused on early intrapersonal experiences and treatment outcome; the association of early interpersonal experiences and early problem actuation and treatment outcome were exploratory hypotheses. Motivational incongruence is caused if basic needs are not met in a person's everyday life like the need for attachment. Previous studies showed that especially patients with psychological problems suffer from high incongruence. The results showed a significant interaction of incongruence on early patient-rated intrapersonal experiences on treat-

ment outcome. Patients with high incongruence showed better treatment outcomes when early intrapersonal experiences were also higher. Early interpersonal experiences were associated with treatment outcome regardless of incongruence.

The third study focused on patients, who discontinued treatment prematurely (dropped out). Drop-out is a problem for psychotherapy because patients might not get better and might even get worse when dropping out. Survival analysis was used as statistical procedure because it could consider time independent (e.g. initial diagnoses) and time varying (e.g. interpersonal experiences) predictors. In addition, patients' individual treatment courses could be considered. Results showed that therapist- and patient-rated interpersonal experiences over the course of treatment were associated with drop-out.

In summary, our studies showed that patient-rated interpersonal experiences was the most consistent predictor. Patient-rated intrapersonal experiences and therapist-rated general change mechanisms showed less consistent associations with treatment outcome. Problem actuation did not show significant results in any of the studies. Thus, routine outcome monitoring of interpersonal experiences over the course of treatment might benefit therapeutic practice. Future research should consider general change mechanisms in more detail (e.g. more often during treatment) to enhance therapeutic processes and treatment outcome.

Inhaltsverzeichnis

Hintergrund	1
Common Factors und Grawes generelle Veränderungsmechanismen als Erfolgsfaktoren in der Psychotherapie	1
Aktueller Forschungsstand zu Grawes generellen Veränderungsmechanismen	9
Die Beziehung der kognitiven Verhaltenstherapie zu Common Factors und Grawes generellen Veränderungsmechanismen	12
Messung von Therapieerfolg	14
Eine Methode zur Erfassung von Grawes generellen Veränderungsmechanismen: der Berner Patienten- und Therapeutenstundenbogen	20
Zielsetzung und Forschungsagenda	24
Publikationen	28
Gmeinwieser, Kuhlencord, Ruhl, Hagmayer und Probst (2019a)	28
Gmeinwieser, Hagmayer, Pieh und Probst (2019b)	66
Gmeinwieser, Schneider, Bardo, Brockmeyer und Hagmayer (in press)	105
Diskussion und Ausblick	140
Gmeinwieser et al. (2019a)	140
Gmeinwieser et al. (2019b)	146
Gmeinwieser et al. (in press)	148
Zusammenfassende Betrachtung	151
Zukünftige Forschung und Ausblick	157
Literaturverzeichnis zum Rahmentext	160
Appendix	174
Appendix A: R-Skript	174
Appendix B: Lebenslauf	179

Tabellenverzeichnis zum Rahmentext

Tabelle 1. Deskriptive Werte der Population und der simulierten Stichprobe.....	15
Tabelle 2. Kategoriale Einteilung der simulierten Stichprobe (s. Jacobson & Truax, 1991)...	17
Tabelle 3. Signifikante Prädiktoren aus Gmeinwieser et al. (2019a).....	141
Tabelle 4. Umrechnung von Cohens d in Odds und Hazard Ratios.....	150

Hintergrund

In diesem Abschnitt werden die grundlegenden Konzepte und Methoden dieser Dissertation vorgestellt. Da das Zentrum dieser Arbeit drei Publikationen zu Common Factors (CFs) bzw. zu generellen Veränderungsmechanismen (general change mechanisms, GCMs) sind, werden GCMs im ersten Abschnitt einführend erklärt. Es wird außerdem ein kurzer Blick auf die Entstehung der Idee von CFs und GCMs eingegangen und einige empirische Befunde beschrieben. Der darauffolgende Abschnitt beschäftigt sich mit Möglichkeiten, Therapieerfolg zu erfassen. Verschiedene Ansätze werden hierbei vorgestellt und kritisch beleuchtet. Jeder der dort vorgestellten Ansätze wurde in einer der Publikationen verwendet. Der nächste Abschnitt stellt das für die drei Publikationen wesentliche Messinstrument zur Erfassung von generellen Veränderungsmechanismen vor: den Berner Patienten- und Therapeutenstundenbogen. Das Instrument wird in diesem Abschnitt detailliert beschrieben und kritisch diskutiert. Zudem werden die Skalen aus dem Berner Stundenbogen vorgestellt. Der letzte Abschnitt befasst sich mit dem Forschungsschwerpunkt und der Fragestellung der jeweiligen Publikationen. Es werden kurz die Gemeinsamkeiten der Stichproben der jeweiligen Publikationen vorgestellt und danach wird auf die jeweiligen Fragestellungen eingegangen.

Common Factors und Grawes generelle Veränderungsmechanismen als Erfolgsfaktoren in der Psychotherapie

Die Wirksamkeit von Psychotherapie ist vielfach untersucht und belegt worden (z.B. Butler, Chapman, Forman & Beck, 2006; Lambert, 2013; Leichenring & Rabung, 2008). Ein Problem des Wirksamkeitsnachweises ist allerdings, dass unklar ist, welche genauen Bestandteile der Psychotherapie Veränderung bewirken. Ein lang geführter Diskurs in der Psychotherapie behandelt deshalb das Thema, welche Faktoren den Erfolg von Psychotherapie wesentlich beeinflussen. Auf der einen Seite stehen dabei die spezifischen Wirkmechanismen, die bei bestimmten Verfahren oder „Schulen“ der Psychotherapie exklusiv Verwendung finden

(Crits-Christoph, 1997; DeRubeis & Crits-Christoph, 1998). Beispielsweise wurde die Wirksamkeit kognitiv-verhaltenstherapeutischer Techniken wie etwa die Verhaltensmodifikation oder die kognitive Umstrukturierung stark untersucht (Task Force on Promotion and Dissemination of Psychological Procedures, 1995). Auf der anderen Seite steht die therapieformübergreifende Perspektive. Diese Perspektive beschäftigt sich mit Faktoren, die allen Therapierichtungen gemeinsam sind: den Common Factors (CFs). Geprägt wurden die Studien zu CFs vom Dodo-Bird Verdict, das zu Beginn von Rosenzweig (1936) als Metapher für CFs herangezogen und später von Luborsky, Singer und Luborsky (1975) wieder aufgegriffen wurde.

Das Dodo-Bird Verdict stammt aus dem Kinderbuch, „Alice’s Adventures in Wonderland“, von Lewis Carroll (1865). Dort beendet der Dodo-Bird ein Wettrennen mit den Worten, „Everybody has won, and all must have prizes“ (S.34). Auf die Psychotherapie in ihren verschiedenen Ausprägungen übertragen, bedeutet diese Aussage, dass jede Form der Therapie letztlich zu einem positiven Ergebnis für Patienten führen kann. Ob der vergleichbare Therapieerfolg verschiedener Therapierichtungen ein Beleg für CFs ist, wurde zunächst theoretisch diskutiert (Rosenzweig, 1936), in einem Review von Luborsky et al. (1975) kritisch untersucht und in späteren Meta-Analysen auch gezeigt (z.B. Ahn & Wampold, 2001; Wampold et al., 1997). Schlussfolgernd beschreiben Luborsky et al. (1975) etwa eine therapeutische Allianz als Gemeinsamkeit in den Studien ihres Reviews, die einen Teil des Effektes von Psychotherapie ausmachen könnte.

Die therapeutische Allianz, d.h. die Beziehung zwischen Therapeut und Patient während der Psychotherapie (im Folgenden nur noch „Allianz“ genannt), ist ein häufig beschriebener CF. Die Allianz wurde zunächst von der psychoanalytischen Literatur aufgegriffen (u.a. Sterba, 1934) und z.B. von Greenson (1965) als Fähigkeit des Patienten beschrieben, zielgerichtet in der Therapiesituation zu arbeiten (vgl. S.157). Gemeinsame Ziele wurden auch von Bordin (1979) als Teil der Allianz definiert, zusammen mit spezifischen Aufgaben in der Therapie wie die Selbstöffnung und -beobachtung, sowie die Bindung, die sich durch gegenseitig-

ges Vertrauen zwischen Patient und Therapeut ausbildet. Der Beitrag des Therapeuten zur Allianz wurde von Rogers (1957) stärker in den Fokus gerückt. In seinem Artikel beschreibt Rogers (1957) Eigenschaften wie die unbedingte Wertschätzung des Patienten, Empathie und Kongruenz seitens des Therapeuten als notwendig für den Aufbau therapeutischer Allianz und sogar hinreichend zum Herbeiführen von positiven therapeutischen Veränderungen. Diese Konzepte bilden die Grundlage aktueller Definitionen und überprüfbarer Konstrukte von therapeutischer Allianz. Hier zeigt sich aber auch, dass die Definitionen einige Unterschiede aufweisen. Während Greenson (1965) die Allianz als zielgerichtetes Arbeiten zusammenfasst, beschreibt Bordin (1979) weitere Teilaspekte der Allianz, wie gemeinsame Ziele und die Selbstöffnung. In einer aktuellen Studie beschreibt Zilcha-Mano (2017) darüber hinaus den Unterschied zwischen einer state- und einer trait-like Allianz, d.h. einer fluktuierenden Komponente im Verlauf der Therapie und der generellen Fähigkeit des Patienten, Beziehungen und Bindungen zu anderen Menschen einzugehen. Diese unterschiedlichen Definitionen haben letztlich zu einer Fülle verschiedener Instrumente zur Messung der therapeutischen Allianz geführt. Horvath, Del Re, Flückiger und Symonds (2011) zeigen allein in ihrer Arbeit, dass die Allianz mit 30 verschiedenen Instrumenten gemessen werden kann und dass ein gemeinsamer Faktor der Messinstrumente, die sichere, gemeinschaftliche Beziehung („confident collaborative relationship“, S.10) lediglich 50% gemeinsame Varianz unter den Instrumenten aufklärt. Trotz der unterschiedlichen Definitionen zeigten bereits frühe Meta-Analysen, die den Zusammenhang von Allianz und Therapieergebnis untersuchten, Effektgrößen zwischen $r = 0,22-0,275$ unter den berücksichtigten Studien (Horvath et al., 2011; Horvath & Symonds, 1991; Martin, Garske & Davis, 2000). Diese Ergebnisse wurden von Flückiger, Del Re, Wampold und Horvath (2018) in einer aktuellen Meta-Analyse weiter gefestigt. Die Meta-Analyse umfasste 295 Studien mit insgesamt über 30.000 Patienten. Die Autoren fanden eine Effektgröße von $r = 0,278$ (umgerechnet ein Cohens d von 0,579), die, nach Cohen (1988), einem mittleren Effekt entspricht. In den Ergebnissen wird unter anderem auch die Perspekti-

ve der Einschätzung untersucht. Typischerweise kann die Allianz nicht nur aus Sicht des Patienten, sondern auch aus Sicht des Therapeuten, eines Beobachters oder einer anderen beteiligten Person (Verwandter/Bekannter des Patienten) eingeschätzt werden. Flückiger et al. (2018) fanden allerdings keine signifikanten Unterschiede zwischen den Perspektiven, konträr zu vorangegangenen Befunden (z.B. Horvath & Symonds, 1991, Mander et al., 2015), in denen die Patientenperspektive stärker mit dem Therapieergebnis assoziiert war.

Eine Besonderheit der Allianz und anderen CFs ist unter anderem, dass sie im zeitlichen Verlauf der Therapie erhoben werden und damit zeitlich variabel sind. Typischerweise sind bestimmte Therapiesitzungen für eine solche regelmäßige Zwischenmessung oder Verlaufsmessung (routine outcome monitoring, ROM) genau festgelegt, z.B. nach jeder Sitzung (Rubel, Rosenbaum & Lutz, 2017) oder nach einer festgelegten Anzahl von Behandlungswochen (Zilcha-Mano, Dinger McCarthy & Barber, 2014). Da auch die allgemeine Symptombelastung zeitlich variiert und häufig im gleichen Takt wie die CFs erfasst wird (z.B. Lutz et al., 2014; Tang, DeRubeis, Beberman & Pham, 2005), diskutieren DeRubeis, Brotman, und Gibbons (2005) kritisch die Frage, ob die Allianz über die Symptombelastung hinaus überhaupt einen Einfluss auf das Therapieergebnis hat. DeRubeis et al. (2005) argumentieren, dass die Assoziation zwischen Allianz und Therapieergebnis lediglich ein uninteressantes Epiphänomen (vgl. S.179) vorangegangener Verbesserung in der Symptombelastung sei und dass, nach Kontrolle dieser frühen Symptomverbesserung, keine Assoziation zwischen Allianz und Therapieergebnis bestünde. Tatsächlich ist die empirische Befundlage zu dieser Frage gemischt. Studien fanden sowohl keinen Einfluss der Allianz auf das Therapieergebnis nach vorheriger Kontrolle der Symptombelastung (Feeley, DeRubeis & Gelfand, 1999), sie fanden eine signifikante Assoziation (Barber, Connolly, Crits-Christoph, Gladis & Siqueland, 2000; Zilcha-Mano et al., 2014) oder es stellte sich heraus, dass Symptombelastung und Allianz sich wechselseitig im Therapieverlauf beeinflussten (Falkenström, Ganström & Holqvist, 2013, 2014).

Weitere Forschung zu dieser Frage ist daher unbedingt nötig. Der Fokus sollte dabei nicht nur auf der Allianz liegen, sondern auf weitere CFs ausgebreitet werden.

Auch wenn die Allianz empirisch am stärksten untersucht ist, wurden weitere CFs in der Literatur beschrieben und auf ihren Einfluss auf das Therapieergebnis untersucht. Cuijpers, Reijnders und Huibers (2018) beschreiben in ihrem Review beispielsweise Empathie und Feedback, für die Wirksamkeitsbelege vorliegen. Weitere CFs wie Einsicht, affektive Erfahrung, Erfolgserfahrungen oder Therapeutenexpertise werden ebenfalls genannt. Die Autoren beschäftigt außerdem die Frage, inwiefern die vergleichbare Wirksamkeit verschiedener Therapieformen tatsächlich auf den Einfluss von CFs zurückzuführen ist. Die Erklärung, dass auch verschiedene Therapieansätze mit unterschiedlichen Methoden, also mit unterschiedlichen spezifischen Wirkmechanismen, zu therapeutischer Veränderung in der Therapie führen könnten, sei aufgrund mangelnder Forschung zu CFs durchaus gerechtfertigt. Allerdings schlagen Cuijpers et al. (2018) auch mögliche Lösungswege vor, wie spezifische Studiendesigns diese Frage untersuchen könnten. Besonders wird dabei auf das Analysieren der zeitlichen Assoziation der CFs verwiesen, entgegengesetzt zu den bisher dominanten korrelativen und querschnittlichen Untersuchungen. Das bedeutet, dass CFs nicht einmalig im Verlauf der Therapie erhoben und analysiert werden sollten, sondern fortlaufend. So würde die Veränderung im zeitlichen Verlauf einen Einblick in die therapeutischen Prozesse ermöglichen. Das abschließende Fazit der Autoren lautet, dass im Bereich Therapieprozessforschung und CFs weiterer Forschungsbedarf besteht.

Grawe (1997) formulierte aus den Überlegungen zu CFs sogenannte Veränderungsmechanismen (vgl. „mechanisms of change“, S.1), die, auf empirischer Basis formuliert, positive Veränderung in der Therapie bewirken sollen. In späteren Arbeiten wurden diese Veränderungsmechanismen general change mechanisms (GCMs¹) genannt (z.B.

¹ Der Begriff GCMs bezieht sich im Folgenden ausschließlich auf die generellen Veränderungsmechanismen nach Grawe

in Gassmann & Grawe, 2006). Grawe (1997) formulierte zunächst vier solcher Mechanismen: Problembewältigung, motivationale Klärung, Problemaktualisierung und Ressourcenaktivierung. Problembewältigung wird beschrieben als der erfolgreiche Umgang des Patienten mit seiner Problematik. Therapeutische Techniken können den Patienten dabei unterstützen. Zum Beispiel kann ein Patient mit sozialer Phobie dazu motiviert werden, einen Vortrag vor einem Publikum zu halten, um die Erfahrung zu machen, dass sich seine spezifischen Ängste, wie Stottern, Erröten, nicht bewahrheiten oder weniger stark als erwartet ausgeprägt sind. In der motivationalen Klärung soll der Patient die Bedeutung und die Motive hinter seiner Problematik verstehen. Der Therapeut kann diesen Prozess beispielsweise fördern, indem er im sokratischen Dialog etwaige dysfunktionale Gedanken oder Verhaltensweisen naiv kritisch hinterfragt oder mit dem Patienten das Pro und Contra der Veränderung seiner Symptomatik versus eine Nicht-Veränderung erarbeiten. Die Problemaktualisierung wird beschrieben als die Erfahrung des Patienten, im geschützten Rahmen der Therapie störungsrelevante Probleme direkt wiederzuerleben, um so eine Veränderung der Probleme zu provozieren. Ein Patient mit Spinnenphobie, der in der Therapie mit einer Spinne konfrontiert wird, soll etwa die Angst in der gleichen Weise wiedererleben, die ihn zuvor veranlasst hat, ein Therapieangebot aufzusuchen. Durch dieses Wiedererleben, zusammen mit der Erfahrung, die Angst bewältigen zu können, soll letztlich eine Veränderung bewirkt werden. Zuletzt spricht Grawe (1997) noch die Ressourcenaktivierung an. In dieser sollen die günstigen Voraussetzungen eines Patienten in der Therapie herausgearbeitet und gestärkt werden. Ressourcen können dabei unterschiedliche Formen annehmen, wie z.B. soziale, finanzielle, charakterliche oder äußerliche. Wichtig ist, dass der Therapeut den Patienten nicht ausschließlich durch eine problemfokussierte Sicht betrachtet, sondern auch durch eine Ressourcenperspektive, die den Patienten eigene latente oder offenkundige Potenziale erkennen lässt. Jeder GCM kann im Therapieprozess an geeigneter Stelle angewendet werden und soll letztlich zu einer positiven Veränderung des

Patienten führen. Auch wenn nicht initial als GCM formuliert, misst Grawe der Allianz ebenfalls einen hohen Stellenwert zu. Unter dem allgemeinen Begriff der interpersonellen Beziehung werden daher die GCMs um eine weitere Perspektive ergänzt, die unter anderem auch die Allianz beinhaltet.

Die Idee hinter den GCMs nach Grawe (1997) ist letztlich eine alternative Form der Therapie. Eine Therapie, die nicht auf schulengebundene Techniken beschränkt ist, sondern sich den individuellen Bedürfnissen eines Patienten anpasst und den Fokus auf die GCMs legt. Diese Ansicht ist gegensätzlich zur behandlungsfokussierten Therapie, die ihren Ursprung im sogenannten medizinischen Modell hat. Nach Elkins (2009) liegt der Fokus des medizinischen Modells bei der Diagnose des Patienten auf Basis der Symptome und eine Behandlung der daraus folgenden Erkrankung. Elkins (2009) beschreibt das medizinische Modell, auch wenn in der Medizin adäquat, als nicht geeignet für die Psychotherapie, welche Therapie als interpersonellen Prozess ansieht und nicht als hierarchischen Prozess zwischen einem höher gestellten Behandler und dem Patient. Einen Prozess, in dem der Patient mit individuellen Problemen eine Behandlung aufsucht. In der Psychotherapie werden nicht reine Symptome oder Erkrankungen behandelt, der Patient wird stattdessen auch zur persönlichen Weiterentwicklung und zum Setzen neuer Ziele ermutigt. Ein Modell, das dagegen der allgemeinen Psychotherapie zugrunde gelegt wurde, ist das konsistenztheoretische Modell (Grawe, 1999, 2004). Dieses Modell greift auf neurowissenschaftliche Befunde zum Lernen zurück. Demnach werden Verhaltensweisen oder Erfahrungen stärker neuronal verankert, je häufiger eine Person diesen ausgesetzt ist. Dies kann sowohl bei funktionalen als auch bei konfliktreichen Erfahrungen passieren. Wiederholte konfliktreiche Erfahrungen können zu neuronalen Bahnungen führen. Eine mögliche Ursache hinter diesen Konflikten können bestimmte Grundbedürfnisse sein, die nicht erfüllt wurden. Grundbedürfnisse zeichnen sich dadurch aus, dass sie allen Menschen gemeinsam sind, wie etwa das Bedürfnis nach Selbstwerterhöhung oder Kontrolle. Zum Schutz dieser Grundbedürfnisse können sich

bestimmte Vermeidungsreaktionen zum Schutz der Grundbedürfnisse ausbilden, welche langfristig problematische Züge annehmen können, da die Grundbedürfnisse zwar geschützt, aber nicht befriedigt werden. Nach Grawe (1999) führt eine Nicht-Befriedigung zu Konflikten, Inkonsistenz und dadurch einem fluktuierenden Spannungsniveau, welches eine neue neuronale Bahnung bilden kann. Diese Bahnung ist später allerdings unabhängig von den Grundbedürfnissen und kann maladaptiv sein. Als Beispiel wird eine Panikattacke genannt. Ziel der Therapie kann es hier sein, z.B. mittels klärender Techniken die maladaptiven Bahnungen aufzuzeigen, um sie in einem weiteren Schritt mittels bewältigungsorientierter Methoden in eine funktionale Richtung zu lenken.

Die GCMs nach Grawe (1997) basieren auf dieser Konsistenztheorie. Um Inkonsistenz im psychischen Geschehen zu reduzieren und sich den persönlichen Bedürfnissen des Patienten anzupassen, bilden die GCMs wichtige Orientierungspunkte. Die Idee einer allgemeinen Therapie ist auch heute noch aktuell. An der Universität Bern (n.d.) wird beispielsweise seit Jahren eine Therapieausbildung auf der Grundlage von Grawes Theorien und Überlegungen angeboten.

GCMs und CFs weisen einige überschneidende Konzepte auf, wie z.B. die Allianz. Ein wesentlicher Unterschied, der GCMs und CFs allerdings voneinander trennt, ist die theoretische Fundierung. CF ist ein Sammelbegriff für eine Vielzahl von Techniken, Prozessen und Methoden zum Herbeiführen therapeutischer Veränderung, von denen viele eigene theoretische Fundierungen oder eigene Forschungswege haben. Feedback im Therapieverlauf hat beispielsweise eine breite empirische Basis (Bergman et al., 2018) und eigene Arbeitsgruppen, die die Nutzung und Verbreitung von Feedback vorantreiben (z.B. Lambert, 2007; Lutz, De Jong & Rubel, 2015; Probst et al, 2013; Shimokawa & Lambert, 2010). GCMs können ebenfalls einem unabhängigen Forschungsweig zugeordnet werden. Prinzipiell basieren die in dieser Arbeit betrachteten GCMs auf den von Grawe (1997, 1998, 1999, 2004) entwickelten Ideen und der theoretischen Fundierung für eine allgemeine Form

der Therapie. Untersucht wird dabei allerdings nicht ein einzelnes Konzept wie etwa Feedback im entsprechenden Forschungszweig, sondern eine Vielzahl von Ansätzen, die aus Grawes (ibid.) Überlegungen stammen.

Aktueller Forschungsstand zu Grawes generellen Veränderungsmechanismen

Abgesehen von der empirischen Forschung zur Allianz wurde zu einigen weiteren von Grawe formulierten GCMs ebenfalls Studien durchgeführt (z.B. D'Errico et al., 2019; Flückiger & Grosse Holtforth; Mander et al., 2013, 2015; Rubel et al., 2017; Sander, Zobel, Dykieriek & Schramm, 2012; Stangier, Consbruch, Schramm & Heidenreich, 2019). Call, D'Errico, Bents und Mander (2018) führten etwa eine Studie mit 870 Patienten aus der ambulanten Routineversorgung durch. Dabei wurden die GCMs Problemaktualisierung, Ressourcenaktivierung, motivationale Klärung und Problembewältigung zu zwei Zeitpunkten in der Therapie überprüft; zur 15. (t1), sowie zur letzten Sitzung (t2) der Therapie. Die zu t1 erfassten GCMs zeigten signifikante Zusammenhänge mit allgemeiner und depressiver Symptombelastung zu t2. Die Veränderung in GCMs außer Problemaktualisierung von t1 zu t2 sagte darüber hinaus depressive Symptombelastung zu Therapieende (t2) vorher. Die Korrelationen waren allerdings klein und schwankten zwischen $r = 0,14$ (Problembewältigung zu t1) und $r = 0,24$ (Differenz Problembewältigung von t2 zu t1). Vogel, Blanck, Bents und Mander (2016) dagegen fanden keine Zusammenhänge zwischen GCMs zum Zwischenmesszeitpunkt und dem Therapieergebnis unter 303 Patienten, die an einer stationären oder ambulanten Gruppentherapie teilnahmen. Die Differenz der GCMs interpersonelles Lernen (bestehend aus Ressourcenaktivierung, Problembewältigung und motivationale Klärung) und Problemaktualisierung zum Zwischen- und zum Postmesszeitpunkt zeigte dagegen signifikante Zusammenhänge in den Maßen zur Symptombelastung (interpersonelles Lernen, Problemaktualisierung) und depressive Symptombelastung (interpersonelles Lernen). Allerdings kann an der Studie von Vogel et al.

(2016) kritisiert werden, dass sich die Patienten zusätzlich zu Gruppentherapie auch in Einzeltherapie befanden, und dass der Einfluss der Einzeltherapie auf das Therapieergebnis unklar ist.

Flückiger, Grosse Holtforth Znoj und Wampold (2013) schlugen weitere GCMs vor, die sich am dualen Modell des therapeutischen Prozesses von Schulte (1996) orientieren. Flückiger et al. (2013) legten den Fokus auf die frühe Therapiephase, in der vor allem interpersonale und intrapersonale Prozesse eine Rolle spielen. Zudem wurde noch die Problemaktualisierung aufgegriffen, die einen wichtigen Beitrag zur Symptomreduktion am Ende der Therapie leisten kann. In der Studie prüften die Autoren also die GCMs interpersonale Erfahrungen (interpersonal experiences), intrapersonale Erfahrungen (intrapersonal experiences) und Problemaktualisierung (problem actuation). Die zentrale Fragestellung der Studie war es, ob diese GCMs das Therapieergebnis vorhersagen, nachdem die Symptombelastung zu Therapiebeginn mitberücksichtigt wurde. Die Ergebnisse zeigten, dass interpersonale Erfahrungen aus Patienten- und Therapeutensicht ein robuster Prädiktor für das Therapieergebnis waren.

Eine weitere Studie, die interpersonale Erfahrungen, intrapersonale Erfahrungen und Problemaktualisierung nutzte war eine Studie von Probst et al. (2016). Ziel der Studie war der Vergleich zwischen verzögerter Videosupervision (delayed video-based, DVB) gegen Supervision über Video in Echtzeit (bug-in-the-eye, BITE) auf den verschiedenen GCM-Dimensionen aus Patienten- und Therapeutensicht in einem ambulanten Psychotherapiesetting. Es zeigte sich, dass interpersonale und intrapersonale Erfahrungen aus Therapeutenperspektive zwischen den Bedingungen nicht unterschiedlich stark bewertet wurden, aber dass Problemaktualisierung stärker in der BITE-Bedingung vorhanden war als in der DVB-Bedingung. Die Patienteneinschätzung der GCMs zeigte keine signifikanten Unterschiede zwischen den Bedingungen.

Keine weiteren Studien haben sich mit den GCM-Dimensionen interpersonale und intrapersonale Erfahrungen beschäftigt. Auch hier besteht demnach weiterer Forschungsbedarf. Um besser an die Studien von Flückiger et al. (2013) und Probst et al. (2016) anzuknüpfen ist es zudem sinnvoll, auch die Problemaktualisierung näher zu beleuchten.

Zusammenfassend lässt sich die momentane Forschung zu GCMs folgendermaßen beschreiben: Auch, wenn es bereits eine breite empirische Basis für die Allianz als Erfolgsfaktor in der Psychotherapie gibt, ist die Forschung zu anderen GCMs, vor allem interpersonalem Erfahrungen, intrapersonaler Erfahrung, Problemaktualisierung, und dem Therapieergebnis deutlich limitierter. Weiterer Forschungsbedarf besteht dabei nicht nur zu diesen GCMs und verschiedenen Therapieergebnismaßen, sondern auch zu GCMs, die über ROM mehrfach im Therapieverlauf erhoben wurden. Weiterhin ist der wechselseitige Einfluss von Symptombelastung und GCMs ein wichtiger limitierender Faktor bisheriger Forschung. Zur Allianz gibt es zwar bereits einige Studien, die den wechselseitigen Einfluss betrachten, allerdings gibt es nur wenige Studien, die neben den GCMs auch zusätzlich die Symptombelastung betrachten. Nachfolgende Studien sollten die Symptombelastung daher ebenfalls im zeitlichen Verlauf erfassen und als Einflussfaktor im Therapieergebnis berücksichtigen.

In dieser Arbeit wurden die inter- und intrapersonalen Erfahrungen, sowie die Problemaktualisierung als Forschungsgegenstand ausgewählt. Der Grund dafür ist, dass diese GCMs, wie auch die zuvor genannten GCMs (Problembewältigung, motivationale Klärung, Problemaktualisierung, Ressourcenaktivierung) auf der von Grawe entwickelten Konsistenztheorie (u.a. Grawe, 2004) beruhen und darüber hinaus eine neuere Entwicklung in der GCM-Forschung darstellen. Zusätzlich werden die GCMs inter- und intrapersonalen Erfahrungen und Problemaktualisierung aus dem dualen Modell der Psychotherapie (Schulte, 1996) abgeleitet, was die theoretische Grundlage weiter festigt.

Die Beziehung der kognitiven Verhaltenstherapie zu Common Factors und Grawes generellen Veränderungsmechanismen

Der folgende Abschnitt beschäftigt sich mit dem Verhältnis der kognitiven Verhaltenstherapie (cognitive behavioral therapy, CBT) zu den Konzepten der CFs und der GCMs. Dabei wird nach einer kurzen Einführung zur CBT vor allem die empirische Forschung zu den GCMs betrachtet, die auch kognitive Verhaltenstherapie untersucht.

Die CBT beruht prinzipiell auf lerntheoretischen Ansätzen zur klassischen (Pavlov, 1927) und operanten Konditionierung (Skinner, 1938), sowie auf kognitiven Verfahren (z.B. Beck, 1970; Ellis, 1957). Interventionen gestalten sich demnach verhaltensnah und problemorientiert und versuchen etwa, die Gedanken des Patienten mit spezifischen Interventionen in eine positive Richtung zu lenken. Beispiele für solche Interventionen sind etwa Rollenspiele oder kognitive Umstrukturierung, bei der dysfunktionale Gedanken mit positiven, funktionalen Gedanken ersetzt werden. In Deutschland ist die CBT ein Richtlinienverfahren. Das bedeutet, dass Krankenkassen die Kosten einer Psychotherapie mit CBT Schwerpunkt übernehmen. Die CBT ist unter anderem auch deshalb ein Richtlinienverfahren, weil eine breite empirische Basis die Effektivität und Effizienz der CBT bestätigt hat (z.B. Hofman, Asnaani, Vonk, Sawyer & Fang, 2012).

Wie bereits im vorhergehenden Abschnitt beschrieben, besteht die stärkste empirische Basis der CFs und GCMs im Bereich der therapeutischen Allianz. Auch die Schnittmenge der Forschung zu CFs bzw. GCMs und der CBT besteht daher aus Forschung zu diesem Bereich. Die Meta-Analyse von Flückiger et al. (2018) zur Allianz berichtet etwa einen Effekt von $r = 0,20$ von Allianz auf das Therapieergebnis in CBT-Studien. In einer vorausgegangenen Meta-Analyse von Horvath et al. (2011) lag dieser Effekt bei $r = 0,35$. Insgesamt lässt sich der Effekt demnach als klein bis mittel einschätzen, d.h. ein Effekt der Allianz auf das Therapieergebnis ist auch in der CBT nicht von der Hand zu weisen.

Meta-Analysen zu anderen von Grawe (1997) beschriebenen GCMs und der CBT sind dagegen nicht vorhanden. Studien, die sich mit GCMs beschäftigen, haben zudem häufig nicht das Ziel, den Einfluss von GCMs auf das Therapieergebnis für eine spezifische Therapieform zu belegen. Häufig wird nur bei einem genaueren Blick deutlich, dass sich eine Studie auf Patienten bezieht, die mittels CBT behandelt wurden. Den zuvor genannten Studien von Call et al. (2018), D'Errico et al. (2019), Sander et al. (2012) und Vogel et al. (2016) ist vor diesem Hintergrund hinzuzufügen, dass sich diese auf Patientenstichproben beziehen, die ausschließlich mittels CBT behandelt wurden. Das bedeutet, dass die gefundenen Zusammenhänge in erster Linie auf andere CBT-Stichproben generalisierbar sein können, auch wenn die ursprüngliche Idee hinter den GCMs natürlich ein verfahrensunabhängiger Einfluss auf das Therapieergebnis ist. Demnach bestehen Belege für den Einfluss von GCMs auf das Therapieergebnis in der CBT für die Konzepte Motivationale Klärung (Call et al., 2018; D'Errico et al., 2019; Sander et al., 2012; Vogel et al., 2016), Problembewältigung (Call et al., 2018; D'Errico et al., 2019; Sander et al., 2012; Vogel et al., 2016), Problemaktualisierung (Call et al., 2018) und Ressourcenaktivierung (Call et al., 2018; D'Errico et al., 2019; Vogel et al., 2016).

Für die von Flückiger et al. (2013) beschriebenen GCMs interpersonale und intrapersonale Erfahrungen ist die empirische Basis wiederum kleiner, vor allem für die CBT. Arbeiten, die sich mit diesen Konzepten beschäftigten, wurden entweder in gemischten Behandlungssettings durchgeführt (Flückiger et al., 2013) oder hatten nicht den Fokus, den Zusammenhang zwischen GCMs und dem Therapieergebnis vorherzusagen (Probst et al., 2016).

Es besteht demnach dringend weiterer Forschungsbedarf zu den GCMs interpersonale Erfahrungen und intrapersonale Erfahrungen in Verbindung mit der Problemaktualisierung und dem Therapieergebnis. Auch wenn einige Studien zu anderen GCMs Hinweise des Einfluss der GCMs auf das Therapieergebnis in der CBT geben, sind auch darüber hinaus

noch weitere Studien notwendig, die stärker den zeitlichen Verlauf der GCMs in der CBT betrachten oder den Fokus stärker auf alternative als auf symptombezogene Maße legen.

Messung von Therapieerfolg

Therapieerfolg kann auf unterschiedliche Arten beurteilt werden. Im folgenden Abschnitt soll der Fokus auf drei verschiedene Arten gelegt werden, Therapieerfolg zu messen: kontinuierliche Werte eines klinisch relevanten Outcomes, kategoriale Einteilung von Therapieerfolg nach klinisch bedeutsamen Kriterien und einseitige Beendigung von Therapie (auch Therapieabbruch oder Drop-out genannt).

Ein vielfach verwendetes Forschungsdesign zur Beurteilung von Behandlungsprogrammen in der klinischen Forschung ist das randomisierte Prä-Post-Kontrollgruppendesign, welches Therapieerfolg als Outcome heranzieht. Patienten werden randomisiert einer von zwei Gruppen zugewiesen. Üblicherweise beinhaltet eine Gruppe die zu untersuchende Behandlung (Experimentalgruppe, EG) und die andere ist eine Form von Kontrollgruppe (KG). KG können dabei Gruppen sein, in denen der Patient keine Behandlung erhält (Warteliste-KG) oder Gruppen, die ebenfalls eine Behandlung beinhalten, von der allerdings ausgegangen wird, dass diese zu einem schlechteren Therapieergebnis führt als die EG. Üblicherweise werden in diesen Gruppen Fragebögen durchgeführt, auf denen z.B. der Patient sein Wohlbefinden oder seine Belastung einschätzt. Diese Fragebögen werden zu einem Zeitpunkt vor der Therapie (Präzeitpunkt) und wiederum nach der Therapie (Postzeitpunkt) erhoben. Erwartet wird hier, dass sich Patienten zum Präzeitpunkt zwischen den Gruppen nicht unterscheiden und dass sich die Patienten in der EG stärker verbessern als in der KG. Zur Veranschaulichung dieses Forschungsdesigns wurde eine Simulation in R (R Core Team, 2018) geschrieben. Der verwendete Code kann Appendix A entnommen werden. Für die Simulation wurden zudem die Packages „MASS“ (Venables & Ripley, 2002), „psych“ (Revelle, 2018) und „lsr“ (Navarro, 2015) herangezogen. Es wurden zwei Stichproben der

jeweiligen Größe von $N = 100$ simuliert. Die Stichproben haben jeweils einen Prä- und einen Postwert. Die deskriptiven Werte der Population (μ , σ) und der einzelnen Stichproben (M , SD) lassen sich Tabelle 1 entnehmen. Da eine der Stichproben zum Postzeitpunkt einen geringeren Mittelwert aufweist, soll diese im Folgenden als die KG und die Stichprobe mit dem höheren Mittelwert zum Postzeitpunkt als die EG bezeichnet werden. Ein wichtiger Unterschied der beiden Stichproben ist, dass sich diese hinsichtlich ihres Postwertes signifikant unterscheiden ($t(195,9) = 5,54$, $p < 0,001$, $d = 0,73$) und dass die Prä- und die Postwerte ebenfalls einen signifikanten Unterschied in beiden Stichproben aufweisen ($t(99) = 11,63$; $p < 0,001$; $d = 1,16$ KG und $t(99) = 21,63$; $p < 0,001$; $d = 2,16$ EG). Die Prä-Post-Differenzen in beiden Gruppen können als großer Effekt nach Cohen (1988) interpretiert werden. Der Unterschied beider Gruppen im Postwert ist dagegen eine Differenz von knapp vier Punkten und weist mit $d = 0,78$ einen mittleren bis großen Effekt auf.

Tabelle 1. Deskriptive Werte der Population und der simulierten Stichprobe

	KG	KG	EG	EG
	Population	Stichprobe	Population	Stichprobe
N	-	100	-	100
μ (σ) bzw. M (SD) Prä	50 (4,47)	50,02 (4,42)	50 (4,47)	49,84 (4,33)
μ (σ) bzw. M (SD) Post	55 (4,47)	55,19 (5,03)	59 (4,47)	58,94 (4,53)
ρ bzw. r Prä-Postwerte	0,5	0,56	0,5	0,55

Notiz: KG = Kontrollgruppe; EG = Experimentalgruppe; M = Mittelwert; SD = Standardabweichung

Es stellt sich die Frage, welche klinische Relevanz die Differenz und Effektgröße zwischen Gruppen tatsächlich hat. Diese Fragen stellten auch Jacobson, Follette und Revensdorf (1984) und aufbauend Jacobson und Truax (1991). In den beiden Arbeiten wurden Ideen entwickelt, Studienergebnisse für die klinische Praxis leichter interpretierbar zu

machen. Die Grundidee war es, für jeden einzelnen Patienten zu beurteilen, ob seine Veränderung überzufällig und klinisch relevant war. Die Beurteilung einer Veränderung eines individuellen Patienten erfolgte dabei auf der Basis von zwei Regeln: die Veränderung vom Prä- zum Postzeitpunkt muss reliabel sein und der Wert des Patienten soll sich zum Postzeitpunkt näher an einer gesunden Stichprobe befinden als an einer klinischen Stichprobe. Damit eine Veränderung als reliabel beurteilt werden kann, wurde der Reliable Change Index (RCI) definiert. Anhand des Standardfehlers des verwendeten Tests lässt sich eine Differenz bestimmen, die ein Patient von Prä zu Post mindestens überschreiten muss, damit die Veränderung als reliabel und damit als nicht-zufällig gilt. Der Postwert muss dabei außerhalb des 95%-Konfidenzintervalls des individuellen Präwerts liegen. Ob der Patient zum Postzeitpunkt dann zusätzlich noch näher an einer gesunden Stichprobe liegt als an einer klinischen Stichprobe, wurde mit sogenannten Cut-off-Werten bestimmt. Cut-offs trennen eine klinische von einer gesunden Stichprobe. Die genauen Kriterien, wie ein solcher Cut-off bestimmt wird, können sich, je nach Vorhandensein entsprechender Normstichproben, unterscheiden. Eine genaue Beschreibung kann Jacobson und Truax (1991) entnommen werden. Verbessert sich nun ein Patient um einen Wert größer als der RCI, befindet sich zum Präzeitpunkt unter dem Cut-off, d.h. näher an einer klinischen Stichprobe, und zu Postzeitpunkt über dem Cut-off, d.h. näher an einer gesunden Stichprobe, wird dies als klinisch signifikante Verbesserung bezeichnet. Nach Jacobson und Truax (1991) lassen sich so vier unterschiedliche Kategorien bilden: Klinisch signifikante Verbesserung (RCI + Cut-off-Kriterium), reliable Verbesserung (RCI ohne Cut-off-Kriterium), keine Veränderung (Prä-Post-Differenz kleiner als RCI) und reliable Verschlechterung (negativer RCI).

In der Simulation soll der RCI auf einen Wert von $RCI = 8,1^2$ festgelegt werden. Das bedeutet, dass jeder Patient, der sich von Prä zu Post um 8,1 Punkte verändert, sich auch

² Der RCI wurde auf der Grundlage der KG-Stichprobe und auf Basis von Jacobson und Truax (1991) bestimmt. Als Retest-Reliabilität diente dabei die Korrelation der Prä-Post-Werte der KG.

reliabel verändert. Der Cutoff zur gesunden Stichprobe soll auf 58,86³ festgesetzt werden. In Bezug auf die simulierten Daten bedeutet das, dass sich Patienten in der KG im Mittel nicht reliabel verbessern ($M_{KGPost} - M_{KGPrä} < 8,1$), die Patienten der EG dagegen schon ($M_{EGPost} - M_{EGPrä} > 8,1$). Allerdings liegt die KG zum Postzeitpunkt noch im klinischen Bereich und damit unter dem Cutoff. Bedeutet das nun auch, dass sich kein Patient in der KG klinisch signifikant verbessert? Wie sich die simulierten Patienten in die von Jacobson und Truax (1991) vorgeschlagenen Gruppen einteilen, kann Tabelle 2 entnommen werden. Es fällt auf, dass zwar im Mittel keine klinisch signifikante Verbesserung der KG zu verzeichnen ist, dass sich aber dennoch einzelne Patienten klinisch signifikant verbessern. Zudem zeigt sich, dass mehr Patienten in der EG eine reliable und eine klinisch signifikante Verbesserung aufweisen als Patienten in der KG. Der einfache Gruppenvergleich verschleiert den Umstand, dass sich dennoch Patienten in der KG reliabel und klinisch signifikant verbessern.

Tabelle 2. Kategoriale Einteilung der simulierten Stichprobe (s. Jacobson & Truax, 1991)

	Reliable Verslechterung	Keine Veränderung	Reliable Verbesserung	Klinisch signifikante Verbesserung
KG	0	81	8	11
EG	0	38	22	40

Notiz: KG = Kontrollgruppe; EG = Experimentalgruppe

Seit der Publikation von Jacobson und Truax (1991) haben die Konzepte des RCI und der klinischen Signifikanz eine breite Anwendung in Wissenschaft und Praxis gefunden (vgl. Ogles, Lunnen & Bonesteel, 2001). Allerdings wurden die Konzepte auch kritisiert. Ein

³ Der Cutoff wurde auf Grundlage der KG-Stichprobe und auf Basis von Jacobson und Truax (1991) bestimmt. Der Cutoff wurde auf einen Wert von zwei Standardabweichungen über dem Mittelwert der KG zum Präzeitpunkt festgelegt.

Kritikpunkt etwa, der von Speer (1992) angeführt wird, ist die Vernachlässigung der Regression zur Mitte. Im Detail bedeutet das, dass Patienten unabhängig von einer Intervention reliable Prä-Post-Veränderungen zeigen können, besonders, wenn initial ein stark von der Norm abweichender Fragebogenwert berichtet wird. Trotz dieser Kritik bietet sich mit dem RCI ein leicht bestimmbares Maß zur Beurteilung individueller Therapieverläufe. Dies ist gerade in der psychotherapeutischen Praxis von Vorteil, die Veränderung eines einzelnen Patienten über den Verlauf der Therapie im Vordergrund steht und Vergleiche zu anderen Patienten oder sogar einer kompletten Patientenstichprobe eher die Ausnahme bilden.

Neben der Benutzung kontinuierlicher Outcomes und der kategorialen Einteilungen von Therapieerfolg wird in vielen Studien auch der Drop-out als Erfolgskriterium herangezogen (z.B. Altmann et al., 2014; Cinkaya, Schindler & Hiller, 2011; Fernandez, Salem, Swift & Ramtahal, 2015; Issakidis & Andrews, 2004; Lutz et al., 2018; Sharf, Primavera & Diener, 2010; Simon & Ludman, 2010).

Laut Swift und Greenberg (2012) ist die implizite Definition von Drop-out über Studien hinweg vergleichbar: Patienten erreichen nicht ihr therapeutische Ziel und können keinen vollständigen Nutzen aus der Therapie ziehen. Operationalisiert wird Drop-out dagegen zum Teil unterschiedlich. Zum Beispiel gilt eine Therapie in Studien als abgebrochen, wenn Patienten einen vorgesehenen Behandlungsplan vorzeitig beenden oder Therapeuten treffen am Ende der Therapie eine Beurteilung, ob die Therapie als Abbruch gilt oder nicht. Allein diese unterschiedlichen Definitionen führen zu stark unterschiedlichen Abbruchraten in der Meta-Analyse von Swift und Greenberg (2012), die 669 Studien mit insgesamt 83.834 Patienten umfasste. Der Abbruch eines vorgesehenen Behandlungsplans zeigte dabei eine Abbruchrate von 18%, wohingegen das Therapeutenurteil die von allen berücksichtigten Moderatoren höchste Abbruchrate von 38% aufwies. Auch in einer früheren Meta-Analyse, von Wierzbicki und Pekarik (1993), die 125 Studien umfasste, konnten

signifikant unterschiedliche Raten von Drop-out abhängig von der Operationalisierung gezeigt werden.

Dass Drop-out trotz unterschiedlicher Definitionen mit negativen Konsequenzen einhergehen kann, konnte allerdings in vielen Studien gezeigt werden. Zum einen verlässt der Patient häufig die Therapie, ohne sich hinreichend verbessert zu haben (Knox et al., 2011, Saatsi, Hardy & Cahill, 2011). Dabei zeigten Swift und Greenberg (2012), dass vor allem schwerwiegendere Störungen von Therapieabbruch betroffen sein können. So zeigten Patienten mit einer Persönlichkeitsstörung eine Abbruchrate von 26%, Patienten mit einer Angststörung dagegen eine geringere Abbruchrate von 16%. Zum anderen kann auch der Therapeut durch Drop-out demotiviert werden, erneut Patienten zu behandeln (Piselli, Halgin & Macewan, 2011). Zudem bedeutet ein Therapieabbruch auch einen finanziellen Verlust für die Krankenkassen, die die Therapie bezahlen, sofern der Patient bis zum Abbruch nicht gebessert ist (Altmann et al., 2018).

Ob Drop-out allerdings in jedem Fall als therapeutischer Misserfolg gewertet werden kann, wurde von Altmann et al. (2014) kritisch diskutiert. Die Autoren untersuchten die Gründe für Therapieabbruch in einer Stichprobe von 584 Patienten mit unterschiedlichen Diagnosen. Sie unterschieden dabei zwischen problematischen und unproblematischen Drop-outs. Die problematischen Drop-outs zeigten dabei zum Ende der Therapie erhöhte Symptombelastung in den Bereichen Depressivität, Angst und Stress. Zudem galt der Großteil der Dropouts als problematisch. Unter der Drop-out-Rate von 44%, wurden 70% der Abbrüche als problematisch eingestuft. Es lässt sich also zeigen, dass zwar nicht alle, aber zumindest ein Großteil der Therapieabbrüche problematisch sein können und mit negativen Konsequenzen für den Patienten assoziiert sind.

Therapieerfolg kann folglich auf unterschiedliche Arten erfasst werden (kontinuierliche Werte, kategoriale Beurteilungen, Drop-out). Den kontinuierlichen Prä-Post- oder Gruppenvergleich zu nutzen, ist dabei die Methode, die am stärksten in der empirischen

Literatur vertreten ist. Um Forschungsergebnisse aber auch für Praktiker interpretierbarer zu machen, empfiehlt es sich, das Konzept der klinischen Signifikanz und der reliablen Veränderung von Jacobson und Truax (1991) zu nutzen. Darüber hinaus ist der Drop-out ein ebenfalls häufig untersuchtes Maß. Vergangene Studien konnten zeigen, dass Drop-out gravierende Folgen für Patienten und Therapeuten haben kann.

Eine Methode zur Erfassung von Grawes generellen Veränderungsmechanismen: der Berner Patienten- und Therapeutenstundenbogen

Messinstrumente zur Erfassung der therapeutischen Allianz allein sind zahlreich, wie etwa das Helping Alliance Questionnaire (Bassler, Potratz & Krauthauser, 1995) oder das Working Alliance Inventory (Horvath & Greenberg, 1989). Seltener sind Instrumente zur Erfassung von GCMs in der Literatur zu finden, wie etwa die Scale for the Multiperspective Assessment of General Change Mechanisms in Psychotherapy von Mander et al. (2013) oder der Individual Therapy Process Questionnaire von Mander et al. (2015). Die Publikationen in dieser Dissertation nutzen das Instrument, auf dem die Maße von Mander et al. (2013, 2015) aufbauen: den Berner Stundenbogen, welcher aus einem Patientenstundenbogen (PSTB) und einem Therapeutenstundenbogen (TSTB) besteht (Flückiger, Regli, Zwahlen, Hostettler & Caspar, 2010). Es ist daher naheliegender, von den Berner Stundenbögen zu sprechen und damit beide Teile anzusprechen. Die Stundenbögen sind bereits im Jahr 2000 entstanden und wurden seitdem in der Psychotherapie und der Psychotherapieforschung genutzt (Flückiger et al., 2010).

Die Berner Stundenbögen wurden mit dem Ziel entwickelt, den therapeutischen Prozess durch ROM zu erfassen. Zum einen sollte dies der Qualitätssicherung der Psychotherapie dienen und zum anderen sollten in den Stundenbögen die von Grawe formulierten GCMs erfasst werden. Ein weiterer Punkt der Stundenbögen ist, dass diese den Therapieprozess aus der Perspektive des Therapeuten und aus Sicht des Patienten abbilden.

An einer Stichprobe von 429 Patienten prüften Flückiger et al. (2010) die psychometrische Qualität der Stundenbögen, auf die im Folgenden eingegangen wird.

Der PSTB enthält 22 Items, die auf einer 7-stufigen Likert-Skala von -3 bis 3 aus Patientensicht beurteilt werden (-3 = *überhaupt nicht*, 0 = *weder noch*, 3 = *ja, ganz genau*). Der Fragebogen ist in 8 Dimensionen aufgeteilt: Therapiebeziehung, Selbstwerterfahrung, Bewältigungserfahrungen, Klärungserfahrungen, Therapiefortschritte, Aufgehoben sein, Direktivität des Therapeuten plus Kontrollerfahrungen, sowie Problemaktualisierung. Dabei enthält die Dimension Therapiefortschritte nur ein Item. Inhaltlich bestehen die Items aus einzelnen Aussagen, denen der Patient mehr oder weniger zustimmen kann. Ein Beispiel für ein Item aus der Dimension Therapiebeziehung lautet etwa, „Die Therapeutin und ich verstehen einander“ oder ein Item aus der Dimension Bewältigungserfahrungen lautet, „Ich traue mir jetzt mehr zu, meine Probleme aus eigener Kraft zu lösen“. Die interne Konsistenz der Skalen wurde für verschiedene Settings (Einzel- und Paarsetting), sowie für verschiedene Therapiephasen (Sitzungen 1-9, 10-19, >19) überprüft und ist als gut zu bewerten (Skalen $\alpha > 0,70$). Einzig die Skala Problemaktualisierung zeigte in den Sitzungen 1-9 eine unzureichende interne Konsistenz von $\alpha = 0,69$. Die konfirmatorische Faktorenanalyse zeigte einen akzeptablen Fit der Faktorenstruktur für den Fit-Index Root Mean Square Error of Approximation (RMSEA, $RMSEA < 0,08$). Der Standardized Root Mean Square Residual Index (SRMR, $SRMR < 0,08$) zeigte sich allerdings im Mehrpersonen-Setting auffällig ($SRMR = 0,099$). Der Goodnes of Fit Index (GFI, $GFI > 0,95$) und der Comparative Fit Index (CFI, $CFI > 0,95$) war in allen untersuchten Stichproben unzureichend. Die Konventionen für die verschiedenen Fit-Indizes können Schreiber, Amaury, Stage, Barlow und King (2010) entnommen werden. Die faktorielle Struktur des PSTB ist nach Schreiber et al. (2010) daher als nicht bestätigt zu bewerten.

Der TSTB enthält 27 Items. Die ersten 12 Items werden auf derselben 7-stufigen Likert-Skala beurteilt wie auch der PSTB. Die übrigen Items werden dagegen auf einer 6-

stufigen Likert-Skala von 0-5 beurteilt (0 = *überhaupt nicht*, 5 = *stimmt vollkommen*). Der TSTB wird von Therapeuten ausgefüllt und enthält 11 Dimensionen: Therapiebeziehung, Offenheit, Therapiefortschritte, Interaktionelle Schwierigkeiten, Problembewältigung, Bezug zur realen Lebenssituation, Motivationale Klärung, Ressourcenaktivierung, Problemaktualisierung, Anstrengungsbereitschaft und Interaktionelle Perspektive. Die Dimension interaktionelle Schwierigkeiten besteht im TSTB aus einem Item. Einige Items sind dabei äquivalent aus Therapeutesicht formuliert, z.B. „Die Patientin und ich verstehen einander“ aus der Dimension Therapiebeziehung, andere Items sind dagegen stärker umformuliert. Das Gegenstück zum Beispiel aus der PSTB-Bewältigungserfahrungsdimension findet sich in der Therapeutenversion in der Skala Problembewältigung wieder und lautet: „Ich habe heute darauf hingearbeitet, dass die Patientin sich einem bestimmten Problem besser gewachsen fühlen kann als bisher.“. Die interne Konsistenz ist auch im TSTB überwiegend als gut zu bewerten, allerdings zeigen sich hier mehr Auffälligkeiten im Cronbachs Alpha. Unzureichende Werte zeigen sich z.B. im Paarsetting in der Dimension motivationale Klärung, die Skala Anstrengungsbereitschaft zeigt dagegen in jeder Stichprobe außer dem Paarsetting einen unzureichenden Cronbachs Alpha Wert. Der TSTB zeigt sich auch in der konfirmatorischen Faktorenanalyse deutlich auffälliger als der PSTB. Auch im TSTB ist der GFI und der CFI unzureichend hoch, allerdings zeigen darüber hinaus der RMSEA und SRMR im Mehrpersonensetting unzureichend hohe Werte. Auch im TSTB kann die Faktorenstruktur also nicht bestätigt werden.

Flückiger et al. (2013) schlugen daher die am dualen Modell des therapeutischen Prozesses von Schulte (1996) orientieren GCMs interpersonale Erfahrungen und intrapersonale Erfahrungen vor und berücksichtigten zusätzlich die Problemaktualisierung. Im dualen Modell werden zwei interagierende Prozesse in der Therapie beschrieben: Beziehung und Methoden. Die Beziehung bezeichnet dabei das Verhältnis von Therapeut und Patient, das in dieser Arbeit zuvor mit der Allianz beschrieben wurde. Der Beziehungsprozess soll die

Therapiemotivation verstärken und den Patienten dazu anregen, das in der Therapie neu erlernte Verhalten zu zeigen (von Schulte [1996] Basisverhalten genannt). Dieses Verhalten ist letztlich wieder rückgekoppelt mit den therapeutischen Methoden. Im Methodenprozess passt der Therapeut die ihm zur Verfügung stehenden Techniken auf das individuelle Störungsmodell des Patienten an. Dadurch soll das beim Patienten vorliegende Problemverhalten bearbeitet werden.

Flückiger et al. (2013) begründeten den Fokus auf interpersonale und intrapersonale Erfahrung vor Allem damit, dass diese in der frühen Therapiephase besonders relevant sind. Übertragen auf das duale Modell von Schulte (1996) passen die interpersonalen Prozesse dabei zur Beziehung und die intrapersonalen Prozesse zu den Methoden. Flückiger et al. (2013) greifen zudem noch die Problemaktualisierung auf, die einen wichtigen Beitrag zur Symptomreduktion am Ende der Therapie leisten kann. Die Faktoren wurden dabei aus den Berner Stundenbögen extrahiert. Interpersonale Erfahrungen beinhalten in der Therapeutenversion die Faktoren Therapiebeziehung (global alliance) und Offenheit (openness) und in der Patientenversion die Faktoren Therapiebeziehung (global alliance) und Aufgehoben sein (contentment). Intrapersonale Erfahrungen enthalten die Dimensionen Problembewältigung (mastery) und motivationale Klärung (clarification) in der Therapeutenversion und Bewältigungserfahrung (mastery) und Klärungserfahrung (clarification) in der Patientenversion. Die Skala Problemaktualisierung wurde den beiden originalen Stundenbogen-Versionen entnommen. Diese extrahierten Sekundärskalen zeigen insgesamt eine gute interne Konsistenz mit Cronbachs-Alpha-Werten zwischen $\alpha = 0,74$ (intrapersonale Erfahrungen Therapeut) und $\alpha = 0,89$ (interpersonale Erfahrungen Patient). Allerdings findet sich keine Aussage zur genauen exploratorischen oder konfirmatorischen Faktorenanalyse, die zur Auswahl der Faktoren geführt hat. Ob die Faktoren interpersonale Erfahrungen, intrapersonale Erfahrungen und Problemaktualisierung in der Patienten- oder Therapeutenversion einen ausreichenden Fit auf die Daten erzeugen, bleibt daher unklar.

Auch die Frage, ob diese Sekundärskalenstruktur der ursprünglichen Primärskalenstruktur überlegen ist, kann nicht beantwortet werden.

Zielsetzung und Forschungsagenda

Im folgenden Abschnitt werden Zielsetzung und Forschungsagenda der jeweiligen Publikationen beschrieben. Ziel der Studien war es jeweils, den Einfluss von GCMs auf das Therapieergebnis in der CBT zu bestimmen.

Fragestellung Publikation 1 (Gmeinwieser, Kuhlencord, Ruhl, Hagmayer & Probst, 2019a): In vorausgegangenen Arbeiten konnte gezeigt werden, dass GCMs mit dem Therapieergebnis assoziiert sind (Flückiger et al., 2013; Mander et al., 2013; Rubel et al., 2017). Bisher wurde allerdings die frühe Veränderung von GCMs in der Therapie als möglicherweise wichtiger Prädiktor für den Erfolg eher vernachlässigt. Das erste Ziel der Studie war es, diese Lücke zu schließen und die frühen Veränderungen in den GCMs zu betrachten. Zudem wurden in früheren Studien lediglich kontinuierliche Outcomes untersucht und nicht die reliable Verbesserung nach Jacobson und Truax (1991). Das Therapieergebnis wurde daher in dieser Publikation kontinuierlich (Postmessung) und dichotom betrachtet (reliable Verbesserung vs. keine reliable Verbesserung). Ein weiteres Ziel der Publikation war es, den Einfluss von frühen Veränderungen in GCMs auf das Therapieergebnis zusätzlich zur frühen Veränderung in der Symptombelastung zu bestimmen. Das Verhältnis zwischen GCMs und der Symptombelastung wurde bisher nur für die Allianz stärker untersucht (z.B. Zilchamano et al., 2014). Zur Erfassung des Therapieergebnisses wurde dabei nicht nur die subjektive Symptombelastung herangezogen, sondern auch die Patientenbeurteilung der allgemeinen Lebenszufriedenheit. Zur Bestimmung der frühen Veränderung der GCMs und der Symptombelastung wurden die ersten drei ROM-Messungen herangezogen (Prä, Ende der Probatorik, 10. Therapiesitzung).

Fragestellung Publikation 2 (Gmeinwieser, Hagmayer, Pieh & Probst, 2019b): Grosse Holtforth, Grawe, Fries und Znoj (2008) untersuchten differenzielle Effekte intrapersonaler Erfahrungen auf das Therapieergebnis. In ihrer Studie wurden Patienten zufällig zwei Behandlungen zugeordnet. Eine Behandlung fokussierte stärker auf Bewältigungserfahrung und die andere Behandlung legte den Fokus stärker auf Bewältigungserfahrungen und die motivationale Klärung (diese bilden zusammen den Faktor interpersonale Erfahrungen). Es zeigte sich, dass die Therapie mit dem stärkeren Fokus auf Bewältigungserfahrung und motivationale Klärung zu einem besseren Therapieergebnis in den Maßen interpersonelle Probleme und retrospektiv erfasste Verbesserung führte. Dies war allerdings nur für diejenigen Patienten der Fall, die stärkere Vermeidungsziele angaben. Grawe (2004) beschrieb in seiner Konsistenztheorie den Unterschied zwischen Patienten, die eher Annäherungsziele und Patienten, die eher Vermeidungsziele fokussierten. Patienten sind in ihrem Handeln und Erleben demnach stärker inkongruent, je stärker sie Vermeidungsziele angeben, da diese langfristig zu mehr Leidensdruck führen würden. Zum Beispiel kann ein Patient das Vermeidungsziel haben, öffentliche Orte zu meiden. Dies widerspricht allerdings dem Ziel, soziale Kontakte zu knüpfen und fördert damit die Inkongruenz. Die Publikation von Gmeinwieser et al. (2019b) baut auf den Überlegungen von Grawe (2004) zur Inkongruenz und der Studie von Grosse Holtforth et al. (2008) auf. Ziel des Artikels war es, die Hypothese zu prüfen, dass frühe intrapersonale Erfahrungen abhängig vom Inkongruenzniveau des Patienten anders mit dem Therapieergebnis assoziiert sind. Als explorative Analysen sollten zudem geprüft werden, ob diese Interaktion auch für die frühen GCMs interpersonale Erfahrungen und Problemaktualisierung besteht.

Fragestellung Publikation 3 (Gmeinwieser, Schneider, Bardo & Hagmayer, in press): Ein großer Teil der Studien zu GCMs, zu denen auch die beiden zuvor genannten Publikationen (Gmeinwieser et al., 2019a; Gmeinwieser et al., 2019b) zählen, beschäftigten sich lediglich mit Patienten, die die Therapie erfolgreich beendet hatten (Flückiger et al.,

2013; Grosse Holtforth et al., 2008). Es wurden nur wenige Studien durchgeführt, die sich mit dem Zusammenhang von GCMs und Therapieabbruch beschäftigten (z.B. Flückiger et al., 2011; Kegel & Flückiger, 2015). Therapieabbruch ist allerdings, wie bereits beschrieben, ein ernstzunehmendes Problem für die Psychotherapie mit verschiedenen problematischen Konsequenzen (Knox et al., 2011; Altmann et al., 2018). Ziel dieser Studie war es daher, die differenziellen Effekte von GCMs für das Risiko eines Therapieabbruchs zu bestimmen. Die in dieser Studie verwendete Survival Analyse (SA) konnte berücksichtigen, ob sich Prädiktoren im Verlauf der Zeit und damit im Verlauf der Therapie veränderten (GCMs, Symptombelastung) oder ob diese konstant blieben (z.B. Behandlungsdiagnose, Alter bei Therapiebeginn) und wie sich diese Prädiktoren auf das Risiko auswirkten, die Therapie abubrechen. Dabei konnte für jeden Patienten der individuelle Therapieverlauf mit mehr oder weniger vorhandenem ROM-Sitzungen betrachtet werden. Die untersuchte Hypothese war, dass die GCMs prädiktiv für das Risiko eines Therapieabbruchs sind.

Die in dieser Arbeit genutzten Publikationen nutzen Teilstichproben derselben Gesamtstichprobe aus einer poliklinischen Institutsambulanz, dem Therapie- und Beratungszentrum des psychologischen Instituts der Universität Göttingen. Das Therapie- und Beratungszentrum ist eine Ausbildungsambulanz für psychologische Psychotherapie mit dem Schwerpunkt CBT. Die durchgeführten Therapien werden daher zum Großteil von Psychotherapeutinnen und Psychotherapeuten in Ausbildung durchgeführt. Voraussetzung in Deutschland für Psychotherapeuten in Ausbildung ist die Supervision durch einen approbierten Psychotherapeuten jede vierte ambulante Therapiesitzung und mindestens 1 ½ Jahre vorausgegangene klinische Erfahrung. Daten wurden im Zeitraum von 2005 bis 2017 in der Routineversorgung erhoben. Patienten erhielten vor dem Einstieg in die Therapie ein 30-minütiges Gespräch mit einem approbierten Psychotherapeuten und ein vollstandardisiertes, vollstrukturiertes Interview (Composite International Diagnostic Interview; WHO, 1990), das von geschulten wissenschaftlichen Hilfskräften durchgeführt wurde. Ziel dieser ersten Phase

war die Erfassung eines ersten diagnostischen Eindrucks und einer möglichen Indikation für Psychotherapie. Patienten wurden danach ihren behandelnden Therapeuten zugeteilt. Jeder Patient begann die Therapie mit initialen fünf Sitzungen, der sogenannten Probatorik. Diese fünf Sitzungen wurden nicht bei der Krankenkasse beantragt und standen jedem Patienten zu. Die Probatorik diente dazu, einen Behandlungsplan und feste Behandlungsdiagnosen zu formulieren. Zudem dienten die Sitzungen dazu, eine therapeutische Allianz zwischen Patient und Therapeut aufzubauen. Mit dem Behandlungsplan wurde daraufhin die Therapie bei der jeweiligen Krankenkasse des Patienten beantragt. Die Länge der Therapien konnte dabei je nach Diagnose und Schwere der Störung variieren. Nachdem die Krankenkasse die Therapie bewilligt hatte, konnte mit den Therapiesitzungen begonnen werden. In der ersten Probatoriksitzung wurde die Prämessung durchgeführt. Das Ende der Probatorik, fünf Sitzungen in die Therapie und danach jede zehnte Sitzung diente dem ROM. Da ROM als Empfehlung galt, wurde dieses nicht immer vollständig durchgeführt. Die Postmessung konnte bis zu fünf Sitzungen vor Beendigung der eigentlichen Therapie durchgeführten werden, um Therapeuten die Gelegenheit zu geben, die Diagnostik mit ihren Patienten zu besprechen. Prä- und Postmessung enthielten Messinstrumente, die unter anderem die Symptombelastung (Symptom Checklist 90, SCL-90R, Franke, 2002) und die Lebenszufriedenheit (Fragebogen zur Lebenszufriedenheit, Henrich & Herschbach, 2000) erfassten. ROM-Sitzungen enthielten einen Fragebogen zur Symptombelastung (Symptom Checklist 14, SCL-14, Harfst et al., 2002), sowie die Erfassung von GCMs mithilfe des PSTB und des TSTB (Flückiger et al., 2010). Die GCMs, die in den Publikationen untersucht wurden, waren interpersonale Erfahrungen, intrapersonale Erfahrungen und Problemaktualisierung.

Publikationen

Gmeinwieser, Kuhlencord, Ruhl, Hagmayer und Probst (2019a)

Gmeinwieser, S., Kuhlencord, M., Ruhl, U., Hagmayer, Y., & Probst, T. (2019). Early developments in general change mechanisms predict reliable improvement in addition to early symptom trajectories in cognitive behavioral therapy. *Psychotherapy Research*. Advance online publication. <https://doi.org/10.1080/10503307.2019.1609709>

Title:

Early developments in general change mechanisms predict reliable improvement in addition to early symptom trajectories in cognitive behavioral therapy

Running Title:

General change mechanisms predict outcome

Authors:

Sebastian Gmeinwieser¹, Merle Kuhlencord¹, Uwe Ruhl¹, York Hagmayer^{2*}, Thomas Probst^{3*}

* = shared senior authorship

Affiliations:

¹ Department for Clinical Psychology and Psychotherapy, Therapy and Counseling Center, University of Goettingen, Germany

² Department for Cognitive and Decision Sciences, University of Goettingen, Germany

³ Department for Psychotherapy and Biopsychosocial Health, Danube University, Krems, Austria

Corresponding author:

Sebastian Gmeinwieser, M.Sc.

Department for Clinical Psychology and Psychotherapy, Therapy and Counseling Center, University of Goettingen, Gosslerstrasse 14, D-37073, Germany

E-Mail: Sebastian.Gmeinwieser@psych.uni-goettingen.de

Abstract

General change mechanisms (GCMs) have been shown to predict treatment outcome in cognitive behavioral therapy (CBT). Early changes in GCMs (early-change-GCMs) and in symptom severity were also linked to treatment outcome, but evidence is still limited. We investigated whether early-change-GCMs from patients' and therapists' perspectives in addition to early changes in symptom severity predict treatment outcome. Data from a university-based outpatient clinic was analyzed (N=911 completers). Most treatments were provided by therapists in CBT training. Outcomes were symptom severity and life satisfaction. Continuous post-treatment scores and categories of reliable improvement (improved vs. not improved) were investigated. Some patient-rated early-change-GCMs and early changes in symptom severity proved to be predictors for continuous outcomes ($p < .05$) in symptom severity and life satisfaction. In addition, some patient-rated early-change-GCMs predicted reliable improvement in life satisfaction. By contrast, therapist-rated early-change-GCMs and early symptom severity change predicted reliable improvement in symptom severity. Early-change-GCMs explained between two and four percent additional variance in outcomes. Thus, it is important to consider early changes in GCMs in addition to early changes in symptom severity in routine monitoring.

Clinical or Methodological Significance of this Article:

Early changes in general change mechanisms (GCMs) predicted outcomes in addition to early changes in symptom severity. Therefore, it is advisable to monitor general change mechanisms and their progress on a regular basis.

Keywords:

cognitive behavioral therapy, outcome research, process research, general change mechanisms, early change

Word Count: 8498

Disclosure of interest

The authors report no conflict of interest.

Introduction

Psychotherapeutic treatment outcome is influenced by many factors, including disorder-specific interventions and common factors as well as their interactions (cf. Cuijpers, Reijnders & Huibers, 2019; Tschacher, Junghan & Pfamatter, 2014; Wampold, 2015). One conceptualization of common factors refers to general change mechanisms (GCMs, Grawe, 1997).

The therapeutic alliance is probably the most researched common factor and GCM. An early meta-analysis by Horvath and Symonds (1991) including 24 studies found a correlation between the alliance and psychotherapy outcome of $r = 0.26$ averaging over various outcome measures, rater perspectives, and types of treatment. An updated meta-analysis by Flückiger, Del Re, Wampold and Horvath (2018) including 295 studies found a similar correlation of $r = 0.278$. Seventy-two CBT studies were identified for the meta-analysis and a correlation of $r = 0.2$ was found which was not significantly different from other treatment types. The question whether the alliance influences symptom change or vice versa has also been addressed in some studies. Falkenström, Granström, und Holmqvist (2013) reported a reciprocal causal relationship between symptom severity and alliance. Zilcha-Mano, Dinger, McCarthy, and Barber (2014) found that alliance preceded symptom change. There is also some evidence that a strong alliance is differentially associated with the outcome depending on patients' characteristics (e. g., Lorenzo-Luaces, DeRubeis & Webb 2014; Probst et al., 2017).

According to the framework of GCMs provided by Grawe (1997), alliance belongs to interpersonal experiences (Flückiger, Grosse Holtforth, Znoj, Caspar, & Wampold, 2013). Besides interpersonal experiences, higher-order GCMs are intrapersonal experiences and problem actuation (Flückiger et al., 2013). Intrapersonal experiences comprises clarification of meaning and mastery. Mastery refers to developing the patients' coping mechanisms when experiencing difficult situations. Clarification of meaning means that a patient acquires new insights into his/her behaviors, plans, and motives. Problem actuation occurs when patients

experience problem-related emotions during psychotherapy sessions. There are several studies supporting the relevance of these GCMs for psychotherapy outcome (e. g., Flückiger et al., 2013; Mander et al., 2013; 2015; Rubel, Rosenbaun, & Lutz, 2017). Flückiger et al. (2013) found that patient- and therapist-rated interpersonal experiences in the early treatment phase predict the outcome even when controlling for early changes in symptom distress. Flückiger et al. (2013) measured early changes in symptom distress as individual slopes over the first five sessions of treatment.

We use early change as a general term including changes for the good or for the worse in early phases of treatment. A more common researched term that focuses solely on early improvements is early response. Early response is an important concept in many studies. It has been studied in a variety of settings including individual psychotherapy (Koffmann, 2018; Rubel, Lutz & Schulte, 2015), group treatment (Flood, Page & Hooke, 2018), and web-based interventions (Lutz et al., 2017) and was found to predict treatment outcome. The assessment of early response varies, however, from study to study. It ranges from daily changes (Flood, Page & Hooke, 2018), changes from session to session (Lutz et al., 2014) to changes after a fixed amount of sessions (Nordberg, Castonguay, Fisher, Boswell & Kraus, 2014) or weeks (Lutz, Stulz & Köck, 2009). Nordberg et al (2014) applied growth mixture modelling and found distinct groups of early responders, which predicted treatment outcome, dependent on their change-trajectories, such as high initial symptom clients with a rapid change or low symptom clients with no change. Studies that apply the clinical significance approach according to Jacobson and Truax (1991) to early response could also show that early response predicted treatment outcome (Flood et al., 2018). Furthermore, a comparison of growth mixture modelling and clinical change criteria by Flood et al. (2018) found the clinical significance method to be superior to growth mixture modelling. Additionally, studies that defined early response as 50% symptom severity improvement (Schlagert & Hiller, 2017) or used alliance as an early response measure (Zilcha-Mano & Errázuriz, 2017) could also show

that it predicts treatment outcome. These findings illustrate that the assessment of early change varies widely across studies and that early change has to be defined carefully.

The current study builds upon the one by Flückiger et al. (2013) by analyzing the interplay between early changes in GCMs, early changes in symptom severity, and outcome. We extend the findings by Flückiger et al. (2013) by analyzing cognitive behavioral therapy (instead of integrative therapy), by focusing on early changes in GCMs (early-change-GCMs; instead of mean GCMs in the early treatment phase) as predictors of CBT outcome, and by categorizing the outcome into reliable vs. non-reliable improvement (in addition to continuous post-treatment outcome).

To be more precise, our primary research question was whether early-change-GCMs predict treatment outcome in addition to early changes in symptom severity. As GCMs, we considered interpersonal experiences, intrapersonal experiences, and problem actuation. Our secondary research questions were (1) whether patient- and therapist rated early-change-GCMs are differentially predictive, (2) whether early-change-GCMs are differently predictive when the outcome is measured as continuous post-treatment outcome or as reliable improvement, and (3) whether early-change-GCMs are differentially predictive for life satisfaction and symptom severity as outcomes. Based on the findings of Flückiger et al. (2013), we hypothesized that early-change-GCMs in addition to early changes in symptom severity predict continuous post-treatment outcome as well as reliable improvement for symptom severity and life satisfaction.

Method

Setting

Therapies were delivered at a therapy and counseling center of the University of Goettingen, Germany. The center is an outpatient clinic for patients with mental health problems. It is a part of the Institute of Psychology and is involved in the post-graduate training program in CBT. Licensed psychotherapists and psychologists in post-graduate

training both deliver treatment. Requirements for trainees to give treatment are a diploma or master degree in psychology and at least 1.5 years of previous psychotherapeutic experience. Supervision is mandatory for trainees with one supervision session every four treatment sessions.

Participants

Data of 1,129 patients, who completed treatment at the therapy and counseling center between 2005 and 2017 and did not drop out at any point of treatment (36% drop-out), were considered. Patients, who dropped out, did not significantly differ from patients who completed treatment in outcome variables at baseline. Drop-outs were excluded because they were less likely to complete treatment monitoring. Of the 1,129 patients, all patients were excluded, who had more than 20% missing raw values in pre-treatment outcome measures or in post-treatment outcome measures. The resulting sample consisted of 911 patients treated by 145 therapists.

Patients' mean age at pre-treatment was $M = 39.29$ years ($SD = 12.91$). Female were 544 (59.7%) patients. Only one diagnosis for treatment was assigned to 338 (41.1%) patients by the attending therapist, 287 (34.9%) were given one comorbid disorder, and 197 (24%) had three or more diagnoses. All diagnoses were based on the ICD-10 (WHO, 1992). The most common types of diagnoses were affective disorders (521 patients, 57.2%) closely followed by anxiety disorders (480 patients, 52.7%). Four-hundred-fifty-nine (50.4%) patients had a comorbid affective disorder. Patients' and therapists' demographics and baseline values of outcomes with means and standard deviations in parentheses are displayed in Table 1.

Measures

Standardized questionnaires were used to assess symptom severity, life satisfaction, and GCMs.

Symptom severity: Symptom Checklist (SCL-90R; Derogatis, 1977; German version Franke, 2002). The SCL-90R is a 90-item patient-rated measure consisting of items assessing

various areas of symptom severity on a 5-point-likert-scale ranging from 0 (*not at all*) to 4 (*extremely*). The Global Severity Index (GSI) of the German SCL was used to measure psychological symptom severity. It is the mean of all items. A reliability analysis of the GSI of the SCL-90R showed a Cronbach's alpha of $\alpha = 0.97$ in a representative German sample (Franke, 2002) and in our sample. The SCL-14, a German short version consisting of 14 Items (Harfst et al., 2002) has an acceptable internal consistency ranging from $\alpha = 0.81$ (somatization) to $\alpha = 0.89$ (depression; Prinz et al., 2008). In our sample the GSI of the SCL-14 had a Cronbach's alpha value of $\alpha = 0.86$. The long version of the SCL was used at pre- and post-treatment, while the short form was used in the monitoring sessions (treatment sessions 5 and 10) during treatment.

Life Satisfaction: Questions on Life Satisfaction^{Modules} (Fragebogen zur Lebenszufriedenheit, QLS^M, German questionnaire by Henrich & Herschbach, 2000). The QLS^M consists of two 8 item modules. For the present study, only the module on life satisfaction was considered. Items are first rated for importance and then on the degree of satisfaction. Importance and satisfaction are rated by patients on a 5-point-likert scale ranging from *not important* to *very important* and *not satisfied* to *very satisfied* respectively. Every satisfaction rating is then weighted by its corresponding importance rating. The sum of the weighted ratings is the index of General Life Satisfaction (GLS). Index values range from -96 to 160. In a study by Daig et al. (2011) the module on life satisfaction showed an acceptable reliability of $\alpha = 0.83$ in a representative German sample. In our sample the internal consistency was $\alpha = 0.68$. The QLS^M was administered to the patients at pre- and post-treatment.

GCMs: Bern Post Session Reports for Patients and Therapists, short form 2000 (BPSR-Patient/Therapist; Flückiger, Regli, Zwahlen, Hostettler & Caspar, 2010). The BPSR has a patient and a therapist version assessing GCMs according to Grawe (1997), and inquires about the last attended therapeutic session. The patient version has 22 items rated on a 7-

point-likert-scale ranging from -3 (*not at all*) to 3 (*very much*). The therapist version consists of 27 items. Items 1-12 are also rated on a 7-point-likert-scale, items 13-27 are rated on a 5-point-likert-scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*). Since some of the subscales had problems with internal consistency, Flückiger et al. (2013) suggested the secondary scales interpersonal experiences, intrapersonal experiences, and problem actuation. In the patient version, the interpersonal experiences scale consists of the alliance and contentment scales. In the therapist version, it comprises the alliance and the openness scales. The intrapersonal experiences scale consists of the mastery and the clarification of meaning scales in both, the patient and the therapist version. Internal consistency analyses by Probst et al. (2017b) and Flückiger et al. (2013) showed acceptable Cronbach's alpha values (lowest $\alpha = .74$, highest $\alpha = .89$). In our sample consistencies ranged between $\alpha = .71$ and $\alpha = .85$. The BPSR was administered after the 5th and 10th session.

Procedure

All data used for this study was collected during routine practice. Patients first contacted the center through telephone. Demographic data and a problem description were obtained at this point. The second step was a 30-minute-session with a licensed therapist to explore the patient's problems and to provide information on treatment at the center. When patients decided to sign up for treatment, they were asked for their written informed consent to use their routinely collected data anonymously for research purposes. At the end of the session potential diagnoses were assigned. Next, a trained research assistant conducted the Composite International Diagnostics Interview (CIDI, WHO, 1990). Then, the patient was assigned to a therapist, most of which were in post-graduate training.

The first five sessions in Germany are dedicated to make a comprehensive diagnostic assessment, to find out whether there is any indication for treatment, and to decide whether therapist and patient agree on working together. During this period, a written assessment and treatment proposal has to be submitted to the patient's health insurance company to secure

funding. Upon approval, the treatment starts officially. It is terminated, when patient and therapist agree to end the treatment (i.e., completion of therapy), the last session funded by the patient's health insurance ends, or the patient unilaterally ends the treatment (i.e., drops out). In the present study only data from completers was used.

Data was collected routinely at various points during treatment (see Table 2). Pre-treatment measures were administered after the first session and consisted of the SCL-90R and the QLS^M. The fifth session was the first monitoring session at which the SCL-14 and the Bern Post Session reports were assessed. At this point, therapists had to decide on at least one diagnosis for treatment. The second monitoring session was the 10th session. After this, therapists were advised to conduct a monitoring session every 10th session. Post-treatment assessments were made five sessions before the treatment ended to give therapists the chance to discuss the outcome with their patient. The SCL-90R and the QLS^M were administered and the therapist re-evaluated the given diagnoses.

Outcomes

Post-treatment outcome. We used the post-treatment self-assessments of symptom severity and life satisfaction as continuous outcomes.

Reliable improvement. Reliable improvement was defined by a reliable change for the better according to the RCI (Jacobson & Truax, 1991). No improvement was defined as a non-reliable change or a reliable change for the worse. For the GSI of the SCL-90R, the RCI was ± 0.43 (Schauenburg & Strack, 1998). In the current sample, 41.6% of patients improved reliably. For the GLS of the QLS^M the RCI was calculated based on the pre-treatment data of the present study. The RCI was 40.03. According to this RCI, 45.8% of patients improved their life satisfaction reliably.

Early change

Early symptom change. For symptom severity early change, the difference between the GSI (SCL-90R) at pre-treatment and the GSI (SCL-14) at the second monitoring session

(10th session) was calculated. We used the pre-treatment session because many therapists did not collect data on symptom severity after the 5th session. A positive difference indicates an improvement of symptom severity over the course of time.

Early-change-GCMs. For early change of GCMs the difference between the secondary scales of the BPSR at the second monitoring session (10th session) and these scales at the first monitoring session (5th session) was calculated. We measured the early-change-GCMs this way because neither patients nor therapists rated GCMs after the first session. A positive difference indicates an improvement in GCMs.

For consistency, we used continuous difference scores for early symptom change (GSI) as well as early-change-GCMs (BPSR).

Statistical Analysis

Statistical analyses were conducted using R (R Core Team, 2017) and the packages “psych” (Revelle, 2017), “car” (Fox & Weisberg, 2011), “lme4” (Bates, Maechler, Bolker & Walker, 2015), “lmerTest” (Zeileis & Hothorn, 2002), “fmsb” (Nakazawa, 2018), “rcompanion” (Mangiafico, 2018) and “QuantPsyc” (Fletcher, 2012).

Model analyses and selection. For continuous outcomes, we used single-level and multilevel modeling because patients were nested within therapists. We compared a single-level fixed effect model consisting only of covariates with a model also containing all predictor variables and the interaction between predictor variables and the number of sessions. As a third model, we fitted a multilevel random intercept model with covariates, predictors, and interactions. We did so to test for therapist effects by comparing the multilevel models to the matching single-level fixed effect model. For comparison, we used likelihood-ratio tests and the Akaike information criterion (AIC). As effect size, we used R^2 and Nagelkerke’s R^2 . Here, we report the single-level models in detail, because our results showed a worse fit for the multilevel random intercept models for both continuous outcomes.

For reliable improvement as a binary outcome (improvement vs. no improvement), binary logistic regression was used. We compared a model including only covariates with a model containing all covariates and predictor variables and the interaction between number of sessions and predictors. We refrained from calculating a multilevel model because it did not provide a good fit for the continuous outcomes. For model comparisons, the AIC and likelihood-ratio tests were used. As effect size, we used Nagelkerke's pseudo- R^2 and Cox and Snell's pseudo- R^2 .

For all models, multicollinearity was tested using Pearson correlation coefficients (see supplement Table 1 for a full correlation matrix) and variance inflation factors.

Variables. The following variables were used as covariates to compute a first model: Number of diagnoses, initial symptom severity (pre GSI), initial life satisfaction (pre GLS), patients' age and sex, initial interpersonal, initial intrapersonal experiences and initial problem actuation from patients' and therapists' perspectives (5th session BPSR values), number of sessions attended, and whether at least one problematic diagnosis was assigned (substance dependency, substance abuse, schizophrenia and related disorders, bipolar disorder, personality disorder). Diagnoses were classified as problematic based on previous research showing these diagnoses to be negatively related to treatment outcome (Barrett et al., 2008; Swift & Greenberg, 2012). We also used early changes in symptom severity as well as its interaction with number of sessions as a covariate.

The following variables were analyzed as predictors: early changes in interpersonal experiences, early changes in intrapersonal experiences, and early changes in problem actuation rated by therapists and patients. In addition, we included their interactions with the number of sessions.

Missing values for variables collected in monitoring sessions (SCL-14 and BPSR) were replaced by the last completed observation. Thus, SCL-14 missing values at the 10th session were replaced by values from pre-treatment SCL-90R. BPSR missing values at the

10th session were replaced by BPSR values from the 5th session. Missing data of the outcome variables (QLS^M and SCL-90R) were handled as follows: Only patients who had less than 20% missing raw values in the pre/post outcome measures were included. GLS missing values were replaced with the median from the same category (e.g. missing values in the importance rating were replaced with the median of all importance ratings of the respective patient). The GSI of the SCL was calculated by removing missing values.

Results

Covariates and predictors showed no multicollinearity after we centered the variables according to Robinson and Schumacker (2009).

Outcome 1: Symptom severity

Continuous post-treatment outcome. Detailed results are shown in Table 3. The single-level model including covariates and predictors had the best fit for our data when compared to the model only including covariates ($\chi^2(12) = 38.3, p < 0.001$) and when compared to a multilevel random intercept model ($\chi^2(1) = 217.33, p < 0.001$). It had the lowest AIC value (model with covariates AIC = 689, model with covariates and predictors AIC = 675, random intercept model with covariates and predictors AIC = 894). R^2 of the model containing covariates and predictors was $R^2 = 0.48$ (Nagelkerke $R^2 = 0.59$), of the model only containing covariates $R^2 = 0.44$ (Nagelkerke $R^2 = 0.55$). Therefore, we proceeded with the single-level model including covariates and predictors. Significant covariates were initial symptom severity, patient age, initial patient-rated intrapersonal experience and early changes in symptom severity. Predictors showed the following results. Early changes in patient-rated intrapersonal experiences were significant ($\beta = -0.12, p = 0.003$). This indicates that an early increase in intrapersonal experiences was associated with lower symptom severity at post-treatment. In addition, the interactions of early changes in patient-rated interpersonal and intrapersonal experiences with number of sessions showed significant results with beta values of $\beta = -0.009$ ($p = 0.005$) for early changes in interpersonal and $\beta =$

0.005 ($p < 0.001$) for intrapersonal experiences. A closer look revealed that a shorter treatment was associated with a better outcome especially for patients experiencing a lower early change in interpersonal experience (see supplement Figure 1) and a higher early change in intrapersonal experience (see supplement Figure 2). None of the therapist-rated change-GCMs showed significant results.

Reliable improvement. Table 4 shows the results of the binary logistic regression analysis. A comparison of the model with early changes in GCMs as predictors and covariates with the model only including covariates proved that considering early changes in GCMs improved the predictions of reliable improvement in symptom severity, $\chi^2(12) = 31.07$, $p < 0.001$ (AIC of 620 and 627 respectively). Nagelkerke's pseudo- R^2 for the model containing covariates and predictors was $R^2 = 0.71$, (Cox and Snell's pseudo- $R^2 = 0.60$), of the model only including covariates $R^2 = 0.69$ (Cox and Snell's pseudo- $R^2 = 0.58$). Three covariates were statistically significant predictors of reliable improvement in symptom severity: initial symptom severity ratings, early changes in symptom severity and patient age. Initial ratings of GCMs showed no significant results. The following were significant predictors of reliable improvement: Early changes in therapist-rated interpersonal ($OR = 1.691$, $p = 0.011$) as well as early changes in therapist-rated intrapersonal experiences ($OR = 0.596$, $p = 0.025$). A closer look into early changes in therapist-rated intrapersonal experiences showed that not a decrease in intrapersonal experiences was associated with reliable improvement, as the result may suggest. It turned out that a smaller increase in therapist-rated intrapersonal experiences was more positively associated with reliable improvement. Changes in patient-rated intrapersonal experiences showed a significant interaction with number of sessions ($OR = 0.98$, $p = 0.005$). A closer look into the interaction showed that reliable improvement was associated with a stronger increase in intrapersonal experiences in shorter treatments and a lower increase or decrease in longer treatments. No other early-change-GCMs were significant predictors.

Outcome 2: Life satisfaction

Continuous post-treatment outcome. Results can be seen in Table 3 right hand side. The single-level model containing covariates and predictors had the best fit for our data compared to the model containing only covariates ($\chi^2(12) = 38.65, p < 0.001$). When compared to a multilevel random intercept model the test showed no significant difference ($\chi^2(1) = 0.43, p = 0.51$). We proceeded with the single-level model because it was the less complex of the two and had the overall lowest AIC (model with covariates AIC = 6010, model with covariates and predictors 5995, random intercept model with covariates and predictors AIC = 5996). R^2 of the model containing covariates and predictors was $R^2 = 0.39$ (Nagelkerke's $R^2 = 0.39$) of the model only containing covariates $R^2 = 0.35$ (Nagelkerke's $R^2 = 0.35$). Number of diagnoses, initial life satisfaction, initial patient-rated inter- and intrapersonal experiences, and early changes in symptom severity were significant covariates. Significant predictors were early changes in patient-rated interpersonal ($\beta = 0.117, p = 0.003$, that is, an increase in interpersonal experiences predicted a higher life satisfaction at post-treatment) and early changes in patient-rated intrapersonal experiences ($\beta = 0.11, p = 0.01$, that is, an increase in intrapersonal experiences predicted higher life satisfaction at post-treatment). The interaction of early changes in therapist-rated interpersonal experiences and number of sessions was also significant with $\beta = -0.097 (p = 0.004)$. A closer look into the interaction indicated that for longer treatments life satisfaction was lower especially for lower changes in patient-rated intrapersonal experiences. No other predictors showed significant results.

Reliable improvement. Results are displayed in Table 5. The model comparison showed that early changes in GCMs significantly improved the prediction ($\chi^2(12) = 23.32, p = 0.025$) but the best model according to AIC was the model only including covariates (model containing covariates AIC = 777 and AIC = 778 model containing covariates and predictors). Nagelkerke's pseudo- R^2 of the model containing covariates and predictors was $R^2 = 0.58$ (Cox

and Snell's pseudo- $R^2 = 0.49$), of the model only including covariates $R^2 = 0.56$ (Cox and Snell's pseudo- $R^2 = 0.47$). Therefore, we considered the model including covariates and predictors to be the best model. The analysis found the following statistically significant covariates: Initial ratings of life satisfaction, number of diagnoses and initial patient-rated interpersonal and intrapersonal experiences. Significant predictors were early changes in patient-rated inter- and intrapersonal experiences ($OR = 1.622, p = 0.014$ and $OR = 1.374, p = 0.04$ respectively). The results indicate that an early improvement in inter- and intrapersonal experiences was associated with reliable improvement. Considering the interaction terms the only significant interaction was between number of sessions and early changes in patient-rated interpersonal experiences with an odds ratio of $OR = 1.019 (p = 0.049)$. A closer look indicated that an early improvement in patient-rated interpersonal experiences was more strongly associated with a reliable improvement for patients receiving a higher number of sessions. By contrast, early changes in therapist-rated GCMs showed no significant results.

Discussion

We investigated the relationship between early-change-GCMs, early change in symptom severity, and CBT outcomes. As outcome, we used continuous post-treatment scores and categories of reliable improvement (i.e. improved vs. not-improved according to the RCI). The primary research question was if early-change-GCMs predict treatment outcomes in addition to early change in symptom severity. To answer this question, we controlled for a number of covariates including patient variables (age and sex), the number of diagnoses assigned for treatment, and initial ratings of general change mechanisms. Our findings indicate that early changes in at least some GCMs do predict CBT outcome in addition to early changes in symptom severity. They did so regardless of whether outcome was measured continuously or as reliable improvement. They also did so with respect to reductions in symptom severity and improvements in life satisfaction. However, which early-change-GCMs were predictive varied depending on the outcome used (see below). The only

exception were early changes in problem actuation, which were never predictive. Effect sizes turned out to be rather small in all analyses, with early changes explaining an additional two to four percent of variance.

With regard to our first secondary research question, we found that patient-rated and therapist-rated early-change-GCMs differentially predicted outcomes. Patient-rated (but not therapist-rated) early changes in interpersonal experiences predicted continuous post-treatment life satisfaction scores and reliable improvement in life satisfaction. By contrast, therapist-rated (but not patient-rated) early changes in inter- and intrapersonal experiences predicted reliable improvement in symptom-severity. Effect sizes were small (patient- and therapist-rated interpersonal, therapist-rated intrapersonal experiences) or less than small (patient-rated intrapersonal experiences) according to Borenstein, Hedges, Higgins and Rothstein (2009).

The results summarized in the previous paragraph are also relevant for our other secondary research questions. They indicate that early-change-GCMs were differently predictive for the different outcomes. Most notably, early patient-rated positive changes in inter- and intrapersonal experiences were consistently related to more positive outcomes in life-satisfaction, regardless whether life satisfaction was analyzed continuously or as reliable improvement. By contrast, findings for symptom severity were more mixed. Considering continuous post-treatment outcomes, patient-rated early changes in intrapersonal experiences were predictive, but this relation was qualified by an interaction with the number of treatment sessions. With respect to reliable improvement, therapist-rated early changes in inter- and intrapersonal experiences were found to be predictors.

These findings indicate that life satisfaction is not just reversed symptom severity but a concept by itself. They also point out that more research is needed to better understand the relation between early-change-GCMs and different treatment outcomes.

Our findings are partially in line with previous results. Considering early changes in symptom severity, we could replicate previous findings showing that it predicts and probably influences treatment outcome when outcome is measured by reductions in symptom severity or distress (Haas, Hill, Lambert & Morrell, 2002; Koffmann, 2018; Lambert, 2005; Lutz et al. 2009; Lutz et al., 2017; Rubel et al. 2015; Schindler, Hiller & Witthöft, 2013; Schlagert & Hiller, 2017). Going beyond these findings, our results showed that this effect is less clear for life satisfaction as an outcome. In the study we found no significant effect of early changes in symptom severity on reliable improvement in life satisfaction. Further research will have to investigate this point again.

With respect to GCMs, our findings partially agree with findings of previous research. Mander et al. (2013) found that all GCMs according to Grawe (1997), except problem actuation were correlated with the outcome in a sample of 269 inpatients. We clearly replicated the finding for problem actuation. However, Mander and colleagues did not consider early-change-GCMs. Flückiger et al. (2013) investigated the relationship between the means of early GCMs, intake distress and treatment outcome in an outpatient setting. They found that patient- and therapist-rated interpersonal experiences predicted the outcome for different distress-measures. When conducting a similar analysis on our data (see supplement Table 2), we found that the mean of early patient-rated inter- and intrapersonal experiences predicted post-treatment symptom severity, but therapist-rated GCMs did not. Also in contrast to our findings, early changes in symptom severity did not predict the outcome in the study by Flückiger et al. (2013) when GCMs were taken into account. Note that Flückiger et al. (2013) used a multilevel analysis with post-treatment distress as outcome, while we used initial GCM-levels as covariates and early changes as predictors. Although we fitted multilevel models to our data we continued with single-level models because the resulting multilevel models turned out to be a worse fit for our data.

Limitations

There are some limitations of the present study. First, we limited our analysis to treatment completers as did Flückiger et al. (2013). A substantial number of patients dropped out during treatment. The drop-outs might show other GCMs than the completers and this might affect the process outcome associations and might bias the results (Flückiger et al., 2013).

A second limitation is that we used reliable improvement based on the RCI as a categorical outcome. We first planned to have four categories of outcomes (reliable deterioration, no change, reliable improvement, clinically significant improvement). As reliable deterioration was very rare in our sample (< 5%) we decided for the binary outcome reliable improvement vs. no reliable improvement. Another argument for this binary outcome variable is that it divided our sample evenly into two groups. It should, however, be kept in mind that for clinicians it might be more informative if the predicted categories were post-treatment scores under or above the cut-off of clinical significance. To add to the body of research on treatment failure (Lilienfeld, 2007; Shimokawa, Lambert & Smart, 2010), it would be interesting to analyze an even larger sample to investigate the predictive value of early-change-GCMs and early change in symptom severity for reliable deterioration.

Another limitation is that the present study was based on field data and had no control group. Therefore, it is unclear whether improvements in symptom severity or life satisfaction were due to the given treatment including GCMs or other factors. Patients in the sample, however, had to wait on average 7.4 month before treatment started and still a treatment was indicated. Hence, the patients included in the present study at least suffered from enduring problems.

A further limitation is that post-treatment outcome ratings were collected before the treatment officially ended (up to five sessions before) and that no follow-up data were available for analysis. The early assessment of post-treatment ratings is an internal policy of

the center to enable therapists to address a patient's ratings in the final session. This limitation entails that no later changes for the better or worse could be considered in our analyses.

In addition, missing data was a concern in our data set. We decided to carry the last observation forward for monitoring sessions. Using this method we could include more patients into our analyses but at the same time we were unable to distinguish between patients whose GCMs did not change over time and patients who were only missing an observation. Also, we had not enough data for the GSI at session five, so we decided to measure early change by analyzing the change from the pre-treatment session and the 10th session. This means that early symptom change scores were derived from the GSI of the long form of the SCL and of the short version of the SCL. We had to measure early-change-GCMs differently (change from the 5th to the 10th session) because GCMs were not assessed at pre-treatment. This means that early GSI changes and early-change-GCMs are related to different time intervals in the study at hand.

It is also important to emphasize that our findings only indicate associations and not causal relations.

Conclusions and Implications

The present study showed that early-change-GCMs were differentially related to outcomes in a large sample of patients receiving CBT in an outpatient treatment center. Especially early changes in inter- and intrapersonal experiences predicted reductions in symptom severity and improvements in life satisfaction when they were patient-rated. They did so even when early changes in symptom severity were taken into account. Given that our findings only partially corroborate findings of other studies, future research is clearly needed. The practical implications of these findings are that therapists should consider assessing GCMs regularly at least in the early treatment phase from the patient's and therapist's perspective. A further clinical implication is that CBT therapists should not rely on their own estimation of early-change-GCMs as the patients' ratings of early-change-GCMs were more

consistent predictors of the outcome. Therapists should ask their patients on feedback on early-change-GCMs or they could use routine outcome monitoring (ROM) systems. These ROM systems usually include symptom measures and measures on interpersonal experiences such as alliance (Boswell, Kraus, Miller & Lambert, 2015). Our results indicate that it might be fruitful to add measures on intrapersonal experiences to such ROM systems. In addition, different outcomes should be used that are not solely deficit oriented. Measures on well-being and life satisfaction should also be administered to foster a more holistic view of the patient's impairment. As patients with less early symptom improvement and less positive change-GCMs have a less favorable treatment outcome, it is important to identify these patients as early as possible to optimize the treatment.

References

- Barrett, M. S., Chua, W.-J., Crits-Christoph, P., Beth Gibbons, M., Casiano, D., & Thompson, D. (2008). Early Withdrawal from Mental Health Treatment: Implications for Psychotherapy Practice. *Psychotherapy: Theory, Research, Practice, Training*, 45(2), 247-267. <http://dx.doi.org/10.1037/0033-3204.45.2.247>
- Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. <http://dx.doi.org/10.18637/jss.v067.i01>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis* (Reprinted.). Chichester: Wiley.
- Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. *Psychotherapy Research*, 25(1), 6–19. <https://doi.org/10.1080/10503307.2013.817696>
- Cuijpers, P., Reijnders, M., & Huibers, M. J. H. (2019). The Role of Common Factors in Psychotherapy Outcomes. *Annual Review of Clinical Psychology*. Advance online publication. <https://doi.org/10.1146/annurev-clinpsy-050718-095424>
- Daig, I., Spangenberg, L., Henrich, G., Herschbach, P., Kienast, T., & Brähler, E. (2011). Alters- und geschlechtsspezifische Neunormierung der Fragen zur Lebenszufriedenheit (FLZ^M) für die Altersspanne von 14 bis 64 Jahre. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 40(3), 172–178. <https://doi.org/10.1026/1616-3443/a000099>
- Derogatis, L. R. (1977). *SCL-90-R, administration, scoring & procedures manual-I for the R(evised) version*. Johns Hopkins University of Medicine
- Falkenström, F., Granström, F., & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology*, 60(3), 317–328. <https://doi.org/10.1037/a0032258>

- Fletcher, T. D. (2012). QuantPsyc: Quantitative Psychology Tools. R package Version 1.5.
<https://CRAN.R-project.org/package=QuantPsyc>
- Flood, N., Page, A., & Hooke, G. (2018). A comparison between the clinical significance and growth mixture modelling early change methods at predicting negative outcomes. *Psychotherapy Research*, 1–12. <https://doi.org/10.1080/10503307.2018.1469803>
- Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, 55(4).
<https://doi.org/10.1037/pst0000172>
- Flückiger, C., Grosse Holtforth, M., Znoj, H. J., Caspar, F., & Wampold, B. E. (2013). Is the relation between early post-session reports and treatment outcome an epiphenomenon of intake distress and early response? A multi-predictor analysis in outpatient psychotherapy. *Psychotherapy Research*, 23(1), 1–13. <https://doi.org/10.1080/10503307.2012.693773>
- Flückiger, C., Regli, D., Zwahlen, D., Hostettler, S., & Caspar, F. (2010). Der Berner Patienten- und Therapeutenstundenbogen 2000. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 39(2), 71–79. <https://doi.org/10.1026/1616-3443/a000015>
- Fox, J. & Weisberg, S. (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>
- Franke, G. H. (2002). *SCL-90-R: Symptom Checkliste von L.R. Derogatis (deutsche Version)*. Goettingen: Beltz.
- Grawe, K. (1997). Research-Informed Psychotherapy. *Psychotherapy Research*, 7(1), 1–19.
<https://doi.org/10.1080/10503309712331331843>
- 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>

- Harfst, T., Koch, U., Kurtz von Aschoff, C., Nutzinger, D. O., Rüdell, H., & Schulz, H. (2002). Entwicklung und Validierung einer Kurzform der Symptom Checklist-90-R. *DRV-Schriften*, 33, 71-73.
- Henrich, G., & Herschbach, P. (2000). Questions on Life Satisfaction (FLZM) - A Short Questionnaire for Assessing Subjective Quality of Life. *European Journal of Psychological Assessment*, 16(3), 150–159. <https://doi.org/10.1027//1015-5759.16.3.150>
- Horvath, A. O., & Symonds, D. (1991). Relation Between Working Alliance and Outcome in Psychotherapy: A Meta-Analysis. *Journal of Counseling Psychology*, 38(2), 139–149. <https://doi.org/10.1037/0022-0167.38.2.139>
- 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>
- Koffmann, A. (2018). Has growth mixture modeling improved our understanding of how early change predicts psychotherapy outcome? *Psychotherapy Research : Journal of the Society for Psychotherapy Research*, 28(6), 829–841. <https://doi.org/10.1080/10503307.2017.1294771>
- Lambert, M. J. (2005). Early response in psychotherapy: further evidence for the importance of common factors rather than "placebo effects". *Journal of clinical psychology*, 61(7), 855–869. <https://doi.org/10.1002/jclp.20130>
- Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on psychological science : a journal of the Association for Psychological Science*, 2(1), 53–70. <https://doi.org/10.1111/j.1745-6916.2007.00029.x>
- Lorenzo-Luaces, L., DeRubeis, R. J., & Webb, C. A. (2014). Client characteristics as moderators of the relation between the therapeutic alliance and outcome in cognitive

- therapy for depression. *Journal of Consulting and Clinical Psychology*, 82(2), 368–373. <https://doi.org/10.1037/a0035994>.
- Lutz, W., Arndt, A., Rubel, J., Berger, T., Schröder, J., Späth, C., . . . Moritz, S. (2017). Defining and Predicting Patterns of Early Response in a Web-Based Intervention for Depression. *Journal of Medical Internet Research*, 19(6), 206. <https://doi.org/10.2196/jmir.7367>
- Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., . . . Barlow, D. H. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology*, 82(2), 287–297. <https://doi.org/10.1037/a0035535>.
- Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*, 118(1-3), 60–68. <https://doi.org/10.1016/j.jad.2009.01.019>
- Mander, J., Schlarb, A., Teufel, M., Keller, F., Hautzinger, M., Zipfel, S., . . . Sammet, I. (2015). The individual therapy process questionnaire: Development and validation of a revised measure to evaluate GCMs in psychotherapy. *Clinical Psychology & Psychotherapy*, 22(4), 328–345. <https://doi.org/10.1002/cpp.1892>
- Mander, J. V., Wittorf, A., Schlarb, A., Hautzinger, M., Zipfel, S., & Sammet, I. (2013). GCMs in psychotherapy: Multiperspective assessment and relation to outcome. *Psychotherapy research*, 23(1), 105–116. <https://doi.org/10.1080/10503307.2012.744111>
- Mangiafico, S. (2018). *Rcompanion: Functions to Support Extension Education Program Evaluation*. R package version 2.0.3. Retrieved from <https://CRAN.R-project.org/package=rcompanion>
- Nakazawa, M. (2018). *Fmsb: Functions for Medical Statistics Book with some Demographic Data*. R package version 0.6.3. Retrieved from <https://CRAN.R-project.org/package=fmsb>

- Nordberg, S. S., Castonguay, L. G., Fisher, A. J., Boswell, J. F., & Kraus, D. (2014). Validating the rapid responder construct within a practice research network. *Journal of Clinical Psychology, 70*(9), 886–903. <https://doi.org/10.1002/jclp.22077>
- Prinz, U., Nutzinger, D., Schulz, H., Petermann, F., Braukhaus, C., & Andreas, S. (2008). Die Symptom-Checkliste-90-R und ihre Kurzversionen: Psychometrische Analysen bei Patienten mit psychischen Erkrankungen. *Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin, 18*(6), 337–343. <https://doi.org/10.1055/s-0028-1093323>
- Probst, T., Jakob, M., Kaufmann, Y. M., Müller-Neng, J. M. B., Bohus, M., & Weck, F. (2018). Patients' and therapists' experiences of GCMs during bug-in-the-eye and delayed video-based supervised cognitive-behavioral therapy. A randomized controlled trial. *Journal of clinical psychology, 74*(4), 509-522 <https://doi.org/10.1002/jclp.22519>
- Probst, T., Sattel, H., Gündel, H., Henningsen, P., Kruse, J., Schneider, G., & Lahmann, C. (2017). Moderating Effects of Alexithymia on Associations between the Therapeutic Alliance and the Outcome of Brief Psychodynamic-Interpersonal Psychotherapy for Multisomatoform Disorder. *Frontiers in psychiatry, 8*(261). <https://doi.org/10.3389/fpsy.2017.00261>
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>.
- Revelle, W. (2017). Psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, retrieved from <https://CRAN.R-project.org/package=psych> Version = 1.7.5.
- Robinson, C., & Schumacker, R. E. (2009). Interaction Effects: Centering, Variance Inflation Factor, and Interpretation Issues. *Multiple Linear Regression Viewpoints, 35*(1), 6–11.

- Rubel, J., Lutz, W., & Schulte, D. (2015). Patterns of change in different phases of outpatient psychotherapy: A stage-sequential pattern analysis of change in session reports. *Clinical Psychology & Psychotherapy*, 22(1), 1–14. <https://doi.org/10.1002/cpp.1868>
- Rubel, J. A., Rosenbaum, D., & Lutz, W. (2017). Patients' in-session experiences and symptom change: Session-to-session effects on a within- and between-patient level. *Behaviour research and therapy*, 90, 58–66. <https://doi.org/10.1016/j.brat.2016.12.007>
- Schauenburg, H., & Strack, M. (1998). Die Symptom Checklist-90-R (SCL-90-R) zur Darstellung von statistisch und klinischsignifikanten Psychotherapieergebnissen. *PPmP. Psychotherapie, Psychosomatik, medizinische Psychologie*, 48(7), 257–264.
- Schindler, A., Hiller, W., & Witthoft, M. (2013). What predicts outcome, response, and drop-out in CBT of depressive adults? a naturalistic study. *Behavioural and cognitive psychotherapy*, 41(3), 365–370. <https://doi.org/10.1017/S1352465812001063>
- Schlagert, H. S., & Hiller, W. (2017). The predictive value of early response in patients with depressive disorders. *Psychotherapy Research*, 27(4), 488–500. <https://doi.org/10.1080/10503307.2015.1119329>
- 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>
- Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: a meta-analysis. *Journal of Consulting and Clinical Psychology*, 80(4), 547–559. <https://doi.org/10.1037/a0028226>
- Tschacher, W., Junghan, U. M., & Pfammatter, M. (2014). Towards a taxonomy of common factors in psychotherapy-results of an expert survey. *Clinical Psychology & Psychotherapy*, 21(1), 82–96. <https://doi.org/10.1002/cpp.1822>

- Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World psychiatry*, *14*(3), 270–277. <https://doi.org/10.1002/wps.20238>
- World Health Organization (1990). *Composite International Diagnostic Interview (CIDI)*. Geneva: WHO
- World Health Organization (1992). *The ICD-10 Classification of Mental and Behavioral Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: WHO. Retrieved from <http://www.who.int/classifications/icd/en/bluebook.pdf>
- Zeileis, A. & Hothorn, T. (2002). Diagnostic Checking in Regression Relationships. *R News*, *2*(3), 7-10. <https://CRAN.R-project.org/doc/Rnews/>
- Zilcha-Mano, S., Dinger, U., McCarthy, K. S., & Barber, J. P. (2014). Does alliance predict symptoms throughout treatment, or is it the other way around? *Journal of Consulting and Clinical Psychology*, *82*(6), 931–935. <https://doi.org/10.1037/a0035141>.
- Zilcha-Mano, S., & Errázuriz, P. (2017). Early development of mechanisms of change as a predictor of subsequent change and treatment outcome: The case of working alliance. *Journal of Consulting and Clinical Psychology*, *85*(5), 508–520. <https://doi.org/10.1037/ccp0000192>

Table 1

Baseline characteristics of sample. Numbers indicate frequency for sample size, percentages (%), or means with standard deviations in parentheses.

Sample size	911	
Patient age	39.29	(12.91)
Patient sex	59.7% female	
Therapist age	30.53	(6.65)
Therapist sex	85.2% female	
Length of waiting period (telephone contact to first treatment session, days)	220	(90.14)
Length of treatment (days)	525	(261.14)
Number of sessions	45.08	(18.49)
One diagnosis	41.1%	
Two diagnoses	34.9%	
Three or more diagnoses	24%	
Initial symptom severity (Pre GSI)	0.99	(0.62)
Initial life satisfaction (Pre GLS)	17.99	(32.88)

Note. GSI = General Severity Index of the SCL; GLS = General life satisfaction of the QLS^M.

Table 2*Data collected at various points in treatment.*

Session	<i>Pre-treatment</i>	<i>5th session</i>	<i>10th session</i>	<i>Posttreatment</i>
Measures	SCL-90R - Initial GSI QLS ^M - Initial GLS	Bern Post Session Report - Initial interpersonal experiences - Initial intrapersonal experiences - Initial problem actuation	SCL-14 - GSI Bern Post Session Report - Interpersonal experiences - Intrapersonal experiences - problem actuation	SCL-90R - Post GSI QLS ^M - Post GLS

Note. GSI = General Severity Index of the SCL; GLS = General life satisfaction of the QLS^M.

Table 3

Linear regression analysis for continuous post-treatment scores in symptom severity and life satisfaction. Symptom severity is measured by the GSI of the SCL and Life Satisfaction is measured by the GLS of the QLS^M. The most predictive model according to model comparisons is shown.

Variable	Symptom severity		Life Satisfaction	
	β	<i>p</i> -value	β	<i>p</i> -value
Covariates				
Intercept		0.211		< 0.001***
Problematic diagnoses	0.018	0.568	-0.012	0.721
Number of diagnoses	0.049	0.140	-0.135	< 0.001***
Initial symptom severity (pre GSI)	0.612	< 0.001***	-0.036	0.371
Initial life satisfaction (pre GLS)	0.012	0.738	0.401	< 0.001***
Patient age	0.128	< 0.001***	0.012	0.725
Patient sex	0.026	0.410	0.000	0.993
Number of sessions	0.025	0.443	-0.058	0.104
Initial patient interpersonal experiences	-0.076	0.083	0.218	< 0.001***
Initial patient intrapersonal experiences	-0.101	0.013*	0.166	0.008**
Initial patient problem actuation	0.019	0.643	0.002	0.957
Initial therapist interpersonal experiences	-0.040	0.293	-0.078	0.061
Initial therapist intrapersonal experiences	0.023	0.585	-0.010	0.826
Initial therapist problem actuation	0.028	0.505	-0.059	0.193
Early changes in symptom severity (GSI) ^a	-0.220	< 0.001***	0.142	< 0.001***
Interaction early changes in symptom severity (GSI) ^a with number of sessions	-0.001	0.238	-0.001	0.616
Predictors (early changes in ...)				
Patient interpersonal experiences ^b	-0.056	0.125	0.117	0.003**
Patient intrapersonal experiences ^b	-0.120	0.003**	0.110	0.010*
Patient problem actuation ^b	0.065	0.108	-0.016	0.720
Therapist interpersonal experiences ^b	-0.05	0.178	0.005	0.906
Therapist intrapersonal experiences ^b	0.048	0.221	-0.029	0.492
Therapist problem actuation ^b	-0.052	0.194	0.034	0.424
Interaction number of sessions and early changes in ...				
Patient interpersonal experiences ^b	-0.009	0.005**	0.007	0.053
Patient intrapersonal experiences ^b	0.005	< 0.001***	0.001	0.592
Patient problem actuation ^b	-0.042	0.309	-0.036	0.420
Therapist interpersonal experiences ^b	-0.004	0.904	-0.097	0.004**
Therapist intrapersonal experiences ^b	0.000	0.989	0.001	0.649
Therapist problem actuation ^b	0.024	0.506	0.013	0.738

Note. GSI = General Severity Index of the SCL; GLS = General life satisfaction of the QLS^M.^a early changes in GSI are defined as difference between pretreatment and 10th session.

^b early changes in GCMs are defined as difference between ratings at 10th and 5th session

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$

Table 4

Binary logistic regression analysis for reliable improvement in symptom severity. Prediction of reliable improvement versus no reliable improvement. Symptom severity is measured by the GSI of the SCL

Variable	<i>b</i>	<i>OR</i>	<i>95% CI</i>	<i>p-value</i>
Covariates				
Intercept	-1.978	0.138	0.033-0.555	0.006**
Problematic diagnoses	0.043	1.544	0.792-3.006	0.200
Number of diagnoses	-0.224	0.799	0.634-1.001	0.054
Initial symptom severity (pre GSI)	2.453	11.626	7.089-19.793	< 0.001***
Initial life satisfaction (pre GLS)	0.003	1.003	0.995-1.010	0.487
Patient age	-0.022	0.978	0.962-0.995	0.010*
Patient sex	-0.221	0.801	0.513-1.249	0.328
Number of sessions	0.004	1.004		0.500
Initial patient interpersonal experiences	0.295	1.343	0.922-1.971	0.127
Initial patient intrapersonal experiences	0.234	1.263	0.919-1.745	0.152
Initial patient problem actuation	-0.011	0.989	0.806-1.217	0.919
Initial therapist interpersonal experiences	0.239	1.27	0.881-1.840	0.203
Initial therapist intrapersonal experiences	-0.265	0.767	0.497-1.177	0.227
Initial therapist problem actuation	0.033	1.034	0.727-1.472	0.852
Early changes in symptom severity (GSI) ^a	1.093	2.938	1.735-5.293	<0.001***
Interaction early changes in symptom severity (GSI) ^a with number of sessions	0.007	1.007	0.979-1.037	0.625
Predictors (early changes in ...)				
Patient interpersonal experiences ^b	0.353	1.424	0.914-2.240	0.121
Patient intrapersonal experiences ^b	0.197	1.218	0.872-1.706	0.247
Patient problem actuation ^b	-0.197	0.821	0.669-1.005	0.058
Therapist interpersonal experiences ^b	0.525	1.691	1.139-2.549	0.011*
Therapist intrapersonal experiences ^b	-0.517	0.596	0.377-0.934	0.025*
Therapist problem actuation ^b	0.369	1.447	1.068-1.972	0.018
Interaction number of sessions and early changes in ...				
Patient interpersonal experiences ^b	0.016	1.016	0.995-1.380	0.133
Patient intrapersonal experiences ^b	-0.020	0.980	0.966-0.994	0.005**
Patient problem actuation ^b	0.006	1.006	0.997-1.015	0.200
Therapist interpersonal experiences ^b	-0.007	0.993	0.975-1.011	0.477
Therapist intrapersonal experiences ^b	0.003	1.003	0.984-1.022	0.768
Therapist problem actuation ^b	-0.005	0.995	0.982-1.008	0.420

Note. GSI = General Severity Index of the SCL; GLS = General life satisfaction of the QLS^M.^a early changes in GSI are defined as difference between pretreatment and 10th session.

^b early changes in GCMs are defined as difference between ratings at 10th and 5th session

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$

Table 5

Binary logistic regression analysis for reliable improvement in life satisfaction. Prediction of reliable improvement versus no reliable improvement. Life Satisfaction is measured by the GLS of the QLS^M

Variable	<i>b</i>	<i>OR</i>	<i>95% CI</i>	<i>p-value</i>
Covariates				
Intercept	0.098	1.103	0.332-3.693	0.873
Problematic diagnoses	-0.248	0.781	0.411-1.460	0.442
Number of diagnoses	-0.251	0.778	0.641-0.940	0.010*
Initial symptom severity (pre GSI)	-0.233	0.792	0.554-1.124	0.195
Initial life satisfaction (pre GLS)	-0.024	0.976	0.969-0.982	< 0.001***
Patient age	0.009	1.009	0.995-1.024	0.194
Patient sex	-0.199	0.819	0.558-1.201	0.307
Number of sessions	0.001	1.001	0.991-1.011	0.845
Initial patient interpersonal experiences	0.588	1.801	1.310-2.495	< 0.001***
Initial patient intrapersonal experiences	0.452	1.571	1.195-2.081	0.001**
Initial patient problem actuation	-0.110	0.896	0.751-1.068	0.223
Initial therapist interpersonal experiences	-0.109	0.897	0.660-1.219	0.486
Initial therapist intrapersonal experiences	-0.132	0.876	0.611-1.252	0.469
Initial therapist problem actuation	-0.107	0.899	0.666-1.211	0.483
Early changes in symptom severity (GSI) ^a	0.411	1.508	0.996-2.309	0.055
Interaction early changes in symptom severity (GSI) ^a with number of sessions	-0.010	0.990	0.967-1.012	0.358
Predictors (early changes in ...)				
Patient interpersonal experiences ^b	0.484	1.622	1.107-2.390	0.014*
Patient intrapersonal experiences ^b	0.317	1.374	1.017-1.864	0.040*
Patient problem actuation ^b	-0.055	0.946	0.797-1.123	0.528
Therapist interpersonal experiences ^b	0.055	1.057	0.766-1.463	0.736
Therapist intrapersonal experiences ^b	-0.203	0.816	0.560-1.185	0.287
Therapist problem actuation ^b	0.071	1.074	0.831-1.391	0.586
Interaction number of sessions and early changes in ...				
Patient interpersonal experiences ^b	0.019	1.019	1.000-1.039	0.049*
Patient intrapersonal experiences ^b	-0.004	0.996	0.983-1.009	0.531
Patient problem actuation ^b	-0.002	0.998	0.991-1.005	0.585
Therapist interpersonal experiences ^b	-0.011	0.989	0.975-1.003	0.145
Therapist intrapersonal experiences ^b	-0.003	0.997	0.981-1.013	0.720
Therapist problem actuation ^b	0.001	1.001	0.990-1.012	0.876

Note. GSI = General Severity Index of the SCL; GLS = General life satisfaction of the QLS^M.^a early changes in GSI are defined as difference between pretreatment and 10th session.

^b early changes in GCMs are defined as difference between ratings at 10th and 5th session

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$

Supplement

Supplement Table 1

Correlation matrix of covariates and predictors used in the analyses.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Covariates																				
(1) Problematic diagnoses	1.00	0.20	0.08	0.12	0.00	0.05	0.03	0.15	0.07	0.02	0.12	0.08	0.02	0.01	0.04	0.06	0.01	0.04	0.10	0.03
(2) Number of diagnoses	0.20	1.00	0.27	0.19	0.08	0.08	0.19	0.03	0.10	0.00	0.03	0.04	0.04	0.03	0.00	0.00	0.04	0.01	0.06	0.04
(3) Initial symptom severity (pre GSI)	0.08	0.27	1.00	0.36	0.07	0.02	0.16	0.10	0.10	0.07	0.02	0.09	0.06	0.32	0.04	0.06	0.03	0.01	0.03	0.01
(4) Initial life satisfaction (pre GLS)	0.12	0.19	0.36	1.00	0.03	0.08	0.17	0.16	0.14	0.03	0.13	0.10	0.08	0.02	0.02	0.07	0.01	0.05	0.05	0.00
(5) Patient age	0.00	0.08	0.07	0.03	1.00	0.01	0.07	0.08	0.04	0.19	0.02	0.04	0.03	0.05	0.03	0.12	0.02	0.03	0.01	0.02
(6) Patient sex	0.05	0.08	0.02	0.08	0.01	1.00	0.05	0.11	0.06	0.06	0.16	0.09	0.07	0.02	0.09	0.03	0.05	0.08	0.01	0.02
(7) Number of sessions	0.03	0.19	0.16	0.17	0.07	0.05	1.00	0.06	0.06	0.03	0.00	0.09	0.09	0.01	0.06	0.05	0.04	0.02	0.05	0.05
(8) Initial patient interpersonal experiences	0.15	0.03	0.10	0.16	0.08	0.11	0.06	1.00	0.29	0.09	0.30	0.11	0.11	0.03	0.39	0.02	0.02	0.03	0.03	0.00
(9) Initial patient intrapersonal experiences	0.07	0.10	0.10	0.14	0.04	0.06	0.06	0.29	1.00	0.21	0.17	0.17	0.06	0.12	0.11	0.46	0.07	0.01	0.12	0.02
(10) Initial patient problem actuation	0.02	0.00	0.07	0.03	0.19	0.06	0.03	0.09	0.21	1.00	0.06	0.03	0.18	0.02	0.01	0.10	0.57	0.01	0.05	0.06
(11) Initial therapist interpersonal experiences	0.12	0.03	0.02	0.13	0.02	0.16	0.00	0.30	0.17	0.06	1.00	0.28	0.16	0.05	0.02	0.00	0.03	0.39	0.14	0.06
(12) Initial therapist intrapersonal experiences	0.08	0.04	0.09	0.10	0.04	0.09	0.09	0.11	0.17	0.03	0.28	1.00	0.35	0.01	0.01	0.01	0.05	0.02	0.56	0.05
(13) Initial therapist problem actuation	0.02	0.04	0.06	0.08	0.03	0.07	0.09	0.11	0.06	0.18	0.16	0.35	1.00	0.01	0.00	0.00	0.12	0.03	0.12	0.52
(14) Early changes in Symptom severity	0.01	0.03	0.32	0.02	0.05	0.02	0.01	0.03	0.12	0.02	0.05	0.01	0.01	1.00	0.02	0.07	0.13	0.01	0.01	0.00
Predictors (early changes in ...)																				
(15) Patient interpersonal experiences	0.04	0.00	0.04	0.02	0.03	0.09	0.06	0.39	0.11	0.01	0.02	0.01	0.00	0.02	1.00	0.25	0.08	0.19	0.07	0.01
(16) Patient intrapersonal experiences	0.06	0.00	0.06	0.07	0.12	0.03	0.05	0.02	0.46	0.10	0.00	0.01	0.00	0.07	0.25	1.00	0.12	0.16	0.14	0.05
(17) Patient problem actuation	0.01	0.04	0.03	0.01	0.02	0.05	0.04	0.02	0.07	0.57	0.03	0.05	0.12	0.13	0.08	0.12	1.00	0.04	0.00	0.21
(18) Therapist interpersonal experiences	0.04	0.01	0.01	0.05	0.03	0.08	0.02	0.03	0.01	0.01	0.39	0.02	0.03	0.01	0.19	0.16	0.04	1.00	0.20	0.07
(19) Therapist intrapersonal experiences	0.10	0.06	0.03	0.05	0.01	0.01	0.05	0.03	0.12	0.05	0.14	0.56	0.12	0.01	0.07	0.14	0.00	0.20	1.00	0.18
(20) Therapist problem actuation	0.03	0.04	0.01	0.00	0.02	0.02	0.05	0.00	0.02	0.06	0.06	0.05	0.52	0.00	0.01	0.05	0.21	0.07	0.18	1.00

Note. GSI = General severity index of SCL-90R; GLS = General life satisfaction of the QLS^M.

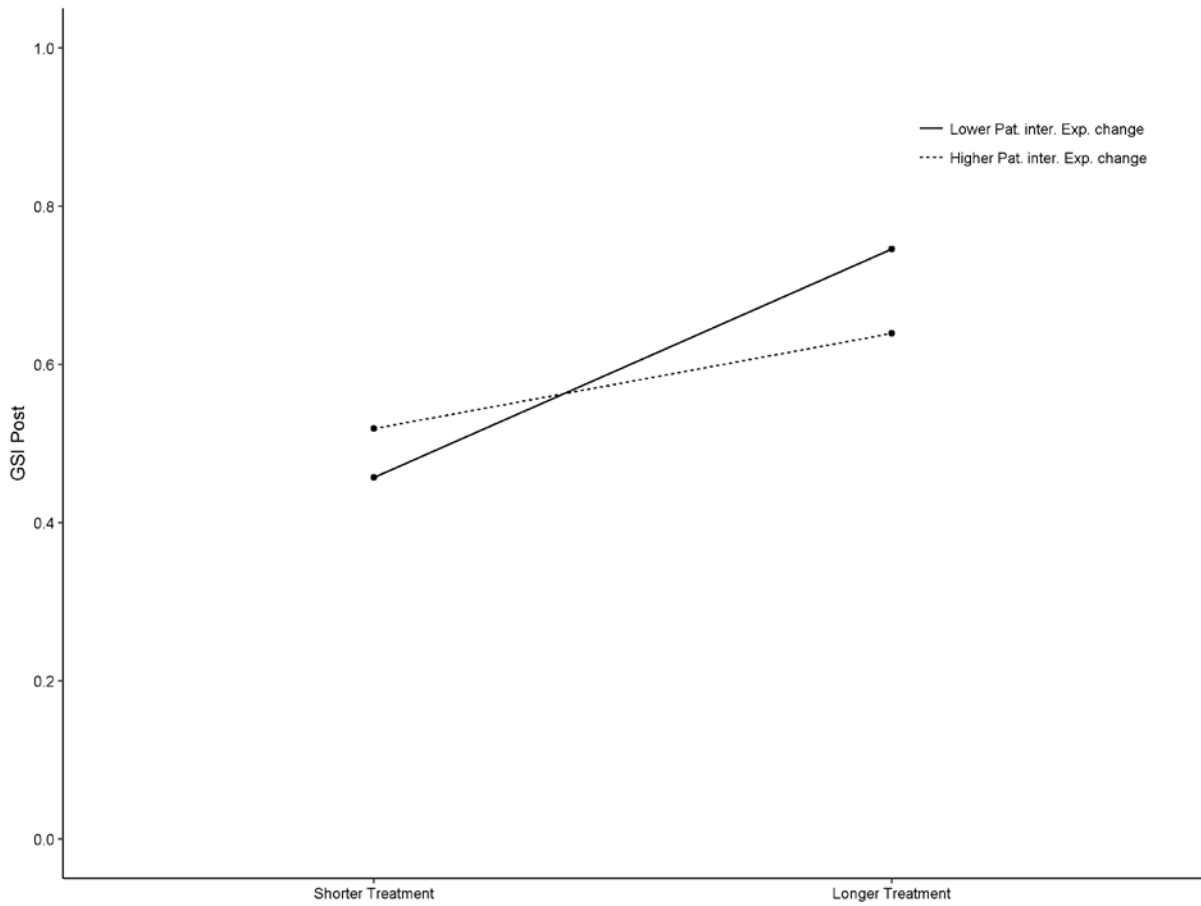
Supplement Table 2

Linear regression analysis for continuous post-treatment scores in symptom severity and life satisfaction. Symptom severity is measured by the GSI of the SCL-90R and Life Satisfaction is measured by the GLS of the QLS^M. The most predictive model according to model comparisons is shown.

Variable	GSI		GLS	
	β	<i>p</i> -value	β	<i>p</i> -value
Covariates				
Intercept		0.435		< 0.001***
Problematic diagnoses	0.019	0.542	-0.005	0.872
Number of diagnoses	0.041	0.217	-0.125	< 0.001***
Initial symptom severity (pre GSI)	0.033	0.672	0.308	< 0.001***
Initial life satisfaction (pre GLS)	0.003	0.935	0.413	< 0.001***
Patient age	0.131	< 0.001***	0.002	0.945
Patient sex	0.022	0.497	-0.007	0.84
Number of sessions	0.023	0.478	-0.063	0.078
Means in symptom severity (GSI) ^a	0.55	< 0.001***	-0.324	< 0.001***
Predictors (means in)				
Patient interpersonal experiences ^b	-0.076	0.034*	0.213	< 0.001***
Patient intrapersonal experiences ^b	-0.085	0.018*	0.101	0.009**
Patient problem actuation ^b	0.012	0.731	0.007	0.853
Therapist interpersonal experiences ^b	-0.041	0.254	-0.072	0.066
Therapist intrapersonal experiences ^b	0.018	0.622	0.011	0.773
Therapist problem actuation ^b	0.02	0.558	-0.054	0.148

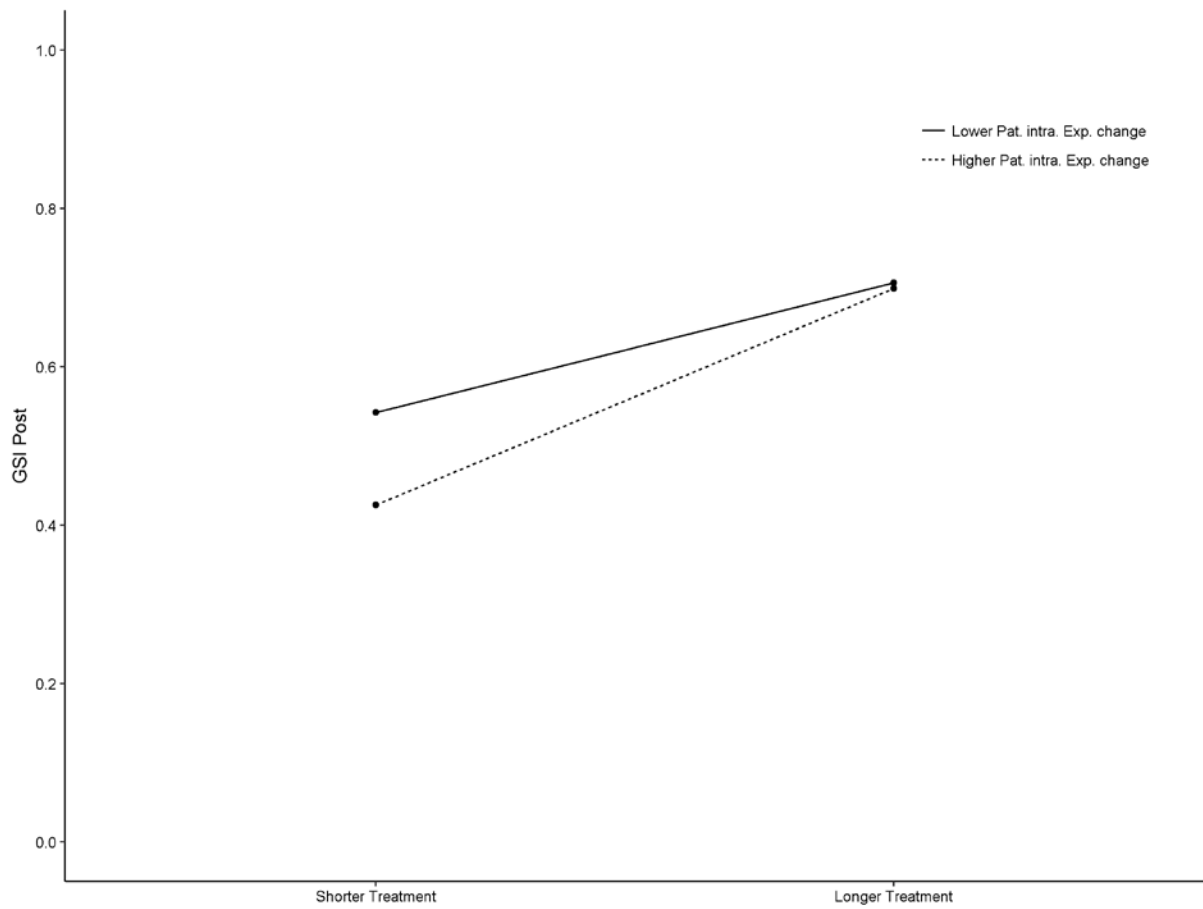
Note. GSI = General Severity Index of the SCL-90R; GLS = General life satisfaction of the QLS^M. ^a mean GSI is defined as the mean between pretreatment and 10th session. ^b mean GCMs are defined as the mean between ratings at 5th and 10th session. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. R^2 GSI models $R^2 = 0.38$ (Nagelkere's pseudo- R^2 , $R^2 = 0.46$) model containing only covariates, $R^2 = 0.45$ (Nagelkere's pseudo- R^2 , $R^2 = 0.56$) model containing covariates and predictors, $R^2 = 0.46$ (Nagelkere's pseudo- R^2 , $R^2 = 0.56$) model containing covariates, predictors and interactions with number of sessions, AICs 742, 679, 685 respectively; R^2 GLS models $R^2 = 0.29$ (Nagelkere's pseudo- R^2 , $R^2 = 0.29$) model containing only covariates, $R^2 = 0.37$ (Nagelkere's pseudo- R^2 , $R^2 = 0.37$) model containing covariates and predictors, $R^2 = 0.38$ (Nagelkere's pseudo- R^2 , $R^2 = 0.38$) model containing covariates, predictors and interactions with number of sessions AICs 6052, 5994 respectively.

Supplement Figure 1



Supplement Figure 1. Interaction of treatment length and early changes in patient-rated interpersonal experiences with post symptom severity as outcome. GSI = Global Severity Index, Pat. inter. exp. change = early changes in patient-rated interpersonal experiences.

Supplement Figure 2



Supplement Figure 2. Interaction of treatment length and early changes in patient-rated intrapersonal experiences with post symptom severity as outcome. GSI = Global Severity Index, Pat. intra. exp. change = early changes in patient-rated intrapersonal experiences.

Gmeinwieser, Hagmayer, Pieh und Probst (2019b)

Gmeinwieser, S., Hagmayer, Y., Pieh, C., & Probst, T. (2019). General change mechanisms in the early treatment phase and their associations with the outcome of cognitive behavioural therapy in patients with different levels of motivational incongruence. *Clinical Psychology & Psychotherapy*, 26(5), 550–561. <https://doi.org/10.1002/cpp.2381>

Title:

General change mechanisms in the early treatment phase and their associations with the outcome of Cognitive-Behavioral Therapy in patients with different levels of motivational incongruence

Running Title:

Incongruence, GCMs, and outcome

Authors

Sebastian Gmeinwieser¹, York Hagmayer², Christoph Pieh³, Thomas Probst³

Affiliations:

¹ Department for Clinical Psychology and Psychotherapy, Therapy and Counseling Center, University of Göttingen, Germany

² Department for Cognitive and Decision Sciences, University of Göttingen, Germany

³ Department for Psychotherapy and Biopsychosocial Health, Danube University, Krems, Austria

Corresponding author:

Sebastian Gmeinwieser, M.Sc.

Department for Clinical Psychology and Psychotherapy, Therapy and Counseling Center, University of Göttingen, Gosslerstrasse 14, D-37073, Germany

E-Mail: Sebastian.Gmeinwieser@psych.uni-goettingen.de

Abstract

Early general change mechanisms (GCMs) have been shown to be associated with psychotherapy outcome but it remains unclear which specific patients benefit from which GCM. This study explored whether the patients' characteristic motivational incongruence moderates the effect of GCMs in the early treatment phase on the outcome. Three early GCMs (interpersonal experiences, intrapersonal experiences, problem actuation) were evaluated after five therapy-preparing sessions by patients as well as therapists. Based on previous work we assumed that the association between intrapersonal experiences and outcome is moderated by incongruence. A total of 524 patients completing outpatient cognitive behavioral therapy (CBT) were investigated. The patient-reported outcome measure was psychological symptom severity. The patients' motivational incongruence was assessed with the incongruence questionnaire short form at baseline. Results showed that therapists' ratings of all three early GCMs and patients' ratings of early problem actuation were not associated with CBT outcome. By contrast, positive patients' ratings of early interpersonal as well as early intrapersonal experiences were beneficial for CBT outcome (both $p < .05$). Only the association between patients' ratings of early intrapersonal experiences and CBT outcome was moderated by incongruence ($p < .05$). The higher the patients' baseline incongruence was, the more beneficial early intrapersonal experiences in the patients' perspective were for a good outcome. These findings entail that increasing early intrapersonal experiences from the patients' perspective is particularly important in patients with motivational incongruence but working on early interpersonal experiences in the patients' perspective is important in patients with all levels of incongruence to reach a good CBT outcome.

Key Practitioner Message

- Therapists should work on the patients' early interpersonal experiences (including the therapeutic alliance) in patients with all levels of motivational incongruence to reach a good CBT outcome
- Strengthening the patients' early intrapersonal experiences (i.e., mastery and clarification of meaning) is particularly important in patients with medium to high levels of incongruence to improve CBT outcome.

Keywords

General change mechanisms, motivational incongruence, moderator analysis, cognitive behavioral therapy, treatment outcome

Conflict of interest statement

The authors report no conflict of interest.

Introduction

To personalize psychotherapy, it is important to know which treatment options and which psychotherapeutic process variables (e. g. common factors like the therapeutic alliance, see Cuijpers, Reijnders, & Huibers, 2018) work best for which patient. In this context, it is important to go beyond diagnostic categories and take further patient characteristics into account. In the last years, several studies were conducted to identify the treatment options patients with specific characteristics benefit from the most (see the reviews by Cohen, & DeRubeis, 2018, and Schneider, Arch, & Wolitzky-Taylor, 2015). Less research investigated the question how specific psychotherapeutic process variables are correlated with the outcome for patients with specific characteristics.

The most researched psychotherapeutic process variable is the alliance. A recent meta-analysis on the association between alliance and psychotherapy outcome included 295 studies with more than 30,000 patients and found a correlation of $r = .278$ corresponding to an effect size of $d = .579$ (Flückiger, Del Re, Wampold, & Horvath, 2018). Although the meta-analysis by Flückiger et al. (2018) reported that the alliance-outcome association was not influenced by patient variables, some recent research showed that specific patient variables moderate the relationship between alliance and outcome (e. g., Lorenzo-Luaces, DeRubeis & Webb, 2014; Probst et al., 2017; Zilcha-Mano & Errázuriz, 2015; Zack et al., 2015; Lorenzo-Luaces et al., 2017; Zilcha-Mano, 2018). Lorenzo-Luaces et al. (2014), for example, found that the number of previous depressive episodes moderates the association between alliance and outcome in cognitive behavioral therapy (CBT). A stronger alliance was beneficial for a good outcome in patients with 0-2 previous episodes, but not in patients with more previous depressive episodes. In psychodynamic interpersonal psychotherapy, a stronger alliance in the therapists' perspective was more beneficial for a good outcome for those patients with multisomatoform disorder being more alexithymic at baseline than for those being less alexithymic (Probst et al., 2017). One explanation for these inconsistent results regarding patients' characteristics as

moderators of the alliance-outcome association might be that some studies focused on between-patient comparisons (i. e., a comparison of outcomes between patients with different alliance levels) whereas other studies focused on within-patient comparisons (i. e., a comparison of outcomes after sessions with different levels of alliance).

The alliance is a process variable and a general change mechanism (GCM) according to the theoretical framework of Grawe (1997). Besides alliance, Grawe identified the following GCMs based upon an extensive review of the psychotherapy research literature: mastery/coping, clarification of meaning, problem actuation, and resource activation. Mastery/coping refers to the patient's mastery and coping abilities in everyday tasks and difficult situations. In treatment, therapists achieve better mastery/coping by working on the self-efficacy of the patient and strengthening functional strategies to deal with situations related to the problem (e. g., training of problem-solving skills). Clarification of meaning stands for the patient acquiring insights into the meaning behind pathologic symptoms and dysfunctional behavior. In treatment, therapists can improve clarification by linking symptoms and behaviors to the patient's underlying motivational goals. It is analyzed whether behaviors are compatible with personal goals and other behaviors. As soon as the link becomes explicit for the patient and the underlying conflicts become apparent, an important step for a healthy change can be taken. Problem actuation means that the patient experiences disorder related problems within the treatment session. There are various strategies for therapists to help patients re-experiencing their problems such as exposing patients to fear inducing stimuli. Resource activation is defined as the activation of change-supporting abilities and strengths of the patients. These abilities are then used in treatment to further strengthen mastery/coping strategies.

According to Flückiger, Grosse Holtforth, Znoj, Caspar and Wampold (2013), the GCMs can be subsumed under three secondary factors. Alliance belongs to interpersonal experiences whereas mastery as well as clarification of meaning are summarized as

intrapersonal experiences. Inter- as well as intrapersonal experiences are experiences within the treatment session, e.g. interpersonal experiences refer only to the patient-therapist dyad and not to interpersonal experiences in general. The third secondary factor is problem actuation. The secondary GCMs described by Flückiger et al. (2013) do not include resource activation. As the present study uses the secondary factors reported by Flückiger et al. (2013), resource activation was not analyzed in this study.

The GCMs proposed by Grawe (1997) were associated with psychotherapy outcome in several studies (e.g., Call, D’Errico, Bents, & Mander, 2018; Mander et al., 2013; Mander et al., 2015b; Flückiger et al., 2013; Gassmann & Grawe, 2006; Rubel, Rosenbaum, & Lutz., 2017; Stangier, von Consbruch, Schramm & Heidenreich, 2010). Flückiger et al. (2013), for example, highlighted the importance of GCMs in the early phase of psychotherapy and found that early interpersonal as well as early intrapersonal experiences predict the outcome of psychotherapy. Mander et al. (2015b) conducted analyses of the association between several GCMs and several outcome measures and found that results vary dependent on the GCM and the selected outcome. While problem actuation was not associated with any outcome, therapists’ ratings of clarification were associated with changes in perceived stress but neither with changes in psychological symptom severity nor with changes in interpersonal problems. Therapists’ ratings of mastery, on the other hand, were positively associated with pre-post changes in perceived stress and symptom severity differences. Thus, the relationship between GCMs and therapy outcome turned out to be complex and inconsistent across studies. This indicates that there are probably moderators of the relationship. Therefore it is important to address the question “which specific patients benefit from which GCM”?

Although a few studies have investigated patient characteristics as moderators between alliance and outcome (see above), there appears to be a research gap for other GCMs. Such research could be useful for clinical decision-making. Some GCMs might be beneficial only for patients with specific characteristics whereas patients without these characteristics might

not benefit from these GCMs but from other GCMs. Knowing which GCMs work best for whom might guide therapists to focus on those GCMs that have been shown to predict a favorable outcome in patients with characteristics similar to their patient.

Patient characteristics that interact with GCMs are motivational schemas (Mander et al., 2015a) and motivational goals (e. g., Grosse Holtforth, Grawe, & Castonguay, 2006). Moreover, interpersonal problems, which have been shown to influence the GCM alliance (e.g., Muran, Segal, Samstag, & Crawford, 1994), were associated with motivational incongruence (Grosse Holtforth, Pincus, Grawe, Mauler, & Castonguay, 2007).

Motivational schemas, motivational goals, motivational incongruence, and their relation to mental health are described by Grawe's consistency theory (2004). According to Grawe, humans strive to satisfy basic psychological needs (e.g. bonding) in order to stay healthy. Motivational schemas develop from the basic needs to satisfy or protect these needs. Motivational goals, in turn, emerge in order to translate the schemas into everyday life. There are two types of motivational goals, approach goals and avoidance goals. If the environment is supportive of the basic needs and motivational schemas, approach goals are developed and the basic needs are satisfied. But if the environment is non-supportive, avoidance goals are developed and the focus becomes protecting the motivational schemas from further harm. Although avoidance goals can be adaptive and healthy to a certain degree, they become harmful in the long term and prevent basic needs from being satisfied. For example, staying indoors to avoid panic attacks reduces anxiety and panic attacks in the short term, but prevents the person from bonding with other people and forming interpersonal relationships. If an approach or avoidance goal is not being met in everyday life, incongruence results. Incongruence is the discrepancy of the goals being set and the actions taken to meet these goals. Humans pursue consistency in their everyday actions by comparing them – knowing- or unknowingly – to their basic needs. This pursuit is regulated by consistency mechanisms (e.g. coping mechanisms, defense mechanisms). If these consistency regulating mechanisms

are maladaptive, emotional and psychological distress can result. Hence, incongruence may be important to address in psychotherapeutic treatment.

Research showed that motivational incongruence is associated with more psychological distress and less well-being (Grosse Holtfort & Grawe, 2003). Moreover, changes in incongruence during inpatient CBT were correlated with treatment outcomes (Berking, Grosse Holtfort, & Jacobi, 2003). Grosse Holtforth, Grawe, Fries and Znoj (2008) investigated incongruence and its two aspects (discontentment with the way approach goals are met vs. occurrence of avoidance goals) as moderators of treatment outcomes in a randomized-controlled trial (RCT) comparing a combined experiential and CBT psychotherapy with CBT alone. While CBT focused mainly on mastery, the combined experiential and CBT approach had a stronger focus on integrating clarification and mastery (summarized as the secondary factor intrapersonal experiences). Patients experiencing more avoidance goals benefitted more from the combined psychotherapy than from CBT alone, whereas patients experiencing less avoidance goals benefitted more from CBT alone than from the combined psychotherapy. However, this was the case only for the outcomes interpersonal problems and retrospectively assessed improvements but not for psychological symptoms or well-being.

To advance the findings of Grosse Holtforth et al. (2008), we examined whether the effect of early intrapersonal experiences (mastery + clarification of meaning) on CBT outcome is moderated by motivational incongruence. Based on the theoretical model of Grawe (2004) and the findings of Grosse Holtforth et al. (2008), we hypothesized for our theory-guided moderation analysis that patients with high motivational incongruence would benefit more from early intrapersonal experiences than less incongruent patients.

We found no studies on interactions between incongruence and the GCMs interpersonal experiences and problem actuation. However, incongruence was shown to be related to interpersonal problems (Grosse Holtforth et al., 2007). Interpersonal problems in

turn are related to in-session interpersonal experiences such as the therapeutic alliance (e. g., Muran, et al., 1994). Furthermore, incongruence is correlated with psychological problems (Grosse Holtfort, & Grawe, 2003) and the actuation of problems might be differently important for a good outcome in patients with more or less incongruence related problems. Therefore, we considered it as interesting to explore potential interactions of incongruence and interpersonal experiences as well as problem actuation with respect to CBT outcome. In two explorative analyses, we therefore analyzed incongruence as a moderator of the effect of early interpersonal experiences and early problem actuation on CBT outcome.

Method

Participants

Patients were psychotherapy completers with mental health problems treated between 2007 and 2017 in an outpatient psychotherapy center at Göttingen, Germany. The outpatient clinic is a part of the Department for Clinical Psychology and Psychotherapy of the University of Göttingen.

Our sample consisted of 1772 patients that ended treatment. However, 629 patients were excluded from the analysis because they ended treatment unilaterally (35% drop-out). Another 616 (35%) patients were excluded because they did not complete the incongruence measure at baseline, the pre- and post-treatment outcome measure, or the patients as well as their therapists did not provide ratings of early GCMs after five therapy-preparing sessions. Another three patients (< 0.1%) were excluded because they had an unknown treatment termination status.

Five-hundred-twenty-four patients (30% of the original sample) fulfilled the inclusion criteria, 308 (58.4%) of which were female. Patients' age was $M = 39.97$ ($SD = 13.12$). The patients received $M = 49.29$ ($SD = 17.22$) sessions (including the 5 therapy-preparing sessions). A more detailed description of the patients is displayed in Table 1.

The patients were treated by 111 therapists (86.2% female) with $M = 4.75$ ($SD = 3.33$) patients per therapist. Most therapists were in post-graduate training in CBT. In Germany, every fourth therapy session has to be supervised during post-graduate training. Therapists' age was $M = 30.94$ ($SD = 7.39$) years. Therapists assigned $M = 1.99$ ($SD = 1.03$) diagnoses per patient. Depressive disorders and anxiety disorders were the most common diagnostic categories with 68.4% and 46.9% respectively. Diagnoses were given based on therapists' clinical judgement following the criteria of the ICD-10 (WHO, 1992). All patients gave written informed consent to their data being used anonymously for scientific research purposes.

Measures

We used standardized measures for symptom severity, GCMs, and incongruence. See supplement Table 1 for the complete list of the variables that were routinely measured at the outpatient clinic.

Symptom severity: The Symptom Checklist (SCL-90R; Derogatis, 1977; German version by Franke, 2002) measures psychological symptom severity with 90 items. Symptoms are rated on a 5-point-likert scale ranging from 0 (*not at all*) to 4 (*extremely*). We used the Global Severity Index (GSI) as the patient-reported outcome measure which is the mean of all 90 items. Higher GSI values indicate more severe psychological symptoms. In a representative German sample, the GSI showed good reliability (Cronbach's Alpha of $\alpha = .97$ [Franke, 2002], $\alpha = .98$ in this study). The SCL-90R was administered to the patients at pre- and post-treatment. The post-treatment assessment took place 5-1 sessions before the actual end so that therapists could discuss the results in the last session(s).

GCMs: We used the Bern Post Session Report Patient- and Therapist-version (BPSR; Flückiger et al., 2010). The patient-version consists of 22 items and the therapist version of 27 items. Items in both versions are rated on a 7-point-likert-scale (-3: *not at all* to 3: *very much*) except for the items 13-27 in the therapist version. These are rated on a 5-point-likert scale (0:

strongly disagree to 4: *strongly agree*). For both versions, we calculated the secondary factors (see Flückiger et al., 2013) “interpersonal experiences” (consisting of the scales alliance and contentment in the patient version and of the scales alliance and openness in the therapist version), “intrapersonal experiences” (comprising the mastery and clarification scales in both versions), and problem actuation. Higher values on the secondary scales indicate a more positive realization of the GCMs. Reliability analyses by Flückiger et al. (2013) and Probst et al. (2017) showed Cronbach’s alpha values ranging between $\alpha = .74$ and $\alpha = .89$ for the secondary factors. Cronbach’s alpha values in the current sample ranged from $\alpha = .73$ (problem actuation) to $\alpha = .85$ (interpersonal experiences). The GCMs were assessed after five therapy-preparing sessions. Patients and therapists were asked to rate the GCMs for the last session. The first five sessions have been used to define the early phase of psychotherapy in previous studies as well (e. g., Flückiger et al., 2013; Hayes, et al., 2007; Huber et al., 2018; Kleinstäuber, Lambert, & Hiller, 2017) although it should be kept in mind that there is a considerable heterogeneity in definitions. For example, Bradford et al. (2011) defined the early treatment phase as the first month in treatment, Van et al. (2008) as the first two months in treatment, or Flood, Page, and Hooke (2018) as the first five days of treatment.

Incongruence: Motivational incongruence was operationalized with the incongruence questionnaire short form (K-INK, Grosse Holtfort, & Grawe, 2003; Grosse Holtfort, Grawe & Tamcan, 2004). The K-INK is a self-report that consists of 23 statements. The item content and the scales of the K-INK were derived from the “Inventory of Approach and Avoidance Motivation” (IAAM; Grosse Holtfort & Grawe, 2000; 2002). The IAAM measures the intensity of motivational goals and the K-INK measures the degree to which these motivational goals are insufficiently met (discontentment with the way approach goals are met vs. occurrence of avoidance goals). The statements are divided into approach goals (items 1-14) and avoidance goals (items 15-23). Statements in the approach goal-part are rated on a five point-likert scale ranging from 1 (*barely*) to 5 (*sufficiently*) and in the avoidance goals-

part from 1 (*not at all*) to 5 (*very much*). The total score of the K-INK is the average of the ratings with respect to avoidance goals and the reversed ratings regarding approach goals. Thus, higher K-INK total score values indicate stronger incongruence. The total score showed a high reliability with a Cronbach's alpha of $\alpha = .91$ in the current sample and in the sample of Grosse Holtforth et al. (2004). In the current study, the patients' K-INK total score at pre-treatment was used.

Procedure

To apply for treatment, patients first had to contact the center via telephone. Then patients were invited to a 30-minute session with a licensed therapist for problem exploration and first potential diagnoses. In a separate session, a trained research assistant conducted the Composite International Diagnostic Interview (CIDI, WHO, 1990) for a standardized assessment. After this session, patients were assigned to a therapist. Most of the therapists were in post-graduate training. All had at least 1.5 years of clinical experience. The therapists then conducted five therapy-preparing sessions and started the psychotherapeutic treatment afterwards.

In Germany, patients get five therapy-preparing sessions during which an indication for treatment is determined and a comprehensive diagnostic assessment is made. First, a thorough analysis of the patient's problems is conducted. In the analysis, the therapist seeks to get a clear picture of the patient's problems. The therapist asks for the starting point of the problems, their severity, their course, and the resulting impairments in everyday life. This is followed by a behavioral analysis and an analysis of the resources the patient already utilizes to cope with his/her problems. At the end of the preparing sessions patient and therapist define realistic and achievable goals for treatment. The diagnoses are assigned after these five therapy-preparing sessions. They are based on the clinical judgement of the therapist and classified according to the ICD-10.

Therapists have to send a written assessment and treatment plan derived from the preparing sessions to the patient's health insurance company to secure funding of the subsequent psychotherapeutic treatment. When the health insurance approves the planned treatment, the actual psychotherapy starts. The number of sessions depends on the patients' mental health problems and their severity.

Statistics

SPSS (version 25, IBM Corp., 2017) was used for all statistical analyses. Means (M), standard deviations (SD), frequencies (n), percentages (%), and partial correlations (r) were calculated. Moderation analyses were performed with the SPSS macro PROCESS v3.0 (Hayes, 2017). Early GCMs (secondary scales of the BPSR after five therapy-preparing sessions) were entered as the independent variable (X-variable), incongruence (K-INK total score at pre-treatment) as the moderator variable (W-variable), and psychological symptom severity at treatment termination (post-treatment GSI of the SCL-90R) as the outcome variable (Y-variable). Covariates were psychological symptom severity at baseline (pre-treatment GSI of the SCL-90R) as well as the number of psychotherapy sessions. A statistically significant interaction between an independent variable (here: early GCMs) and a moderator variable (here: incongruence) indicates a moderation effect, i.e. that the strength of the effect of an early GCM on the outcome depends on incongruence. For each statistically significant interaction effect between early GCMs and incongruence, the Johnson–Neyman Technique was applied to identify the threshold(s) of the moderator (incongruence) where the association between the independent variable (early GCMs) and the outcome (psychological symptom severity at post-treatment) transition(s) between statistical significance and non-significance. The significance level in these moderation analyses was set to $p < .05$ (two-tailed). No correction for multiple testing was applied in the moderation analyses, because we only had one theory-guided hypothesis which concerned early intrapersonal experiences. The

moderation analyses with respect to early interpersonal experiences and early problem actuation were exploratory.

Furthermore, we performed partial correlation analyses between the six GCMs (patients' and therapists' ratings of early interpersonal and intrapersonal experiences, and early problem actuation) and CBT outcome as well as between patients' baseline incongruence and CBT outcome controlling for the number of sessions and baseline symptom severity. We used a Bonferroni correction for multiple testing for these seven partial correlation analyses and set the significance level to $p < .007$ for the two-tailed tests.

The number of sessions was added as a covariate in the moderation analyses as well as the partial correlation analyses, since we wanted to control for the high variance of the number of sessions attended by our sample (see Table 1).

As patients were nested within therapists, we tested for therapist-level effects by performing a likelihood-ratio-test comparing a single level nullmodel and a multilevel nullmodel containing also the therapist level (i.e., random intercept conditional on therapist). The likelihood-ratio-test was significant in favor of the single level ($p < 0.001$) and the single level nullmodel had the lower AIC (921 single level nullmodel vs. 928 multilevel nullmodel). The therapist-level variance in the multilevel model was 0. Therefore we refrained from performing multilevel analyses on our data.

Results

Correlations between incongruence, GCMs, and CBT outcome

To explore how incongruence (K-INK total score at pre-treatment) and early GCMs (secondary factors of the BPSR after five therapy-preparing sessions) are correlated with CBT outcome (GSI of the SCL-90R at post-treatment), partial correlations (with GSI of the SCL-90R at pre-treatment and the number of sessions as control variables) were calculated. The results are presented in Table 2. It turned out that higher patients' ratings of interpersonal experiences and intrapersonal experiences were significantly associated with less severe

psychological symptoms at post-treatment across all patients after correcting for multiple testing (both $p < .001$).

Results for the theory-guided moderation analysis

Incongruence as a moderator of the relation between early intrapersonal experiences and CBT outcome

The results of the moderation analyses for the early GCM intrapersonal experiences are given in Table 3. Incongruence significantly moderated the association between patients' ratings of early intrapersonal experiences and CBT outcome ($p = .039$). Within our sample, the correlation between intrapersonal experiences (patients' perspective) and CBT outcome was not statistically significant in patients with low ($M - 1\text{ SD}$) incongruence (K-INK score 2.160) ($b = -.038$; $p = .249$). By contrast, higher intrapersonal experiences were significantly associated with less severe psychological symptoms at post-treatment in patients with moderate (M) incongruence (K-INK score of 2.890) ($b = -.083$; $p = .001$) as well as in patients with high ($M + 1\text{ SD}$) incongruence (K-INK score of 3.619) ($b = -.127$; $p < .001$). The overall main effect of early patient-rated intrapersonal experiences was no longer significant when incongruence was taken into account. Figure 1 displays the interaction of incongruence and patient-rated intrapersonal experiences with respect to the outcome. It can be seen that the outcome improved negligible with increasing patient-rated intrapersonal experiences when incongruence was low, but the outcome improved substantially with patient-rated intrapersonal experiences when incongruence was high. The Johnson-Neyman Technique showed that higher early intrapersonal experiences (from the patients' point of view) were negatively correlated with (i.e., beneficially for) CBT outcome only for patients with a

baseline K-INK total score above 2.434, that is medium to high incongruence in our sample (see Table 1 for the mean and the standard deviation of the K-INK total score). Figure 2 shows a histogram for the K-INK total score with a dashed line at 2.434. The association between early intrapersonal experiences from the therapists' point of view and CBT outcome was, however, not moderated by incongruence ($p = .808$).

Results for the explorative moderation analyses

Incongruence as a moderator of the relation between early interpersonal experiences and CBT outcome

Table 4 summarizes the results of the moderation analyses for the early GCM interpersonal experiences. Incongruence did not moderate the association between early interpersonal experiences and CBT outcome neither when early interpersonal experiences were rated by the patients ($p = .290$) nor when they were rated by the therapists ($p = .868$). The main effect of early interpersonal experiences from the patients' perspective was no longer significant in the model. Figure 3 depicts the non-significant interaction of incongruence and the patient-rated interpersonal experiences with respect to the outcome. It can be seen that lower patient-rated interpersonal experiences tended to be associated with a worse outcome when incongruence was high. When patient-rated interpersonal experiences were high, however, this pattern was reversed and higher patient-rated interpersonal experiences tended to be associated with a better outcome when incongruence was high.

Incongruence as a moderator of the relation between early problem actuation and CBT outcome

The results of the moderation analyses for the early GCM problem actuation are presented in Table 5. Incongruence moderated the association between early problem actuation and CBT outcome neither when early problem actuation was rated by the patients ($p = .890$) nor when it was rated by the therapists ($p = .778$).

Discussion

The present study investigated whether associations between early GCMs (assessed after five therapy-preparing sessions) and the outcome of CBT (psychological symptoms) depend on the patients' pre-treatment levels of motivational incongruence (according to Grawe's consistency theory). Patients' as well as therapists' ratings of three early GCMs (interpersonal experiences, intrapersonal experiences, and problem actuation) were analyzed. Before we discuss the results of the moderation analyses, the associations between early GCMs and the outcome without taking incongruence into account will be shortly summarized here (a more detailed integration of these results into the existing literature will be given below). Early GCMs in the therapists' perspective were not correlated with the outcome, whereas patient-rated early inter- and intrapersonal experiences predicted CBT outcome. This shows that it is important for therapists not to rely on their perception of such early in-session experiences and that monitoring of early GCMs from the patient's perspective is advisable. In the following, we start the discussion with the results of the theory-guided moderation analysis of incongruence as a moderator of the relation between early intrapersonal experiences and the outcome. After that, the results of the explorative moderation analyses will be considered.

As hypothesized, baseline incongruence moderated the association between early intrapersonal experiences and CBT outcome. Yet, this was the case only for patient-rated early intrapersonal experiences but not for therapist-rated ones. Higher early intrapersonal experiences as rated by the patients became increasingly beneficial for the outcome of CBT at higher levels of baseline incongruence. For patients with less incongruence at baseline, early intrapersonal experiences from the patients' point of view were not correlated with CBT outcome. Intrapersonal experiences comprise clarification of meaning as well as mastery/coping. Our results indicate that therapists should focus on these processes in the early phase of CBT particularly in patients with medium to high levels of incongruence.

This finding conforms to Grawe's (2004) consistency theory as explained above. Patients with high incongruence have a large gap between motivational goals and coping strategies. They might benefit the most from a focus on intrapersonal experiences because the GCMs mastery/coping and especially clarification should reduce the gap between behavior and motivational goals. By clarifying the meaning behind symptoms and making underlying motivational goals explicit, incongruence can be reduced and – as a result – also psychological symptoms. According to the Transtheoretical Model (Prochaska & Velicer, 1997) highly incongruent patients could be entering treatment in a different motivational phase than patients with lower incongruence and therefore might need different processes of change. Highly incongruent patients could be stuck between contemplation and preparation, still procrastinating actions and lacking consciousness for their underlying problems. These patients are likely to benefit from interventions like clarification which raise consciousness. Patients with lower incongruence, on the other hand, might be closer to the action and maintenance phase, having more insight into their problems and might benefit more strongly from disorder-specific interventions. According to the Rubicon Model (Heckhausen & Gollwitzer, 1987; Grawe, 1998), incongruent patients might not yet have crossed the line between the choosing and pre-action phase and might, thus, benefit from a stronger focus on resolving conflicting and incongruent thoughts and goals. In sum, all three theoretical models (consistency theory, Transtheoretical Model, Rubicon Model) agree that clarification and other interventions addressing inconsistencies between motivational goals and behaviors should be more effective for patients experiencing high motivational incongruence. Our results on early patient-rated intrapersonal experiences concur. Therefore, it is recommended to identify highly incongruent patients to assist treatment planning and to adapt interventions according to the motivational phase the patient enters treatment. A psychometrically sound instrument to measure incongruence is the K-INK but this instrument consists of 23 items and the

development of an even shorter instrument might be necessary to economically measure incongruence under the conditions of routine care.

Considering our explorative analyses, patients' baseline incongruence was a moderator neither between early interpersonal experiences (patients' and therapists' ratings) and the outcome nor between early problem actuation (patients' and therapists' ratings) and the outcome. As patient-rated early interpersonal experiences predicted a more favorable outcome in the total sample, therapists should work on early interpersonal experiences with all patients regardless of baseline incongruence.

The lack of associations between therapists' assessments of early GCMs and CBT outcome across all patients is partially in line with results reported by Flückiger et al. (2013). In their study, therapists' ratings of early problem actuation and therapists' ratings of early intrapersonal experiences were not correlated with the outcome, but therapists' ratings of early interpersonal experiences were. Comparable to the current study, other studies also found that patients' ratings of GCMs are better predictors of the outcome than therapists' ratings (e. g., Mander et al., 2013; 2015b). Therapists' ratings of the alliance, however, have been shown to be stronger predictors of patient-reported outcomes under some circumstances, for example, in more effective therapies (Zilcha-Mano, Muran, Eubanks, Safran, & Winston, 2018) or in patients scoring higher on alexithymia (Probst et al., 2017).

Our finding that neither therapists' nor patients' ratings of early problem actuation were associated with CBT outcome fits to other studies (Flückiger et al., 2013; Mander et al., 2013; 2015b). In contrast, Stangier et al. (2010) found problem actuation to be correlated with the outcome. One explanation for this discrepancy might be that Stangier et al. (2010) investigated a relatively homogenous sample of patients with social phobia, whereas we investigated a more heterogeneous sample comprising patients with different psychological disorders. Problem actuation might also be used more often in the treatment of social phobia.

Specific treatment elements like role plays, behavioral experiments, and exposure are conducive to experiencing the problem during treatment sessions (Clark & Wells, 1995).

Gassmann and Grawe (2006) argue that problem actuation needs to be combined with resource activation in order to reach its therapeutic effect. As we did not analyze resource activation, we cannot know whether patients had the required resources and whether these were activated during the first treatment sessions. This is a limitation of the current study. Rubel and colleagues (2017) found that problem actuation shows differential associations with the outcome between- and within-patients. The fact that we analyzed the between-patient-level without taking the within-patient-level into account can be seen as a limitation of the study. To investigate associations between GCMs and the outcome on a within-patient-level, assessing GCMs and outcome measures in parallel on a session-to-session basis would be optimal. Yet, the monitoring procedure in the investigated outpatient center did not include session-to-session assessments of GCMs and outcome measures.

A further shortcoming is that the post-assessment of the outcome measure was not collected after the end of the psychotherapy, but up to five sessions before the actual termination. Although this procedure is clinically meaningful so that therapists can discuss the results of the assessment with the patients in the last sessions, the post-assessment in the present study can also be seen as a process-measure since some symptom change might also occur in the last sessions.

Furthermore, we only analyzed completers as did Flückiger et al. (2013) who stated that “the lack of information on the rest of the intent-to-treat group may bias the results. For example, it may be that patients with low process evaluations dropped out of treatment before regular termination, which may lead to an underestimation of the process-outcome relationships” (Flückiger et al., 2013, p. 9).

Another limitation is our use of a very general symptom severity measure at pre- and post-treatment. Although this measure applies to a variety of different patients, there are

measures which capture different areas of everyday life or which capture treatment progress in a more general way, such as the Outcome Questionnaire (OQ, Lambert et al., 1996). In our study, a measure like the OQ was not administered to the patients. By using the SCL-90R instead, our outcome is restricted to symptom severity. Whether aspects of everyday life are affected by GCMs and moderated by incongruence has to be evaluated in future studies.

Our GCM-measure can also be seen as a limitation. The BPSR is designed to measure GCMs in the last attended session and not over a longer period of sessions. Since we applied the BPSR only in the fifth treatment preparing session it is unclear how the GCMs developed in the early stage of treatment within patients. This is a major difference between our study and the one conducted by Flückiger et al. (2013) who used the GCM means over the first five sessions to predict outcome.

The major limitation, however, is the correlational analyses, which does not allow to draw causal inferences. The internal validity of the results is, therefore, rather low. We used a longitudinal design with pre-, monitoring- and posttreatment measurements that allows interpreting the directionality of the effect but the best internal validity would be achieved through a RCT. In a RCT, high internal validity would result from randomizing patients with higher baseline incongruence levels either to a treatment including a component to foster early intrapersonal experiences or to a treatment excluding this component. The hypothesis according to our results would be that patients with higher incongruence benefit more from a treatment including a component to foster early intrapersonal experiences. Such a trial would be different from the one conducted by Grosse Holtforth et al. (2008), who compared a treatment (CBT alone) including one part of intrapersonal experiences (mastery) with another treatment (experiential and cognitive-behavioral psychotherapy) comprising the two parts of intrapersonal experiences (mastery and clarification of meaning).

Although the internal validity of the presented correlative study is rather low, its external validity is high. The CBT psychotherapies were conducted under the conditions of

routine practice and are typical for post-graduate training in Germany. However, the average number of sessions was rather high in comparison to other countries. Therefore, the results might be less generalizable to outpatient centers outside of German-speaking countries. As we only investigated CBT, the results are also not representative for other psychotherapeutic approaches.

To conclude, the clinical implications of the present study are that CBT outcome might be improved when (i) therapists foster early intrapersonal experiences in the patients' view specifically in patients with medium to high levels of incongruence and when (ii) therapists foster early interpersonal experiences in the patients' view in all patients regardless of the patients' incongruence.

References

- Berking, M., Grosse Holtforth, M., & Jacobi, C. (2003). Reduction of incongruence in inpatient psychotherapy. *Clinical Psychology & Psychotherapy*, *10*, 86-92. <https://doi.org/10.1002/cpp.357>
- Bradford, A., Cully, J., Rhoades, H., Kunik, M., Kraus-Schuman, C., Wilson, N., & Stanley, M. (2011). Early response to psychotherapy and long-term change in worry symptoms in older adults with generalized anxiety disorder. *The American Journal of Geriatric Psychiatry : Official Journal of the American Association for Geriatric Psychiatry*, *19*, 347–356. <https://doi.org/10.1097/JGP.0b013e3181f18061>.
- Call, M., D’Errico, L., Bents, H., & Mander, J. (2018). General change factors in outpatient training therapy. *Psychotherapeut*, *63*, 347-356. <https://doi.org/10.1007/s00278-018-0310-4>
- Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. G. Heimberg, M. R. Liebowitz, D. A. Hope, & F. R. Schneider (Eds.), *Social phobia: Diagnosis assessment, and treatment* (pp. 69–93). New York: The Guildford Press.
- Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology*, *14*, 209–236. <https://doi.org/10.1146/annurev-clinpsy-050817-084746>
- Cuijpers, P., Reijnders, M., & Huibers, M. J. H. (2018). The Role of Common Factors in Psychotherapy Outcomes. *Annual Review of Clinical Psychology*. *15*, 207-231. <https://doi.org/10.1146/annurev-clinpsy-050718-095424>
- Derogatis, L. R. (1977). *SCL-90-R, administration, scoring & procedures manual-I for the R(evised) version*. Johns Hopkins University of Medicine: Eigendruck.
- Flood, N., Page, A., & Hooke, G. (2018). A comparison between the clinical significance and growth mixture modelling early change methods at predicting negative outcomes.

Psychotherapy Research, epub ahead of print, 1–12.

<https://doi.org/10.1080/10503307.2018.1469803>

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, 55, 316-340.

<https://doi.org/10.1037/pst0000172>

Flückiger, C., Grosse Holtforth, M., Znoj, H. J., Caspar, F., & Wampold, B. E. (2013). Is the relation between early post-session reports and treatment outcome an epiphenomenon of intake distress and early response? A multi-predictor analysis in outpatient psychotherapy.

Psychotherapy Research, 23, 1–13. <https://doi.org/10.1080/10503307.2012.693773>

Flückiger, C., Regli, D., Zwahlen, D., Hostettler, S., & Caspar, F. (2010). The Bern Post Session Report 2000. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, 39, 71–

79. <https://doi.org/10.1026/1616-3443/a000015>

Franke, G. H. (2002). *SCL-90-R: Symptom Checkliste von L.R. Derogatis (deutsche Version)*. Göttingen: Beltz.

Gassmann, D., & Grawe, K. (2006). General change mechanisms: The relation between problem activation and resource activation in successful and unsuccessful therapeutic interactions. *Clinical Psychology & Psychotherapy*, 13, 1–11.

<https://doi.org/10.1002/cpp.442>

Grawe, K. (1998). *Psychologische Therapie*. Göttingen: Hogrefe.

Grawe, K. (1997). Research-informed psychotherapy. *Psychotherapy Research*, 7, 1–19.

<https://doi.org/10.1080/10503309712331331843>

Grawe, K. (2004). *Neuropsychotherapie*. Göttingen: Hogrefe.

Grosse Holtforth, M., & Grawe, K. (2000). Questionnaire for the Analysis of Motivational Schemas. *Zeitschrift für Klinische Psychologie und Psychotherapie: Forschung und Praxis*, 29, 170-179. <http://doi.org/10.1026//0084-5345.29.3.170>

- Grosse Holtforth, M., & Grawe, K. (2002). *FAMOS. Fragebogen zur Analyse Motivationaler Schemata*. Göttingen: Hogrefe.
- Grosse Holtforth, M., & Grawe, K. (2003). The incongruence questionnaire (INK). An instrument for the analysis of motivational incongruence. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *32*, 315-323 <https://doi.org/10.1026/0084-5345.32.4.315>
- Grosse Holtforth, M., Grawe, K. & Castonguay, L. G. (2006). Predicting a reduction of avoidance motivation in psychotherapy: Toward the delineation of differential processes of change operating at different phases of treatment. *Psychotherapy Research*, *16*, 639-644. <https://doi.org/10.1080/10503300600608215>
- Grosse Holtforth, M., Grawe, K., Fries, A., & Znoj, H. (2008). Inconsistency as a criterion for differential indication in psychotherapy. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *37*, 103–111. <https://doi.org/10.1026/1616-3443.37.2.103>
- Grosse Holtforth, M., Grawe, K., Tamcan, Ö. (2004). *INK: Inkongruenzfragebogen*. Göttingen: Hogrefe.
- Grosse Holtforth, M., Pincus, A. L., Grawe, K., Mauler, B. & Castonguay, L. G. (2007). When what you want is not what you get: Motivational correlates of interpersonal problems in clinical and nonclinical samples. *Journal of Social and Clinical Psychology*, *26*, 1095-1119. <https://doi.org/10.1521/jscp.2007.26.10.1095>
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. 2nd Edition*. Guilford Publications.
- Hayes, A. M., Feldman, G. C., Beevers, C. G., Laurenceau, J. P., Cardaciotto, L., & Lewis-Smith, J. (2007). Discontinuities and cognitive changes in an exposure-based cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, *75*, 409-421. <http://doi.org/10.1037/0022-006X.75.3.409>

- Heckhausen, H., & Gollwitzer, P. M. (1987). Thought Contents and Cognitive Functioning in Motivational versus Volitional States of Mind. *Motivation and Emotion, 11*(2), 101–120.
- Huber, J., Born, A. K., Claaß, C., Ehrental, J. C., Nikendei, C., Schauenburg, H., & Dinger, U. (2018). Therapeutic agency, in-session behavior, and patient-therapist interaction. *Journal of Clinical Psychology*, epub ahead of print. <https://doi.org/10.1002/jclp.22700>
- IBM Corp. (2017). IBM SPSS Statistics for Windows [Computer software], New York: IBM Corp.
- Kleinstäuber, M., Lambert, M. J., & Hiller, W. (2017). Early response in cognitive-behavior therapy for syndromes of medically unexplained symptoms. *BMC Psychiatry, 17*, 195. <https://doi.org/10.1186/s12888-017-1351-x>
- Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., & Yanchar, S. C. (1996). The Reliability and Validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy, 3*, 249–258.
- Lorenzo-Luaces, L., DeRubeis, R. J., & Webb, C. A. (2014). Client characteristics as moderators of the relation between the therapeutic alliance and outcome in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 82*, 368–373. <https://doi.org/10.1037/a0035994>.
- Lorenzo-Luaces, L., Driessen, E., DeRubeis, R. J., Van, H. L., Keefe, J. R., Hendriksen, M., & Dekker, J. (2017). Moderation of the alliance-outcome association by prior depressive episodes: Differential effects in Cognitive-Behavioral Therapy and Short-Term Psychodynamic Supportive Psychotherapy. *Behavior Therapy, 48*, 581–595. <https://doi.org/10.1016/j.beth.2016.11.011>
- Mander, J. V., Jacob, G. A., Götz, L., Sammet, I., Zipfel, S., & Teufel, M. (2015a). Associations between Grawe's general mechanisms of change and Young's early

- maladaptive schemas in psychotherapy research: a comparative study of change processes. *Psychotherapy Research*, 25, 249-262. <https://doi.org/10.1080/10503307.2014.889330>
- Mander, J., Schlarb, A., Teufel, M., Keller, F., Hautzinger, M., Zipfel, S., . . . Sammet, I. (2015b). The individual therapy process questionnaire: Development and validation of a revised measure to evaluate general change mechanisms in psychotherapy. *Clinical Psychology & Psychotherapy*, 22, 328–345. <https://doi.org/10.1002/cpp.1892>
- Mander, J. V., Wittorf, A., Schlarb, A., Hautzinger, M., Zipfel, S., & Sammet, I. (2013). Change mechanisms in psychotherapy: Multiperspective assessment and relation to outcome. *Psychotherapy Research*, 23, 105–116. <https://doi.org/10.1080/10503307.2012.744111>
- Muran, J. C., Segal, Z. V., Samstag, L. W., & Crawford, C. E. (1994). Patient pretreatment interpersonal problems and therapeutic alliance in short-term cognitive therapy. *Journal of Consulting and Clinical Psychology*, 62, 185-190. <http://doi.org/10.1037/0022-006X.62.1.185>
- Probst, T., Sattel, H., Gündel, H., Henningsen, P., Kruse, J., Schneider, G., & Lahmann, C. (2017). Moderating effects of alexithymia on associations between the therapeutic alliance and the outcome of Brief Psychodynamic-Interpersonal Psychotherapy for multisomatoform disorder. *Frontiers in Psychiatry*, 8, 261. <https://doi.org/10.3389/fpsy.2017.00261>
- Prochaska, J. O., & Velicer, W. F. (1997). The Transtheoretical Model of Health Behavior Change. *American Journal of Health Promotion*, 12(1), 38–48.
- Rubel, J. A., Rosenbaum, D., & Lutz, W. (2017). Patients' in-session experiences and symptom change: Session-to-session effects on a within- and between-patient level. *Behavior Research and Therapy*, 90, 58–66. <https://doi.org/10.1016/j.brat.2016.12.007>

- Schneider, R. L., Arch, J. J., & Wolitzky-Taylor, K. B. (2015). The state of personalized treatment for anxiety disorders: A systematic review of treatment moderators. *Clinical Psychology Review, 38*, 39–54. <https://doi.org/10.1016/j.cpr.2015.02.004>
- Stangier, U., Consbruch, K. von, Schramm, E., & Heidenreich, T. (2010). Common factors of cognitive therapy and interpersonal psychotherapy in the treatment of social phobia. *Anxiety, Stress, and Coping, 23*, 289–301. <https://doi.org/10.1080/10615800903180239>
- Van, H. L., Schoevers, R. A., Kool, S., Hendriksen, M., Peen, J., & Dekker, J. (2008). Does early response predict outcome in psychotherapy and combined therapy for major depression? *Journal of Affective Disorders, 105*(1-3), 261–265. <https://doi.org/10.1016/j.jad.2007.04.016>
- World Health Organization (1990). *Composite International Diagnostic Interview (CIDI)*. World Health Organization, Division of Mental Health: Geneva
- World Health Organization (1992). *The ICD-10 Classification of Mental and Behavioral Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: WHO. Retrieved from <http://www.who.int/classifications/icd/en/bluebook.pdf>
- Zack, S. E., Castonguay, L. G., Boswell, J. F., McAleavey, A. A., Adelman, R., Kraus, D. R., & Pate, G. A. (2015). Attachment history as a moderator of the alliance outcome relationship in adolescents. *Psychotherapy, 52*, 258–267. <https://doi.org/10.1037/a0037727>
- Zilcha-Mano, S. (2018). Major developments in methods addressing for whom psychotherapy may work and why. *Psychotherapy Research*, epub ahead of print, 1-16. <https://doi.org/10.1080/10503307.2018.1429691>
- Zilcha-Mano, S., & Errázuriz, P. (2015). One size does not fit all: Examining heterogeneity and identifying moderators of the alliance-outcome association. *Journal of Counseling Psychology, 62*, 579–591. <https://doi.org/10.1037/cou0000103>

Zilcha-Mano, S., Muran, J. C., Eubanks, C. F., Safran, J. D., & Winston, A. (2018). When therapist estimations of the process of treatment can predict patients rating on outcome: The case of the working alliance. *Journal of Consulting and Clinical Psychology, 86*, 398-402. <http://doi.org/10.1037/ccp0000293>

Table 1: Sample description with means and standard deviations in parentheses when applicable.

Sample size	524	
Patient age <i>M (SD)</i>	39.97	(13.12)
Patient sex	58.4%	
In a relationship	female	
Unemployed	61%	
Psychopharmacological medication	12%	
Therapist age <i>M (SD)</i>	48%	
Therapist sex	30.94	(7.39)
Patients per therapist <i>M (SD)</i>	86.2%	
Number of sessions <i>M (SD)</i>	female	
At least one comorbid disorder	4.75	(3.33)
Number of diagnoses <i>M (SD)</i>	49.29	(17.22)
Pre GSI <i>M (SD)</i>	61%	
Post GSI <i>M (SD)</i>	1.99	(1.03)
Pre-post effect size (Cohen's <i>d</i>)	1.03	(0.65)
K-INK total score <i>M (SD)</i>	0.59	(0.58)
	0.70	
	2.89	(0.73)

Note: GSI = Global Severity Index of the SCL-90R; K-INK total = total score of the incongruence questionnaire short form at pre-treatment.

Table 2: Partial correlations between incongruence and CBT outcome as well as between early general change mechanisms and CBT outcome (Global Severity Index of the SCL-90R at pre-treatment and number of sessions were control variables).

		Symptom severity at post-treatment ¹
	Incongruence ²	<i>r</i> 0.037 <i>p</i> 0.395
Patients' ratings of	Interpersonal experiences ³	<i>r</i> -0.158 <i>p</i> <0.001
	Intrapersonal experiences ³	<i>r</i> -0.158 <i>p</i> <0.001
	Problem actuation ³	<i>r</i> -0.005 <i>p</i> 0.906
Therapists' ratings of	Interpersonal experiences ³	<i>r</i> -0.080 <i>p</i> 0.068
	Intrapersonal experiences ³	<i>r</i> -0.053 <i>p</i> 0.228
	Problem actuation ³	<i>r</i> 0.022 <i>p</i> 0.616

Note: N = 524, ¹ symptom severity was measured with the Global Severity Index of the SCL-90R; ² incongruence was measured with the total score of the incongruence questionnaire short form at pre-treatment; ³ Interpersonal experiences, intrapersonal experiences and problem actuation were measured with the Bern Post Session Report after five therapy-preparing sessions.

Table 3: Results of the moderation analyses testing incongruence as a moderator of the relation between early intrapersonal experiences and CBT outcome.

Outcome: GSI at post-treatment							
Patients' ratings of early intrapersonal experiences	Model Summary						
	<i>R</i>	<i>R-sq</i>	<i>MSE</i>	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
	0.615	0.378	0.212	63.012	5.000	518.000	<0.001
	Model						
		<i>Coefficient</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>LLCI</i>	<i>ULCI</i>
	Constant	-0.168	0.121	-1.397	0.163	-0.405	0.068
	BPSR IntraExp	0.094	0.090	1.047	0.296	-0.083	0.271
	K-INK total	0.068	0.042	1.626	0.105	-0.014	0.149
	BPSR IntraExp *						
	K-INK total	-0.061	0.030	-2.065	0.039	-0.119	-0.003
GSI at pre-treatment	0.482	0.039	12.240	<0.001	0.405	0.559	
Number of sessions	0.003	0.001	2.069	0.039	<0.001	0.005	
Therapists' ratings of early intrapersonal experiences	Model Summary						
	<i>R</i>	<i>R-sq</i>	<i>MSE</i>	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
	0.600	0.359	0.218	58.131	5.000	518.000	<0.001
	Model						
		<i>Coefficient</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>LLCI</i>	<i>ULCI</i>
	Constant	-0.104	0.302	-0.344	0.731	-0.697	0.489
	BPSR IntraExp	-0.006	0.122	-0.051	0.959	-0.246	0.234
	K-INK total	0.055	0.102	0.536	0.592	-0.145	0.254
	BPSR IntraExp *						
	K-INK total	-0.010	0.042	-0.243	0.808	-0.092	0.072
GSI at pre-treatment	0.490	0.040	12.247	<0.001	0.412	0.569	
Number of sessions	0.002	0.001	1.971	0.049	<0.001	0.005	

Note: GSI = Global Severity Index of the SCL-90R; BPSR IntraExp = intrapersonal experiences scale of the Bern Post Session Report after five therapy-preparing sessions; K-INK total = total score of the incongruence questionnaire short form at pre-treatment. N = 524.

Table 4: Results of the moderation analyses testing incongruence as a moderator of the relation between early interpersonal experiences and CBT outcome.

Outcome: GSI at post-treatment							
Patients’ ratings of early interpersonal experiences	Model Summary						
	<i>R</i>	<i>R-sq</i>	<i>MSE</i>	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
	0.612	0.374	0.213	61.948	5.000	518.000	<0.001
	Model						
		<i>Coefficient</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>LLCI</i>	<i>ULCI</i>
	Constant	-0.120	0.245	-0.491	0.624	-0.602	0.362
	BPSR InterExp	0.012	0.115	0.106	0.916	-0.214	0.239
	K-INK total	0.088	0.080	1.100	0.272	-0.069	0.245
	BPSR InterExp *						
	K-INK total	-0.040	0.038	-1.059	0.290	-0.114	0.034
GSI at pre-treatment	0.494	0.039	12.547	<0.001	0.416	0.571	
Number of sessions	0.003	0.001	2.456	0.014	0.001	0.005	
Therapists’ ratings of early interpersonal experiences	Model Summary						
	<i>R</i>	<i>R-sq</i>	<i>MSE</i>	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
	0.601	0.361	0.218	58.631	5.000	518.000	<0.001
	Model						
		<i>Coefficient</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>LLCI</i>	<i>ULCI</i>
	Constant	-0.014	0.189	-0.076	0.940	-0.385	0.356
	BPSR InterExp	-0.066	0.106	-0.624	0.533	-0.274	0.142
	K-INK total	0.017	0.065	0.259	0.796	-0.110	0.144
	BPSR InterExp *						
	K-INK total	0.006	0.037	0.166	0.868	-0.066	0.079
GSI at pre-treatment	0.498	0.040	12.514	<0.001	0.419	0.576	
Number of sessions	0.003	0.001	2.081	0.038	<0.001	0.005	

Note: GSI = Global Severity Index of the SCL-90R; BPSR InterExp = interpersonal experiences scale of the Bern Post Session Report after five therapy-preparing sessions; K-INK total = total score of the incongruence questionnaire short form at pre-treatment. N = 524.

Table 5: Results of the moderation analyses testing incongruence as a moderator of the relation between early problem actuation and CBT outcome.

Outcome: GSI at post-treatment							
Patients’ ratings of early problem actuation	Model Summary						
	<i>R</i>	<i>R-sq</i>	<i>MSE</i>	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
	0.598	0.358	0.219	57.654	5.000	518.000	<0.001
	Model						
		<i>Coefficient</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>LLCI</i>	<i>ULCI</i>
	Constant	-0.117	0.105	-1.112	0.267	-0.323	0.089
	BPSR PA	-0.009	0.057	-0.157	0.875	-0.122	0.104
	K-INK total	0.029	0.038	0.762	0.446	-0.045	0.102
	BPSR PA * K-INK total	0.003	0.019	0.138	0.890	-0.035	0.041
	GSI at pre-treatment	0.495	0.040	12.407	<0.001	0.417	0.574
Number of sessions	0.003	0.001	2.027	0.043	<0.001	0.005	
Therapists’ ratings of early problem actuation	Model Summary						
	<i>R</i>	<i>R-sq</i>	<i>MSE</i>	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
	0.598	0.358	0.219	57.752	5.000	518.000	<0.001
	Model						
		<i>Coefficient</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>LLCI</i>	<i>ULCI</i>
	Constant	-0.094	0.211	-0.447	0.655	-0.509	0.320
	BPSR PA	-0.014	0.098	-0.142	0.887	-0.205	0.178
	K-INK total	0.013	0.072	0.182	0.855	-0.128	0.154
	BPSR PA * K-INK total	0.009	0.033	0.282	0.778	-0.056	0.075
	GSI at pre-treatment	0.496	0.040	12.442	<0.001	0.418	0.575
Number of sessions	0.002	0.001	1.946	0.052	<0.001	0.005	

Note: GSI = Global Severity Index of the SCL-90R; BPSR PA = problem actuation scale of the Bern Post Session Report after five therapy-preparing sessions; K-INK total = total score of the incongruence questionnaire short form at pre-treatment. N = 524.

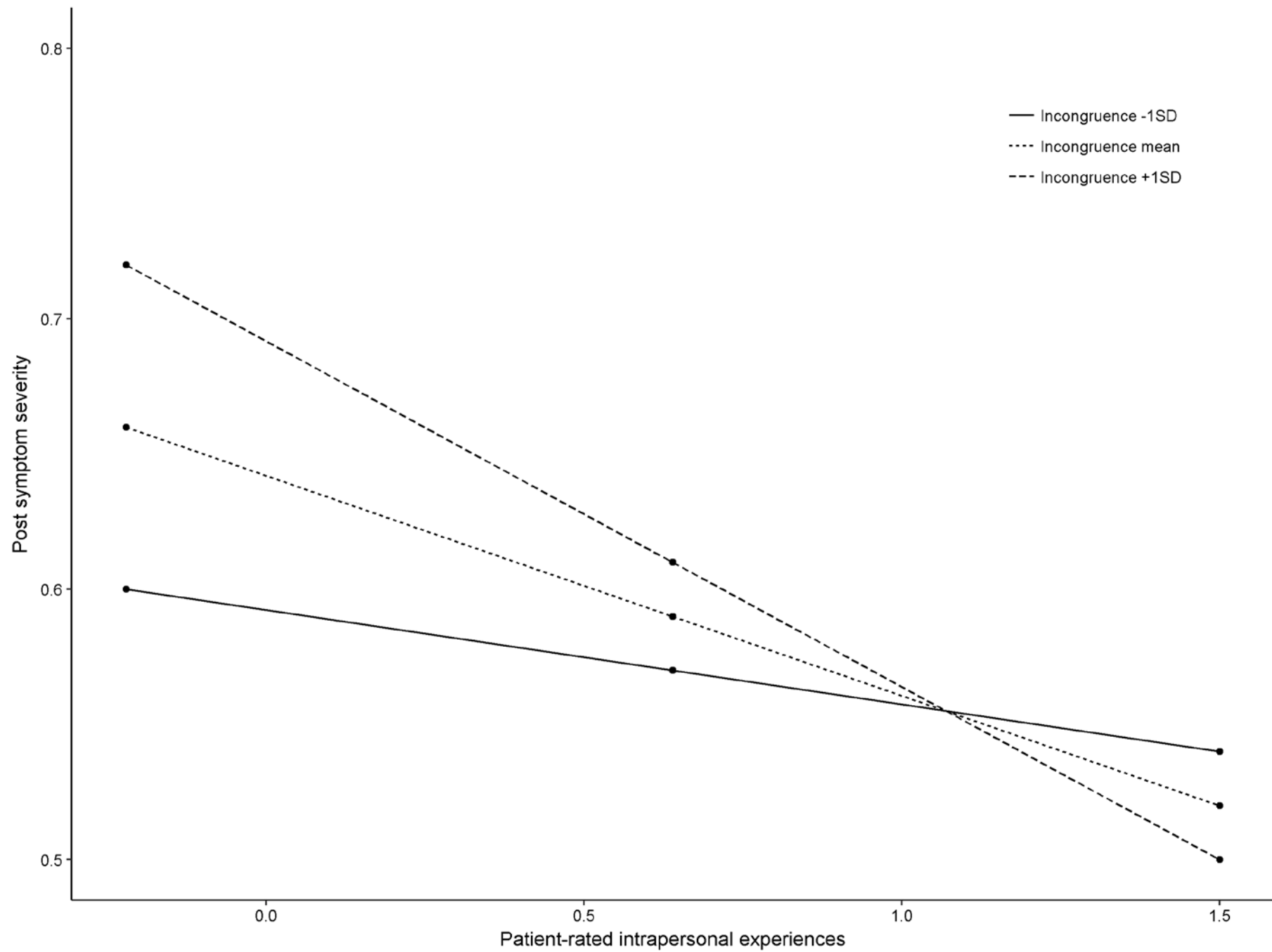


Figure 1. Relationship of early patient-rated intrapersonal experiences and CBT-outcome. The different lines depict different levels of motivational incongruence at baseline. Lines correspond to estimates from the statistical model and do not represent empirical values.

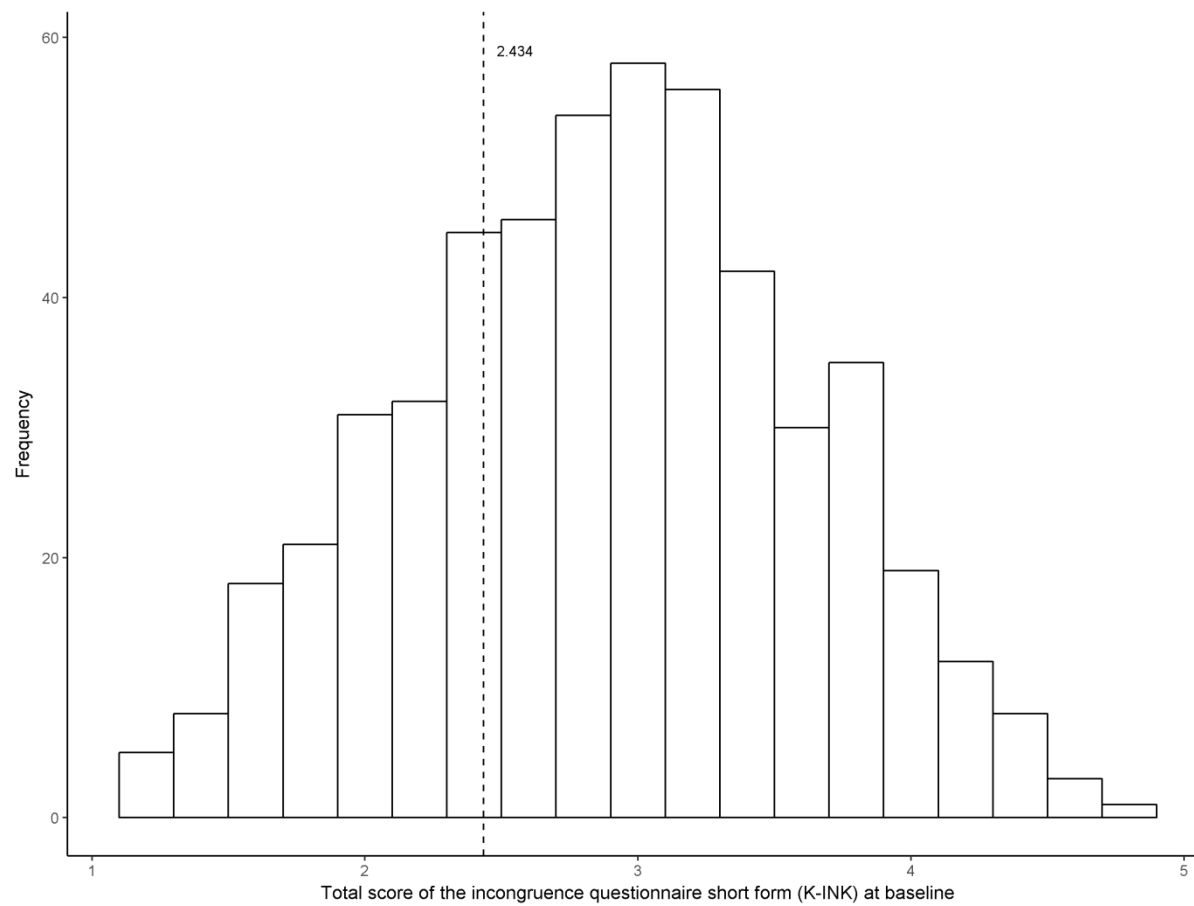


Figure 2. Histogram of motivational incongruence scores at baseline. Early patient-rated intrapersonal experiences were beneficial for the outcome only in patients with values above 2.434 (dashed line).

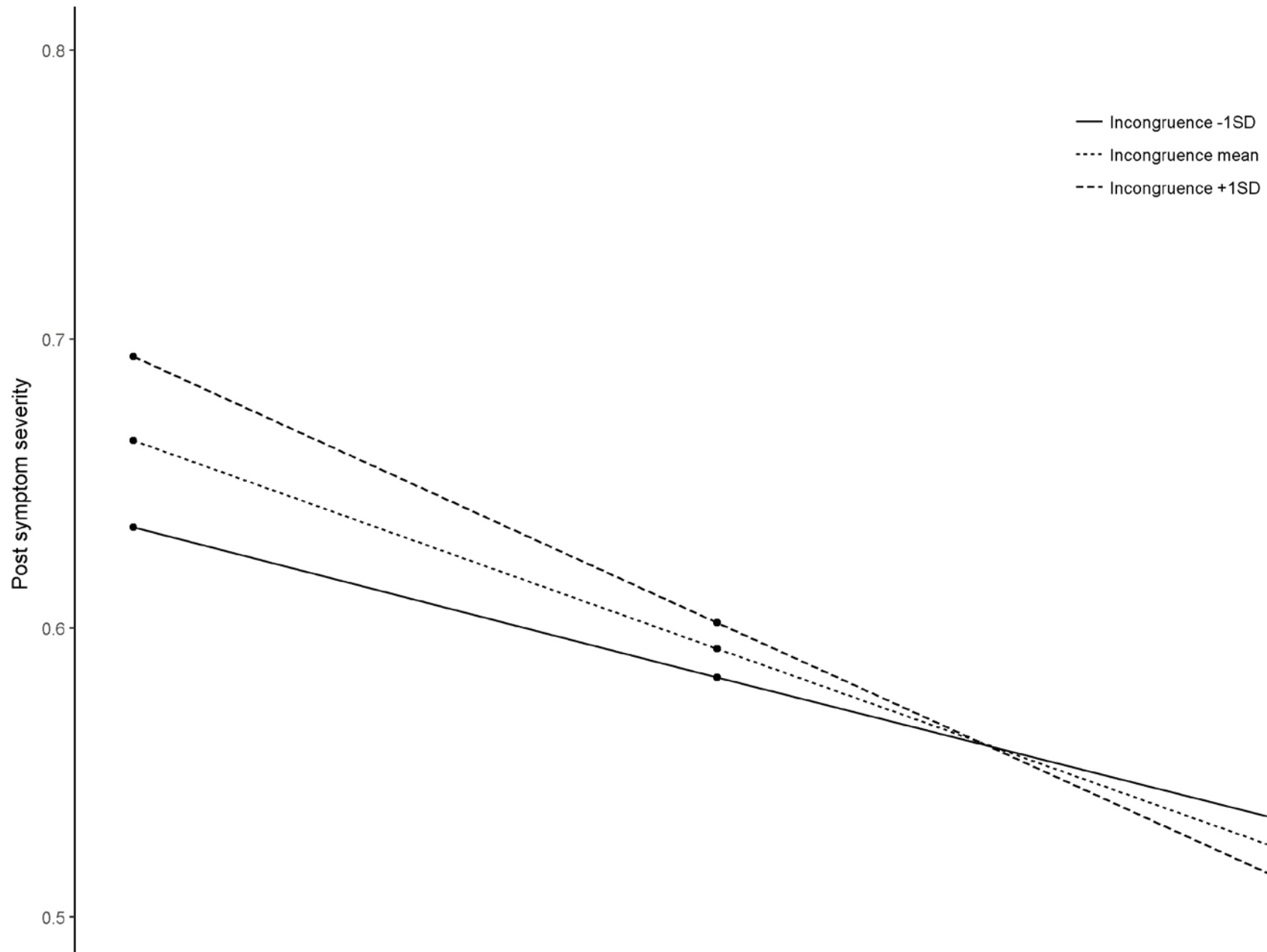


Figure 3. Relationship of early patient-rated interpersonal experiences and CBT-outcome. The different lines depict different levels of motivational incongruence at baseline. Lines correspond to estimates from the statistical model and do not represent empirical values.

Supplemental Material

Supplement Table 1: Applied measures at pre-treatment.

Measure	Abbreviation	Scope
Beck Anxiety Inventory	BAI	Anxiety
Beschwerde Liste (Complaint List)	BL	Chronic Pain
European Quality of Life	EuroQoL	Quality of Life
Allgemeine Depressionsskala Kurzform (General Depression Scale Short Form)	ADS-K	Depression
Incongruence Questionnaire Short Form	K-INK	Incongruence
Pain Disability Index	PDI	Chronic Pain
Patient demographics	-	Demographics
Questions on Life Satisfaction	QLS	Life Satisfaction
Social Interaction Anxiety Scale	SIAS	Social Anxiety
Symptom Checklist	SCL-90R	Symptom Severity
Treatment Goals	-	Treatment Goals
Waiting Time Questionnaire	-	Waiting Time

Gmeinwieser, Schneider, Bardo, Brockmeyer und Hagmayer (in press)

Gmeinwieser, S., Schneider, K. S., Bardo, M., Brockmeyer, T., Hagmayer, Y. (in press). Risk for psychotherapy drop-out in survival analysis: The influence of general change mechanisms and symptom severity. *Journal of Counseling Psychology*.
<https://doi.org/10.1037/cou0000418>

©American Psychological Association, 2019. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/cou0000418

Title:

Risk for psychotherapy drop-out in survival analysis: The influence of general change mechanisms and symptom severity

Running Title:

Drop-out and general change mechanisms

Authors:

Sebastian Gmeinwieser¹, Kai Sebastian Schneider², Maximilian Bardo², Timo Brockmeyer¹,
York Hagmayer³

Affiliations:

¹ Department of Clinical Psychology and Psychotherapy, University of Goettingen, Germany

² Department for Applied Statistics, University of Goettingen, Germany

³ Department for Cognitive and Decision Sciences, University of Goettingen, Germany

Corresponding Author:

Sebastian Gmeinwieser, M.Sc.

Department of Clinical Psychology and Psychotherapy, University of Goettingen,
Gosslerstrasse 14, D-37073, Germany

E-Mail: Sebastian.Gmeinwieser@psych.uni-goettingen.de

Competing interests statement: The authors declare that there are no conflicts of interest.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Abstract

Dropping out of psychotherapeutic treatment, i.e. the patient ending treatment unilaterally poses a problem for patients, therapists and the health care sector. Previous research showed that changes in symptom severity and general change mechanisms (GCMs), such as interpersonal experiences, intrapersonal experiences, and problem actuation, might be related to drop-out. We investigated the relationship of these predictors and drop-out in a sample of 724 patients (21.1% drop-out) receiving cognitive behavioral therapy in routine care from a German outpatient clinic. Survival analysis was used to account for the longitudinal nature of the data created by routine outcome monitoring and to deal with the time varying predictors, GCMs and changes in symptom severity. As outcome, we predicted the risk of dropping out. Results showed that patient- and therapist-rated interpersonal experiences, which include alliance, significantly predicted the risk for drop-out. Contrary to previous research, intrapersonal experiences and symptom severity change did not predict drop-out. Overall, GCMs and symptom severity change accounted for 3.8% of explained variance in the outcome. These results entail that it is important to monitor interpersonal experiences over the course of treatment to identify patients at risk for drop-out.

Public Significance Statement:

Realization of general change mechanisms and changes in symptom severity over the course of treatment were used to predict drop-out in a survival analysis. Interpersonal experiences showed a significant association with drop-out and it is therefore advisable to monitor interpersonal experiences on a regular basis.

Keywords: Survival analysis, general change mechanisms, symptom severity, treatment outcome, drop-out

Introduction

Psychotherapeutic treatments can end in different ways. Treatment can end unilaterally, i.e. one party terminates treatment (usually the patient), or mutually agreed, for example upon reaching a certain therapeutic goal. Unilateral termination of treatment by the patient is also called attrition or drop-out. Drop-out has been shown to be problematic for the patient, the therapist, and even the health care sector. Patients could leave treatment unimproved (Cahill, 2003; Kazdin, 1996; Knox et al., 2011), therapists could become discouraged from treating further patients (Piselli, Halgin & Macewan, 2011), and unsuccessful treatments funded by health insurance agencies are ill-spent financial resources (Altmann et al., 2018).

In a meta-analysis, Swift and Greenberg (2012) identified several factors that are associated with drop-out considering 669 studies with a total sample of 83,834 patients. Results showed that drop-out rates varied significantly among clients with different diagnoses like anxiety, mood, and personality disorders (16-25%). Moreover, client characteristics, such as lower client age and lower education were associated with drop-out. Subgroup analyses of treatment-related factors revealed that treatment setting and time limitations to treatment were related to drop-out (e.g., university-based clinics and no time limitations to treatment had a higher drop-out rate). Roseborough, McLeod and Wright (2016) showed in an archival study of 3,728 clients that patient characteristics were associated with drop-out. Patient demographics such as higher client age and diagnostic categories such as obsessive-compulsive disorder (OCD), personality and depressive disorder were related to a higher risk for drop-out.

According to Roseborough et al. (2016), another factor associated with drop-out is symptom improvement over the course of treatment. The results showed that greater symptom improvement (i.e. symptom reduction) was associated with a lower risk for drop-

out. Previous research that investigated the relationship of symptom severity, symptom severity change or related constructs and drop-out also found a significant association (e.g. Delgadillo et al., 2014; Xiao et al. 2017; Zimmermann, Rubel, Page & Lutz, 2017). Contrary to this finding, Roos and Werbart (2013) described in a narrative review a set of findings that may seem paradoxical at first sight. Both early gains (i.e., symptom reductions) and a lack of symptom improvement were similarly related to higher drop-out. However, these seemingly contradictory findings become perfectly understandable if one considers that drop-out can result from both patient satisfaction (i.e., the patient ending the treatment despite the therapist recommending a continuation of therapy in order to achieve more sustainable results) and patient dissatisfaction. Roos and Werbart (2013) also highlighted that therapeutic alliance and client dissatisfaction were related to drop-out. Furthermore, lack of agreement on core problems, goals and procedures between patient and therapist was associated with drop-out. Sharf, Primavera, and Diener (2010) conducted a meta-analysis solely on the relation of therapeutic alliance and drop-out including 11 studies with 1,301 patients. They found a considerable effect of $d = 0.55$. These findings strongly suggest that process variables beyond symptom relief are important predictors for drop-out.

In this paper, we focus on a specific set of process variables: General change mechanisms (GCMs). According to Grawe's research informed psychotherapy (1997), GCMs are critical for a therapy to be successful. Flückiger, Grosse Holtforth, Znoj, Caspar and Wampold (2013) reviewed the literature and defined three different types of GCMs, interpersonal experiences, intrapersonal experiences, and problem actuation. These GCMs cover basic processes in psychotherapeutic treatment. Interpersonal experiences refer to the patient-therapist dyad. Patient and therapist have to establish a fruitful working alliance. This is necessary for the patient to feel content with treatment and to remain open for therapeutic interventions. For example, depressed patients might feel shame when talking about suicidal

thoughts and tendencies. Through positive interpersonal experiences, patients might be more open when talking about their own thoughts and feelings. This is also the base for positive intrapersonal experiences, which are defined as situations in which the patient feels more control and more understanding concerning their problems. A patient with agoraphobia might get more insight through daily protocols of avoidance behaviors, which could encourage a subsequent behavioral experiment like taking the bus. Through repeated experiments, the patient might get a feeling of mastery of their problems because panic attacks will become less frequent. However, this is only achieved if exposure in treatment elicit an emotional response that is comparable to the panic attacks that are experienced outside treatment. This emotional response is defined as problem actuation. Through interventions like exposure, the patient (re-)experiences similar emotions (e.g. anxiety) as in their daily life. This experience, if followed by positive inter- and intrapersonal experiences, could lead to therapeutic change.

Flückiger et al. (2013) describe these GCMs in the context of the dual model of psychotherapy (Schulte, 1996; Schulte & Eifert 2002). This model proposes two processes that interact with each other. First, a process that acts as a catalyst for specific treatment techniques encompassing the therapeutic alliance and openness to treatment (i.e. interpersonal experiences). Second, the treatment techniques that foster therapeutic change within the patient like encouraging the patient to master problems and to clarify the meaning of these problems (i.e. intrapersonal experiences). These two processes are considered to be intertwined with each other and a therapist might switch back and forth between these two processes within a single treatment session (Schulte & Eifert, 2002). For example, if a patient refuses exposure in vivo, it might be helpful for the therapist to work on the patient's contentment with treatment and the therapeutic alliance to motivate the patient for this intervention instead of switching to exposure in sensu.

Several studies showed that interpersonal and intrapersonal experiences but not problem actuation were related to therapy outcome (Mander et al., 2015; Flückiger et al., 2013, Gmeinwieser, Hagmayer, Pieh & Probst, 2019a; Gmeinwieser, Kuhlencord, Ruhl, Hagmayer & Probst, 2019b). All of these studies used data from routine outcome monitoring (ROM). To the best of our knowledge, there are only two studies which investigated the association of GCMs and drop-out. Kegel and Flückiger (2015) found an association of drop-out with mastery and clarification (i.e., intrapersonal experiences), the therapeutic alliance (part of interpersonal experiences), and self-esteem in a sample of 296 outpatients with mixed disorders who received cognitive behavioral therapy. Flückiger et al. (2011) only found a significant association between drop-out and self-esteem in a sample of 78 patients with binge eating disorder, but none for therapeutic alliance or mastery and clarification. Hence, at present, rather little is known about the relationship of GCMs and drop-out. The present study intends to provide more evidence in this regard by investigating a large sample of patients receiving cognitive-behavioral therapy in routine care.

For our analyses, we considered the realization of GCMs and changes in symptom severity over time as predictors of drop-out. To this end, we decided to use survival analysis. While it has not been used on research into GCMs, it has been applied successfully in previous research on drop-out (e.g., Corning & Malofeeva; 2004; Roseborough et al., 2016). Survival analysis provides some advantages over more traditional statistical approaches for process-outcome research as described by Corning and Malofeeva (2004): First, drop-out and mutually agreed termination can be considered separately. Often these two ways of ending a treatment are treated as opposite poles. This view is problematic because treatments that are terminated mutually agreed might generally take a longer time than treatments where the patient drops out. Therefore, predictors may differ for these two outcomes when time until treatment termination is considered. Second, both time varying and constant predictors can be

considered in survival analysis. In ROM, variables are measured several times over the course of treatment. Hence, some predictors vary in time (e.g. changes in symptom severity) and some are time-independent (e.g. initial diagnoses). Third, patients vary regarding the number of ROM-sessions until the end of treatment. Survival analysis makes use of all available ROM-sessions for each individual without loss of information.

Previous studies that investigated the association of GCMs and drop-out either did not use survival analysis (Flückiger et al., 2011) or did not simultaneously consider symptom improvement and GCMs (Kegel & Flückiger, 2015). On the other hand, Corning and Malofeeva (2004), who used survival analysis, did not investigate process variables like symptom improvement or GCMs. Roseborough et al. (2016), who also used survival analysis, investigated symptom improvement but not GCMs.

To sum up, based on research on GCMs and in particular on the relation of GCMs and drop-out, we hypothesized that GCMs and drop-out from psychotherapeutic treatment are associated. We specifically hypothesized that higher GCMs over the course of treatment are associated with a lower risk for drop-out. As GCMs, we considered interpersonal experiences, intrapersonal experiences and problem actuation. This means that stronger interpersonal experiences (i.e. a better therapeutic alliance, more contentment and openness), stronger intrapersonal experiences (i.e. clarification: greater insight into crucial aspects of symptoms/ problems, and mastery: feeling well equipped to solve problems and to achieve personal treatment goals), and greater problem actuation (i.e. re-experiencing central problems, being emotionally involved/ activated) over the course of treatment are associated with a lower risk for drop-out. In addition, we hypothesized that changes in symptom severity and drop-out are related. More specifically, based on previous research, we expected stronger positive changes in symptom severity to be associated with a reduced risk for drop-out. We tested these hypotheses in a sample of 724 patients from routine care using survival analysis

to better account for the longitudinal nature of the data as well as time varying (GCMs, changes in symptom severity) and time constant (therapist and patient characteristics) predictors.

Method

Setting and Procedure

Patients in the current study received psychotherapeutic treatment in an outpatient clinic in Germany. The treatment conducted for all patients was cognitive behavioral therapy. Therapies were primarily delivered by graduate psychologists during their clinical training. A senior, licensed psychotherapist supervised the therapist every fourth treatment session to provide support and to ensure a high quality of treatment. In Germany, patients start treatment with five preparing sessions. The purpose of this phase is to assign diagnoses that are relevant for further treatment planning, to see if patient and therapist agree on working together and to secure funding from the patient's health insurance agency. The research ethics committee declared this study as exempt because it was based on routinely collected data from an outpatient clinic in Germany.

Sample

The original sample consisted of 1,772 patients who terminated treatment (either mutually agreed or unilaterally). However, a substantial number of patients had to be excluded due to missing data (i.e., missing treatment dates, unknown treatment termination status, missings in covariates) or inconsistent data (e.g. negative time intervals between treatment sessions, for example the second ROM-session date precedes the first ROM-session date). Patients who dropped out during the preparatory treatment phase also were excluded if GCM (general change mechanisms) measures were not provided. This was the case for 275 (15.52%) patients. The remaining sample consisted of 724 patients (40.86% of the original sample) treated by 137 therapists. Average patient age was $M = 38.86$ years ($SD = 12.97$,

median 39) and 56.08% patients were women. The average therapist age was $M = 31.05$ years ($SD = 7.29$, median 29) and 87.5% were women. Patients completed an average of $M = 47.07$ ($SD = 21.13$, median 48) treatment sessions in $M = 433.23$ days ($SD = 228.24$, median 412). In terms of diagnostic categories according to Roseborough et al. (2016), 65.75% of patients had a depressive disorder, 54.14% had an anxiety disorder (excluding obsessive-compulsive disorder), 7.73% had a personality disorder, and 6.08% had an obsessive-compulsive disorder. Patients had an average of $M = 1.95$ diagnoses ($SD = 0.99$, median 2). Patient and therapist demographics are displayed in Table 1. We focused on the same diagnostic categories as Roseborough et al. (2016) because the study was similar to our study.

Measures

Changes in symptom severity and GCMs were assessed using standardized self-report questionnaires.

Symptom severity. Symptom severity at pretreatment was measured using the revised Symptom Checklist (SCL-90R; Derogatis, 1977; German version by Franke, 2002). This questionnaire consists of 90 items rated on a 5-point Likert scale ranging from 0 (*not at all*) to 4 (*extremely*). Chronbach's alpha was 0.98 in this sample. In the ROM-sessions (routine outcome monitoring), we used a short version of the SCL-90R introduced by Harfst et al. (2002). It consists of 14 items rated on the same Likert scale as the SCL-90R. We estimated a Cronbach's alpha value of 0.86 based on the first ROM-session of this sample. The Global Severity Index (GSI) is the total mean score of all items and was used in both versions to measure symptom severity. Prinz et al. (2008) showed a high convergent validity between the GSI in the SCL-90R and the GSI in the SCL-14 of $r = 0.92$ in a sample of 8581 patients. High retest reliability of the GSI in the SCL-90R was reported by Horowitz, Rosenberg, Baer, Ureño and Villaseñor (1988) based on a ten-week waiting period ($r = 0.84$). Information on the retest-reliability of the GSI of the SCL14 was not available. As the

predictor, we used changes in symptom severity from one treatment monitoring session to the next (e.g. GSI second ROM-session -minus GSI first monitoring session).

GCMs. GCMs were measured using the Bern Post Session Reports (BPSR), Patient's and Therapist's version, developed by Flückiger, Grosse Holtforth, Znoj, Caspar and Wampold. (2010). The BPSR was developed to measure GCMs according to Grawe (1997) in a general treatment setting. Thus, it can be used for any kind of psychotherapy including cognitive behavioral therapy. The patient version consists of 22 items rated on a 6-point Likert scale ranging from -3 (*not at all*) to 3 (*yes, exactly*). The therapist version consists of 27 items, in which the first 12 items are rated on the same Likert scale as the patient version and the remaining items are rated on a 5-point Likert scale ranging from 0 (*completely disagree*) to 4 (*completely agree*). The BPSR contains a large number of subscales (eight subscales in the patient version and 11 subscales in the therapist version). Some of these subscales only consist of a single item. It has been shown that these subscales were not completely distinct and reliable (Flückiger et al., 2010). Therefore, Flückiger et al. (2013) introduced 'interpersonal experiences', 'intrapersonal experiences', and 'problem actuation' as secondary scales comprising different subscales of the original version. 'Interpersonal experiences' consists of the 'alliance' and 'contentment' scale of the patient version as well as the 'alliance' and 'openness' scale of the therapist version. Items are, for example, "the relationship with my patient/therapist felt comfortable today", or "I find my therapist's perception of my problems too simplistic". 'Intrapersonal experiences' consists of the scales 'mastery' and 'clarification' of both the patient and the therapist version with items like "I now feel more up to situations I didn't feel I could cope with up until now" and "I feel that I understand myself and my problems better". The problem actuation scale was adopted from the original version and an example item is, "today I was very involved emotionally". In the present study, Cronbach's alpha ranged from 0.74 (therapist-rated problem actuation) to 0.85

(therapist-rated interpersonal experience). Previous studies showed the predictive validity of these GCMs on treatment outcome (Flückiger et al., 2013; Gmeinwieser et al., 2019b).

Drop-out. Whether treatment termination was unilaterally or mutually agreed was documented by the therapists at the end of treatment.

Data collection and missing values

Data collection started at the first treatment preparing session. The first ROM-session was conducted after the initial treatment phase (session five) and after every tenth treatment session (10th session, 20th session, etc.). Whereas the pre-treatment assessment only involved symptom severity (SCL-90R) and other, mostly symptom-related measures, ROM also included GCMs. A list of measures is displayed in Table 2. All measures were routinely applied to all patients at the start or over the course of treatment. Therapists were introduced to and familiarized with ROM and they were encouraged to use and discuss these measures with their patients.

The starting point for the analyses was the first ROM-session. This was done for two reasons. First, in Germany the actual treatment in outpatient psychotherapy usually starts after the preparing phase at the sixth session (after the patient's health insurance agency agreed to pay for treatment). Other studies in German outpatient clinics also assumed that treatment only starts after the initial preparing phase (e.g., Schindler, Hiller & Witthöft, 2011; Schlagert & Hiller, 2016). Second, GCM-measures were first applied at the first ROM-session. Therefore, the first possible starting point for survival analysis with GCMs as time-varying variables was the first ROM-session.

It turned out that symptom severity at the first ROM-session was not collected for most patients (83% missings). For these cases, we used the pre-treatment symptom severity. To check whether this procedure introduced a bias (Lachin, 2016) we compared the distributions between the pre-treatment and the first ROM-session symptom severity visually

and with respect to different distributional measures (i.e. variance, skewness, kurtosis, quantiles). We found that both distributions were fairly equal. In addition, we performed a dependent t-test between pre-treatment symptom severity and the symptom severity at the first ROM-session for patients that had completed measures at both sessions. The t-test revealed a significant increase of symptom severity between pre-treatment and the first ROM-session ($t[116] = 2.19, p = 0.031$) with an effect size of $d = 0.202$, which is considered a small effect according to Cohen (1988). However, a post hoc power analysis revealed a power of 0.58 for a small effect. Furthermore, we compared pre-treatment symptom severity scores of patients, who had scores at the first ROM session, and patients, who did not. An independent t-test showed that there were no significant difference ($t[257.05] = 1.331, p = 0.184$). Given these findings, replacing missings at the first ROM-session by pre-treatment scores is unlikely to introduce a bias that leads to an overestimation of an effect.

No other missings were replaced because survival analysis utilizes the available data independent of the number sessions that were completed.

Statistical analyses

We applied survival analysis to identify factors that predict the risk of drop-out over the course of time. More specifically, we used Cox proportional hazard models which are well suited to analyze variables that might influence the risk of facing an event (i.e. drop-out). As covariates, we used patient and therapist sex as well as patient age. We also used diagnostic categories as covariates, i.e. personality disorder, depressive disorder, obsessive-compulsive disorder and anxiety disorders (excluding obsessive-compulsive disorder). We selected the covariates that were available in our sample based on the literature review by Roseborough et al. (2016) and the meta-analysis by Swift and Greenberg (2012). As predictors, we used changes in symptom severity and patient- and therapist-rated GCMs (interpersonal experiences, intrapersonal experiences and problem actuation). We used

information criteria (AIC, BIC) and a likelihood ratio test to compare a covariates-only model to a model containing covariates and predictors. Lower AIC and BIC values indicate a better model fit. Model fit of the model containing covariates and predictors was estimated using a test for proportional hazards in Cox regression models. The Cox regression model comes with the assumptions of proportional hazards (Collett, 2003). For this study, this means that the risk for drop-out and mutually agreed termination varies in time proportionally and at no point in time the cumulative hazard shifts from drop-out being more likely to drop-out being less likely than mutually agreed termination for variables like changes in symptom severity. We tested for violations against the assumption of proportional hazards (Grambsch & Therneau, 1994; Therneau & Grambsch, 2000). We also used concordance to evaluate the predictive ability of the two models. The concordance summarizes the agreement between observed and predicted survival times and ranges from 0 to 1 where 0.5 is obtained for a model with non-predictive ability and values greater than 0.7 are viewed as reasonably predictive (Harrell, 2015). Pseudo-R² and hazard ratios were used as effect sizes.

We used Cox proportional hazards models with the endpoint drop-out for our survival analysis. According to the model, patients are considered members of a risk set, that is, they are at any point in time to a certain degree likely to experience one of the two possible outcomes (drop-out or mutually agreed termination). Patients are no longer part of the risk set as soon as one of the two outcomes is reached. According to the model, time-varying predictors are considered constant as long as there is no further measurement of the same predictor. The influence of every predictor on the outcome is defined as the log hazard ratio. The log hazard ratio is comparable to the log odds ratio in a logistic regression model. The hazard ratio reflects the gain in risk for two fictional patients scoring equally on all predictors except the one where the hazard ratio is considered. In this predictor, these two patients have a one-point-difference on the respective scale. For example, if the hazard ratio interpersonal

experiences is 0.5, this means that with two fictional patients who score equally on all covariates and GCMs except for a one-point-difference in interpersonal experiences, the patient with the lower score is twice as likely to drop out during treatment.

Our data suggested a hierarchical structure. Measurements were nested within patients and patients were nested within therapists. Therefore, we tested if multilevel modelling was a better fit for our data. We compared three different null-models (i.e. without covariates or predictors) comparing different degrees of nesting: an intercept only cox-model, a two-level mixed effects cox-model with patients nested within therapists and a three-level mixed effects cox-model with measurements nested within patients and patients nested within therapists. We used likelihood ratio tests and information criteria (AIC; BIC) to compare the different models and continued our analysis with the best fitting model.

All analyses were performed using R (R Core Team, 2018) and the additional packages “psych” (Revelle, 2018) and “survival” (Therneau, 2015). Multilevel modelling was performed with the package “coxme” (Therneau, 2018). Figures were created using the packages “cowplot” (Wilke, 2017), “ggplot2” (Wickham, 2009), “gridExtra” (Auguie, 2017) and “survminer” (Kassambra & Kosinski, 2018).

Results

Sample characteristics

Means, medians, standard deviations and correlations between covariates and predictors can be viewed in Table 3. There were no concerning intercorrelations between covariates and/or predictors.

One-hundred-fifty-three patients dropped out during treatment, equaling 21.13% of the total sample. Six of these patients (0.82%) dropped out during the preparatory phase. As to be expected, treatment duration differed between patients who dropped out and patients who ended treatment mutually agreed. Mean treatment duration in days was $M = 306.52$ (SD

= 206.81, 265) for patients with drop-out, and $M = 467.19$ ($SD = 221.82$, median 440) for patients with mutually agreed termination. Mean number of treatment sessions was $M = 28.49$ ($SD = 14.33$, median 25) for patients with drop-out, and $M = 47.69$ ($SD = 15.78$, median 51) for patients with mutually agreed termination. Both differences were significant with p-values of $p < 0.001$ ($t[253.72] = 8.4$ and $t[167.88] = 11.38$, respectively).

Multilevel modelling

We compared an intercept only cox-model, a two-level mixed effects cox-model with patients nested within therapists and a three-level mixed effects cox-model with measurements nested within patients and patients nested within therapists using likelihood ratio tests and information criteria. The AIC and BIC delivered conflicting results with the AIC favoring the two-level model (1836.15 intercept only model, 1830.35 two-level model, 1830.26 three-level model) and the BIC favoring the intercept only model (1836.15 intercept only model, 1896.99 two-level model, 1897.56 three-level model). The likelihood ratio tests revealed no significant differences between the different null-models ($\chi^2[1] = 3.01$, $p = 0.083$ intercept only model vs. two-level model; $\chi^2[2] = 3.05$, $p = 0.218$ intercept only model vs. three-level model; $\chi^2[1] = 0.04$, $p = 0.841$ two-level model vs. three-level model). Based on these results we continued our analyses with a fixed effects model.

Model comparisons

The likelihood ratio test revealed a significant difference between the covariates-only model and the model containing both covariates and predictors ($\chi^2[7] = 82.885$, $p < 0.001$). Information criteria showed an AIC of 1,822 and BIC of 1,843 for the covariates-only model and an AIC of 1,753 a BIC of 1,795 for the model containing both covariates and predictors. These results were further supported by a better fit for the model containing both covariates and predictors. The covariates-only model showed a concordance of $c = 0.623$, the model containing covariates and predictors showed a concordance of $c = 0.738$. Based on the model

analyses we continued with the model containing both covariates and predictors. The test for proportional hazards neither showed significant results for the overall model ($p = 0.69$) nor for any single covariate or predictor (lowest $p = 0.075$ for changes in symptom severity, highest $p = 0.995$ for therapist-rated interpersonal experiences). Thus, the model did not violate the assumption of proportional hazards and fitted the data adequately.

Survival analysis

Detailed results of the model are displayed in Table 4. The pseudo- R^2 of the model was $R^2 = 0.038$ (compared to $R^2 = 0.01$ of the covariates-only model). The only significant covariate was patient age ($HR = 0.981$, $p = 0.007$), which indicates that older patients were less likely to end treatment unilaterally. Significant predictors were patient- and therapist-rated interpersonal experiences. Patient-rated interpersonal experiences showed a hazard ratio of $HR = 0.746$ ($p = 0.022$). Thus, patients with higher interpersonal experiences had a lower risk to drop out. Therapist-rated interpersonal experiences showed a hazard ratio of $HR = 0.589$ with a p-value of $p < 0.001$. This hazard ratio is interpreted in the same way as the hazard ratio for patient-rated interpersonal experiences. Symptom severity change as well as patient- and therapist-rated intrapersonal experiences and problem actuation showed no significant results. Figure 1 (A-D) provides an overview of the estimated hazard ratios for different values of the predictors. More precisely, the lines show how the hazard ratio changes when the values of the predictor changes. For example, Panel A shows that a -2 -point difference in therapist-rated interpersonal experiences increases the risk for drop-out by a factor of $HR = 3$ for the patient with the lower therapist-rated interpersonal experiences.

Discussion

This is the first study that tested whether a range of GCMs (general change mechanisms, i.e. interpersonal experiences, intrapersonal experiences, problem actuation) and changes in symptom severity are associated with drop-out from outpatient psychotherapy

using survival analysis. In our analysis, the model containing GCMs and symptom severity change showed a better fit for the data than a model containing only covariates (diagnostic categories and patient characteristics). Except for patient age (older patients had a lower risk for drop-out), none of the covariates were associated with drop-out.

Our hypotheses regarding GCMs severity were partially supported by our results. In line with our first hypothesis, greater positive patient- and therapist-rated interpersonal experiences (i.e. alliance, contentment, openness) were associated with a reduced risk for drop-out. Counter to our hypotheses, symptom severity change, intrapersonal experiences (i.e. clarification and mastery), and problem actuation were not associated with risk for drop-out.

The results are partially in accordance with previous research. The patient-therapist alliance-outcome relationship has been heavily studied in recent years. A meta-analysis including 295 studies by Flückiger, Del Re, Wampold, Adam and Horvath (2018) showed an average effect size of $r = 0.278$ of alliance on outcome, which is a small but robust effect. Alliance and drop-out had an average correlation of $r = 0.18$ in 27 of the reviewed studies, which is also considered small according to Cohen (1988). Sharf et al. (2010) as well as Roos and Werbart (2013) also reported a negative relation between alliance and drop out in their reviews (i.e. the risk for drop-out is lower for clients with a stronger therapeutic alliance). Roos and Werbart (2013) also identified the patient's contentment as a predictor of dropout. Interpersonal experiences in our analysis comprises alliance and contentment in the patient version and alliance and openness in the therapist version. The respective hazard ratios of $HR = 0.746$ and $HR = 0.589$ are also considered small effect sizes according to Borenstein, Hedges, Higgins and Rothstein (2009). The greater effect size of therapist-rated compared to patient-rated interpersonal experiences can be explained in different ways. On the one hand, therapists might in fact be more reliable to indicate when a patient might drop out from their

interpersonal experiences. On the other hand, therapists with higher interpersonal experiences might be less likely to classify ambiguous patient termination cases as drop-out. In addition, since therapists rated the interpersonal experiences and classified patients as drop-out cases the greater effect size might also be a result from common source variance.

We did not find a significant association between intrapersonal experiences and problem actuation, and drop-out. For problem actuation, this is in line with most research considering GCMs (Flückiger et al., 2013; Gmeinwieser et al., 2019b; Mander et al. 2015). By contrast, intrapersonal experiences have been previously shown to predict treatment outcome (Gmeinwieser et al., 2019a; Grosse Holtforth, Grawe, Fries & Znoj, 2008) and drop-out (Kegel & Flückiger, 2015). In terms of the dual model of psychotherapy, this could mean that interpersonal experiences are the more relevant of the two processes. Interpersonal experiences might also provide the ground on which more demanding interventions such as exposure can be performed effectively. Thus, mastery and clarification (i.e. intrapersonal experiences) as well as problem actuation might only be stronger associated with outcomes when based on solid interpersonal experiences. As can be seen in Table 3, there is less agreement between therapists and patients in terms of their ratings of intrapersonal experiences and problem actuation than in terms of their ratings of interpersonal experiences. Therapists rate intrapersonal experiences and problem actuation much higher than patients do. It might be helpful for patients and therapists to more openly discuss intrapersonal experiences and problem actuation.

Differences in findings may result from the nature of our analysis and our outcome. Most studies did not investigate the relationship between intrapersonal experiences and drop-out (Flückiger et al., 2013; Probst et al., 2018) or they did not investigate this relationship through survival analysis (i.e. Flückiger et al., 2011; Kegel & Flückiger, 2015). There are further differences between the studies that investigated the association between GCMs and

drop-out. Flückiger et al. (2011) investigated the association in a sample of 78 patients with binge eating disorder in 16 weekly group therapy sessions. The sample in the current study consisted of 724 patients from routine care with various diagnoses in an individual treatment setting. Kegel and Flückiger (2015) also examined patients from routine care (N = 296). Patients were categorized as drop-out if they did not complete the post-treatment measures. According to this definition, the drop-out rate in Kegel and Flückiger (2015) was 57.1%. However, the authors reported that, “little effort was put into increasing return rates” (p. 379). Thus, patients might more easily be categorized as drop-outs even if treatment was otherwise completed properly. The sample in this study was with N = 724 more than twice as large and therapists rated whether patients were regarded as drop-out. The drop-out rate was 21.13% and thus much lower than the drop-out rate reported in Kegel and Flückiger (2015). This could be due to the different drop-out definitions in the two studies.

Contrary to previous research, we did not find a significant relationship between symptom severity change and drop-out. Change of symptom severity or sudden gains have been shown to predict outcome in a variety of settings and disorders (Aderka, Nickerson, Bøe & Hofmann, 2012; Stiles et al, 2003; Tang, Luborsky & Andrusyna, 2002, Xiao et al., 2017). However, as Roos and Werbart (2013) already pointed out, the relation between changes in symptom severity during treatment, and drop out may be complex. Patients may drop out because they improved substantially from their subjective point of view or they may leave the treatment because they did not improve or even got worse. It could be that in the current study, two subgroups of patients (i.e. those who improved substantially and dropped-out and those who did not improve and dropped-out) were lumped together and that this resulted in a null finding regarding the association between symptom severity change and drop-out. Another explanation might also be that interpersonal experiences is the stronger predictor of drop-out compared to changes in symptom severity. Sharf and Primavera (2010) found lower

alliance more strongly associated with drop-out in outpatient treatment settings and longer treatments (> 16 treatment sessions). Since we used interpersonal experiences, which comprises the therapeutic alliance, and in our sample, the treatments are predominantly longer than 16 sessions this could be an explanation for our finding. Although the therapeutic relationship is commonly less emphasized in CBT than in other treatment approaches, the current findings suggest that also CBT therapists can benefit from carefully monitoring the therapeutic relationship in order to identify risk of drop-out.

Limitations

There are a several limitations in our study that need to be considered when interpreting the results. First, we did not assess between-session experiences and thus cannot draw any conclusions as to which degree the risk for drop-out was affected by other factors such as life events. Second, data was drawn from outpatient cognitive behavioral therapy, delivered by psychologists in their clinical training, in a routine care setting in Germany. As, for instance, treatment duration is often much longer in Germany than in the U.S. or UK, generalizability of the results to other settings is limited.

Third, as no follow-up data were available for the patients in this study, we cannot draw any conclusions whether drop-out was actually detrimental to the patient and/ or therapist. While there is evidence that dropping-out might result in less or a delayed improvement (Cahill et al., 2003; Lopes, Gonçalves, Sinai & Machado, 2017), there is also counter-evidence that long-term treatment outcomes are at least mixed for patients who dropped out during treatment (Knox et al., 2011). Fourth, we used the GSI of the SCL-90R or SCL-14 as a global measure to calculate symptom severity change. We did not use disorder-specific measures like the GAD-7 (Spitzer, Kroenke, Williams & Löwe, 2006) or the Beck Depression Inventory (Beck, Ward, Mendelson, Mock & Erbaugh, 1961) which could be better suited for specific diagnoses like anxiety or depressive disorders.

Fifth, ROM-sessions (routine outcome monitoring) were carried out every tenth session. With usually weekly treatment sessions, this would lead to a delay of at least 2 ½ months between ROM-sessions. This poses a problem, as processes happening between ROM-sessions could not be tracked. Furthermore, survival analysis models consider time-varying variables like symptom severity change as constant between measurements, which is a limitation. Future studies should apply measures more often in treatment to get a closer look into the fluctuation of the different time-dependent variables.

Sixth, patients who dropped out during the preparatory phase of treatment (i.e. before the first ROM-session) were mostly excluded, because a majority did not provide GCM-measures. Therefore, these early drop-outs are underrepresented in our sample, which is a limitation. In consequence, results should be interpreted cautiously concerning early drop-out in psychotherapy.

Seventh, although patients, who had missing symptom severity measures at the first ROM-session, did not differ significantly in their symptom severity at pre-treatment from patients, who had both measurements, it was shown that patients with both measurements worsened significantly from pre-treatment to the first ROM-session. Replacing missing values in symptom severity at the first ROM-session with pre-treatment symptom severity might therefore have been a too conservative approach when patients show an increase in symptom severity on average.

Eighth, the sample of 724 patients is only 40.86% of the original sample of 1772 patients. Unfortunately, it was not possible to include more patients in our analyses because most of the excluded patients had missing or inconsistent treatment dates or missing measurements. Treatment dates are a fundamental requirement for performing survival analysis. Treatment dates are generally important for comparing drop-out and mutually agreed termination because drop-out usually has a longer treatment duration and cannot,

therefore, be viewed as the opposite pole of mutually agreed termination since it might be associated with different predictors.

Ninth, the factorial validity of the GCM subscales interpersonal experiences, intrapersonal experiences and problem actuation has yet to be demonstrated (but see Flückiger et al., 2013). In addition, it would be important to compare the factorial fit of these secondary subscales to the original subscales as proposed by Flückiger et al. (2010).

Tenth, we could not provide further demographic information of our sample (e.g. education, ethnicity, marital status) because it was not routinely collected in our sample.

Implications and future research

Interpersonal experiences from the patient's and therapist's perspective turned out to be an significant predictor for drop-out. Assessing this factor continuously over the course of treatment using standardized assessment tools might thus help to detect a high risk of the patient to drop-out and to take action. Especially the therapists' assessment of interpersonal experiences seems to be crucial. If the therapist perceives below average interpersonal experiences, there is an increased risk for the patient ending the treatment unilaterally.

Since specific reasons for drop-out might be important in assessing if drop-out is potentially harmful to patients, future research is clearly needed. Building upon this study, it would be helpful to investigate the relationship of these and other GCMs, changes in symptom severity and different reasons for drop-out. In addition, it would be important to know about the short- and long-term outcomes of the patients dropping out. This would shed further light on the processes that lead to drop-out and could potentially help to develop strategies to prevent drop-out negatively impacting patients.

References

- Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: a meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(1), 93–101. <https://doi.org/10.1037/a0026455>
- Altmann, U., Thielemann, D., Zimmermann, A., Steffanowski, A., Bruckmeier, E., Pfaffinger, I., & Strauß, B. (2018). Outpatient psychotherapy improves symptoms and reduces health care costs in regularly and prematurely terminated therapies. *Frontiers in Psychology, 9*, 748. <https://doi.org/10.3389/fpsyg.2018.00748>
- Auguie, B. (2017). gridExtra: Miscellaneous Functions for “Grid” Graphics. Retrieved from <https://CRAN.R-project.org/packages=gridExtra> Version = 2.3
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. (1961).). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571. <https://doi.org/10.1001/archpsyc.1961.01710120031004>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis* (Reprinted.). Chichester: Wiley.
- Cahill, J., Barkham, M., Hardy, G., Rees, A., Shapiro, D. A., Stiles, W. B., & Macaskill, N. (2003). Outcome of patients completing and not completing cognitive therapy for depression. *British Journal of Clinical Psychology, 42*, 133–143.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2.ed.). Hillsdale: Lawrence Erlbaum Associates.
- Collett, D. (2003). *Modelling survival data in medical research* (2.ed.). *Texts in statistical science*. Boca Raton, Fla.: Chapman & Hall/CRC.
- Corning, A. F., & Malofeeva, E. V. (2004). The Application of Survival Analysis to the Study of Psychotherapy Termination. *Journal of Counseling Psychology, 51*(3), 354–367. <https://doi.org/10.1037/0022-0167.51.3.354>

- Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., & Gilbody, S. (2014). Early changes, attrition, and dose-response in low intensity psychological interventions. *The British Journal of Clinical Psychology*, *53*(1), 114–130. <https://doi.org/10.1111/bjc.12031>
- Derogatis, L. R. (1977). *SCL-90-R, administration, scoring & procedures manual-I for the R(evised) version*. Johns Hopkins University of Medicine.
- Flückiger, C., Del Re, 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>
- Flückiger, C., Grosse Holtforth, M., Znoj, H. J., Caspar, F., & Wampold, B. E. (2013). Is the relation between early post-session reports and treatment outcome an epiphenomenon of intake distress and early response? A multi-predictor analysis in outpatient psychotherapy. *Psychotherapy Research*, *23*(1), 1–13. <https://doi.org/10.1080/10503307.2012.693773>
- Flückiger, C., Meyer, A., Wampold, B. E., Gassmann, D., Messerli-Bürge, N., & Munsch, S. (2011). Predicting Premature Termination Within a Randomized Controlled Trial for Binge-Eating Patients. *Behavior therapy*, *42*(4), 716–725.
- Flückiger, C., Regli, D., Zwahlen, D., Hostettler, S., & Caspar, F. (2010). Der Berner Patienten- und Therapeutenstundenbogen 2000. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *39*(2), 71–79. <https://doi.org/10.1026/1616-3443/a000015>
- Franke, G. H. (2002). *SCL-90-R: Symptom Checkliste von L.R. Derogatis (deutsche Version)*. Göttingen: Beltz.
- Gmeinwieser, S., Hagmayer, Y., Pieh, C., & Probst, T. (2019a). General change mechanisms in the early treatment phase and their associations with the outcome of cognitive behavioural therapy in patients with different levels of motivational incongruence.

- Clinical Psychology & Psychotherapy*, 26(5), 550–561.
<https://doi.org/10.1002/cpp.2381>
- Gmeinwieser, S., Kuhlencord, M., Ruhl, U., Hagmayer, Y., & Probst, T. (2019b). Early developments in general change mechanisms predict reliable improvement in addition to early symptom trajectories in cognitive behavioral therapy. *Psychotherapy Research*. Advance online publication.
<https://doi.org/10.1080/10503307.2019.1609709>
- Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3), 515–526.
<https://doi.org/10.1093/biomet/81.3.515>
- Grawe, K. (1997). Research-Informed Psychotherapy. *Psychotherapy Research*, 7(1), 1–19.
<https://doi.org/10.1080/10503309712331331843>
- Grosse Holtforth, M., Grawe, K., Fries, A., & Znoj, H. (2008). Inconsistency as a criterion for differential indication in psychotherapy. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 37(2), 103–111. <https://doi.org/10.1026/1616-3443.37.2.103>
- Harfst, T., Koch, U., Kurtz von Aschoff, C., Nutzinger, D. O., Rüdell, H., & Schulz, H. (2002). Entwicklung und Validierung einer Kurzform der Symptom Checklist-90-R. *DRV-Schriften*, 33, 71-73.
- Harrell, J. F. E. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer Series in Statistics. New York: Springer. <https://doi.org/10.1007/978-3-319-19425-7>
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villaseñor, V. S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, 56(6), 885–892.
<https://doi.org/10.1037/0022-006X.56.6.885>

- Kassambara, A. & Kosinski, M. (2018). survminer: Drawing Survival Curves using 'ggplot2'. Retrieved from <https://CRAN.R-project.org/package=survminer> Version = 2.3
- Kazdin, A. E. (1996). Dropping Out of Child Psychotherapy: Issues for Research and Implications for Practice. *Clinical Child Psychology and Psychiatry*, 1(1), 133–156.
- 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>
- Knox, S., Adrians, N., Everson, E., Hess, S., Hill, C., & Crook-Lyon, R. (2011). Clients' perspectives on therapy termination. *Psychotherapy Research*, 21(2), 154–167. <https://doi.org/10.1080/10503307.2010.534509>
- Lachin, J. M. (2016). Fallacies of last observation carried forward analyses. *Clinical Trials*, 13(2), 161–168. <https://doi.org/10.1177/1740774515602688>.
- Lopes, R. T., Gonçalves, M. M., Sinai, D., & Machado, P. P. (2017). Clinical outcomes of psychotherapy dropouts: Does dropping out of psychotherapy necessarily mean failure? *Revista Brasileira De Psiquiatria*, 40(2), 123–127. <https://doi.org/10.1590/1516-4446-2017-2267>
- Mander, J., Schlarb, A., Teufel, M., Keller, F., Hautzinger, M., Zipfel, S., . . . Sammet, I. (2015). The individual therapy process questionnaire: Development and validation of a revised measure to evaluate general change mechanisms in psychotherapy. *Clinical Psychology & Psychotherapy*, 22, 328–345. <https://doi.org/10.1002/cpp.1892>
- Piselli, A., Halgin, R. P., & Macewan, G. H. (2011). What went wrong? Therapists' reflections on their role in premature termination. *Psychotherapy Research*, 21(4), 400–415. <https://doi.org/10.1080/10503307.2011.573819>

- Prinz, U., Nutzinger, D., Schulz, H., Petermann, F., Braukhaus, C., & Andreas, S. (2008). The Symptom-Check-List-90-R (SCL-90-R) and the Short Versions of the SCL-90-R: Psychometric Analyses of Inpatients with Mental Disorders [Die Symptom-Checkliste-90-R und ihre Kurzversionen: Psychometrische Analysen bei Patienten mit psychischen Erkrankungen]. *Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin*, 18(06), 337–343. <https://doi.org/10.1055/s-0028-1093323>
- Probst, T., Jakob, M., Kaufmann, Y. M., Müller-Neng, J. M. B., Bohus, M., & Weck, F. (2018). Patients' and therapists' experiences of general change mechanisms during bug-in-the-eye and delayed video-based supervised cognitive-behavioral therapy. A randomized controlled trial. *Journal of Clinical Psychology*, 74(4), 509–522. <https://doi.org/10.1002/jclp.22519>
- 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/> Version = 3.5.2
- Revelle, W. (2018). psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston, Illinois, USA. Retrieved from <https://CRAN.R-project.org/package=psych> Version = 1.8.4
- Roos, J., & Werbart, A. (2013). Therapist and relationship factors influencing dropout from individual psychotherapy: A literature review. *Psychotherapy Research*, 23, 394–418. <https://doi.org/10.1080/10503307.2013.775528>
- Roseborough, D. J., McLeod, J. T., & Wright, F. I. (2016). Attrition in Psychotherapy. *Research on Social Work Practice*, 26(7), 803–815. <https://doi.org/10.1177/1049731515569073>
- Schindler, A. C., Hiller, W., & Witthöft, M. (2011). Benchmarking of cognitive-behavioral therapy for depression in efficacy and effectiveness studies--how do exclusion criteria

- affect treatment outcome? *Psychotherapy Research*, 21(6), 644–657.
<https://doi.org/10.1080/10503307.2011.602750>
- Schlagert, H. S., & Hiller, W. (2016). The predictive value of early response in patients with depressive disorders. *Psychotherapy Research*, 27(4), 488–500.
<https://doi.org/10.1080/10503307.2015.1119329>
- Schulte, D. (1996). *Therapieplanung*. Göttingen: Hogrefe.
- Schulte, D., & Eifert, G. H. (2002). What to Do When Manuals Fail? The Dual Model of Psychotherapy. *Clinical Psychology: Science and Practice*, 9(3), 312–328.
<https://doi.org/10.1093/clipsy.9.3.312>
- Sharf, J., Primavera, L.H., & Diener, M. J. (2010). Dropout and therapeutic alliance: a meta-analysis of adult individual psychotherapy. *Psychotherapy Theory, Research, Practice, Training*, 47(4), 637–645. <https://doi.org/10.1037/a0021175>
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder. *Archives of Internal Medicine*, 166, 1092–1097.
- 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>
- Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: a meta-analysis. *Journal of Consulting and Clinical Psychology*, 80(4), 547–559.
<https://doi.org/10.1037/a0028226>
- Tang, T. Z., Luborsky, L., & Andrusyna, T. (2002). Sudden gains in recovering from depression: Are they also found in psychotherapies other than cognitive-behavioral

- therapy? *Journal of Consulting and Clinical Psychology*, 70(2), 444–447.
<https://doi.org/10.1037//0022-006X.70.2.444>
- Therneau, T. (2015). *_A Package for Survival Analysis in S_*. Retrieved from
<https://CRAN.R-project.org/package=survival> Version = 2.3.8
- Therneau, T. (2018). Coxme: Mixed Effects Cox Models. Retrieved from <https://CRAN.R-project.org/package=coxme> Version 2.2-10
- Therneau, T. M., & Grambsch, P. M. (2000). Testing Proportional Hazards. In K. Dietz, M. Gail, K. Krickeberg, J. Samet, A. Tsiatis, T. M. Therneau, & P. M. Grambsch (Eds.), *Statistics for Biology and Health. Modeling Survival Data: Extending the Cox Model* (pp. 127–152). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4757-3294-8_6
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Wilke, C. O. (2017). cowplot: Streamlined Plot Theme and Plot Annotations for ‘ggplot2’. Retrieved from <https://CRAN.R-project.org/packages=cowplot> Version = 0.9.2
- Xiao, H., Castonguay, L. G., Janis, R. A., Youn, S. J., Hayes, J. A., & Locke, B. D. (2017). Therapist effects on dropout from a college counseling center practice research network. *Journal of Counseling Psychology*, 64(4), 424–431.
<https://doi.org/10.1037/cou0000208>
- Zimmermann, D., Rubel, J., Page, A. C., & Lutz, W. (2017). Therapist Effects on and Predictors of Non-Consensual Dropout in Psychotherapy. *Clinical Psychology & Psychotherapy*, 24(2), 312–321. <https://doi.org/10.1002/cpp.2022>

Table 1: Patients and therapists demographics with means, standard deviations in parentheses and medians

Variable	<i>M</i> or percent	<i>SD</i>	Median
Sample size	721		
Patient age	38.86	12.97	39
Patient sex	56.08% women		
Number of therapists	137		
Therapist age	31.05	7.29	29
Therapist sex	87.5% women		
Patients per therapist	5.28	3.51	5
Number of sessions	47.07	21.13	48
Days of treatment	433.23	228.24	412
Number of diagnoses	1.95	0.99	2
Personality disorder	7.73%		
Depressive disorder	65.74%		
Obsessive-compulsive disorder	6.08%		
Anxiety disorder (excluding OCD)	54.14%		

Table 2: Applied measures at pre-treatment and ROM.

Measure	Scope	Reference
Bern Post Session Reports (ROM)	General Change Mechanisms	Flückiger et al. (2010)
Patient demographics (Pre)	Demographics	-
Symptom Checklist-14 (ROM)	Symptom Severity	Harfst et al. (2002)
Symptom Checklist-90R (Pre)	Symptom Severity	Franke (2002)

Note: Pre: applied at the first treatment session; ROM: applied in routine outcome monitoring sessions (sessions 5, 10, 20, 30, etc.).

Table 3: Means, standard deviations and correlations between covariates and predictors.

Variables	<i>M (SD)</i>	Median	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) Patient sex	-		1.00													
(2) Patient age	38.86 (12.97)	39	0.01	1.00												
(3) Therapist sex	-		0.07	0.02	1.00											
(4) Personality disorder	-		0.05	-0.03	-0.05	1.00										
(5) Depressive disorder	-		0.02	0.15	0.02	-0.01	1.00									
(6) Obsessive-compulsive disorder	-		0.02	-0.06	-0.01	0.03	-0.04	1.00								
(7) Anxiety disorder (excluding OCD)	-		0.10	-0.16	0.05	-0.13	-0.25	-0.06	1.00							
(8) Symptom severity ^a	0.92 (0.56)	0.84	0.02	0.11	0.06	0.08	0.23	0.10	0.13	1.00						
Patient-rated																
(9) Interpersonal experiences ^a	1.81 (0.68)	1.86	0.18	0.00	0.09	-0.09	-0.03	-0.06	0.08	-0.16	1.00					
(10) Intrapersonal experiences ^a	0.92 (0.75)	1	0.07	0.02	0.04	-0.05	-0.12	-0.07	0.01	-0.24	0.51	1.00				
(11) Problem actuation ^a	0.71 (1.02)	0.75	0.08	0.24	0.08	-0.03	0.05	0.03	0.02	0.18	0.27	0.32	1.00			
Therapist-rated																
(12) Interpersonal experiences ^a	1.45 (0.74)	1.51	0.20	0.02	0.03	-0.15	0.04	-0.03	0.05	-0.07	0.50	0.38	0.18	1.00		
(13) Intrapersonal experiences ^a	2.52 (0.54)	2.53	0.10	-0.05	0.03	-0.02	0.02	-0.01	0.09	-0.03	0.15	0.22	0.03	0.27	1.00	
(14) Problem actuation ^a	2.04 (0.63)	2	0.08	-0.02	0.00	0.01	0.03	0.03	0.10	0.02	0.14	0.12	0.19	0.10	0.48	1.00

^ameans and standard deviations of the covariates and predictors that vary over time were calculated by first estimating an individual mean per patient over all available routine outcome monitoring sessions and then the overall mean as an average of the individual means. The standard deviations are based on the overall mean.

Table 4: Results of the survival analysis model containing covariates and predictors.

Variable	Perspective	Coefficients	HR	95% CI	z	p-value
Covariates						
Patient sex		-0.234	0.791	0.571-1.098	-1.402	0.161
Patient age		-0.021	0.980	0.966-0.993	-2.887	0.004**
Therapist sex		0.324	1.382	0.839-2.275	1.272	0.203
Personality disorder		0.120	1.127	0.671-1.895	0.452	0.651
Depressive disorder		-0.245	0.783	0.547-1.121	-1.337	0.181
Obsessive-compulsive disorder		-0.166	0.847	0.443-1.619	-0.503	0.615
Anxiety disorder (excluding OCD)		-0.296	0.743	0.528-1.047	-1.699	0.089
Predictors						
Symptom severity		0.128	1.136	0.826-1.563	0.784	0.433
Interpersonal experiences	Patient-rated	-0.296	0.744	0.578-0.956	-2.307	0.021*
	Therapist-rated	-0.530	0.588	0.482-0.718	-5.233	< 0.001***
Intrapersonal experiences	Patient-rated	-0.157	0.854	0.696-1.048	-1.507	0.132
	Therapist-rated	0.046	1.047	0.800-1.371	0.337	0.736
Problem actuation	Patient-rated	-0.126	0.882	0.772-1.007	-1.854	0.064
	Therapist-rated	-0.109	0.896	0.733-1.096	-1.068	0.286

Note: HR = Hazard ratio, CI = confidence interval, OCD = obsessive-compulsive disorder, * p < 0.05, ** p < 0.01, *** p < 0.001

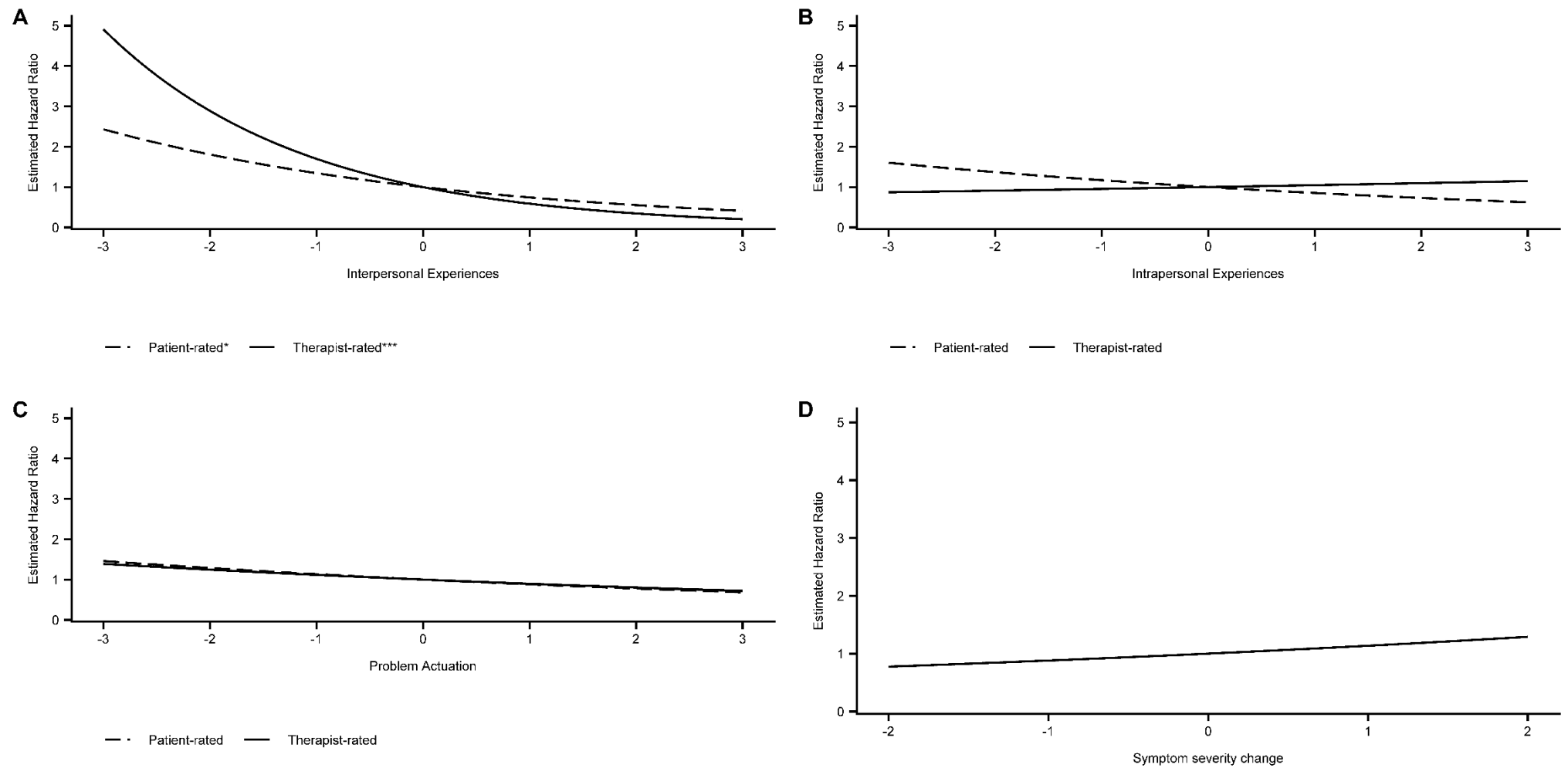


Figure 1. Estimated hazard ratios for the different predictors in the model containing covariates and predictors. A: interpersonal experiences, B: intrapersonal experiences, C: problem actuation, D: symptom severity change. Lines correspond to estimates from the statistical model and do not represent empirical values. The x-axis is interpreted as the difference between two fictional patients who score the same on all other covariates and predictors but have a difference according to the x-axis of the displayed variable. The zero on the x-axis means the (fictional) patients score the same on all covariates and predictors and, therefore, do not differ in estimated hazard ratios (hazard ratio of 1). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Diskussion und Ausblick

Der folgende Abschnitt enthält für jede Publikation eine kurze Zusammenfassung der Ergebnisse. Am Ende des Abschnitts werden die Publikationen dann in Zusammenhang zueinander gesetzt und einer gemeinsamen Diskussion unterzogen.

Gmeinwieser et al. (2019a)

Ziel der Studie war es, den Einfluss von frühen Veränderungen in GCMs auf das Therapieergebnis zu untersuchen. Als Outcomes dienten kontinuierliche und dichotome Symptombelastung, sowie die kontinuierliche und dichotome Lebenszufriedenheit. Die Ergebnisse der Publikation von Gmeinwieser et al. (2019a) sind in Tabelle 3 zusammengefasst. Die Zellen von Tabelle 3 enthalten nur die signifikanten Prädiktoren der jeweiligen Analysen. Tabelle 3 stellt im Vier-Felder-Format kontinuierliche und dichotome Prädiktoren gegenüber. Dadurch wird deutlich, welche Gemeinsamkeiten zwischen kontinuierlichen und dichotomen Outcomes bestehen. Der Rand von Tabelle 3 enthält die signifikanten Prädiktoren unabhängig vom jeweils anderen Outcome. Die dichotome Symptombelastung der ersten Spalte von Tabelle 3 hat z.B. den Prädiktor Pat. Intra. X Sitz mit der kontinuierlichen Symptombelastung gemeinsam. In der letzten Zeile der dichotomen Symptombelastung sind dagegen die Prädiktoren Ther. Inter., Ther. Intra. und Pat. Intra. X Sitz. aufgeführt. Diese Prädiktoren sind in der Analyse der dichotomen Symptombelastung signifikant. Da die Prädiktoren Ther. Inter. und Ther. Intra. nicht bei der Analyse der kontinuierlichen Symptombelastung signifikant geworden sind, sind diese nicht in der oberen rechten Zeile als Gemeinsamkeit zwischen kontinuierlicher und dichotomer Symptombelastung aufgeführt.

Tabelle 3. Signifikante Prädiktoren aus Gmeinwieser et al. (2019a).

Outcome	Dichotom		(kontinuierlich allein)
	Symptombelastung	Lebenszufriedenheit	
Kontinuierlich	Symptombelastung	Pat. Intra. X Sitz.	Pat. Inter.
			Ther. Inter. X Sitz.
			Pat. Inter. X Sitz.
			Pat. Intra. X Sitz.
(dichotom allein)	Lebenszufriedenheit		Pat. Inter.
			Pat. Intra.
			Ther. Inter. X Sitz.
		Ther. Inter. X Sitz.	
	Ther. Inter.	Pat. Inter.	
	Ther. Intra.	Pat. Intra.	
	Pat. Intra. X Sitz.	Ther. Inter. X Sitz.	

Notiz. Pat.: Patienteneinschätzung; Ther.: Therapeuteneinschätzung; Inter.: frühe Veränderungen der interpersonale Erfahrungen; Intra.: frühe Veränderungen der intrapersonalen Erfahrungen; X Sitz.: Interaktion mit der Anzahl der Therapiesitzungen

Insgesamt fällt auf, dass die Patientenperspektive stärker unter den signifikanten Prädiktoren vertreten ist als die Therapeutenperspektive. Die Patienteneinschätzung der frühen Veränderungen der interpersonalen Erfahrungen ist etwa ein signifikanter Prädiktor der kontinuierlichen Symptombelastung und der kontinuierlichen und dichotomen Lebenszufriedenheit. Dagegen sind die frühen Veränderungen der interpersonalen Erfahrungen aus Therapeutenperspektive lediglich ein signifikanter Prädiktor für die

dichotome Symptombelastung. Des Weiteren unterscheiden sich die Ergebnisse auch hinsichtlich der verschiedenen Outcome-Maße. Die Lebenszufriedenheit als dichotomes und kontinuierliches Maß zeigt dabei stärkere Gemeinsamkeiten im Vergleich zur Symptombelastung. Frühe Veränderungen der interpersonalen und intrapersonalen Erfahrungen aus Patientensicht, sowie die Interaktion der frühen Veränderungen der interpersonalen Erfahrungen mit der Anzahl Sitzungen aus Therapeutenperspektive sind sowohl bei der kontinuierlichen als auch bei der dichotomen Lebenszufriedenheit signifikante Prädiktoren. Die einzige Gemeinsamkeit der kontinuierlichen und dichotomen Symptombelastung ist dagegen die Interaktion der frühen Veränderungen der intrapersonalen Erfahrungen mit der Anzahl Sitzungen aus Patientenperspektive. Überschneidungen der verschiedenen Maße und Ergebnisformen werden ebenfalls in Tabelle 3 ersichtlich. So zeigt die dichotome Symptombelastung keine gemeinsamen signifikanten Prädiktoren mit der kontinuierlichen Lebenszufriedenheit. Umgekehrt zeigen sich Überschneidungen beispielsweise bei den frühen Veränderungen der patienteneingeschätzten interpersonalen Erfahrungen. Die unterschiedlichen Ergebnisse von Lebenszufriedenheit und Symptombelastung zeigen, dass die beiden Konstrukte keine entgegengesetzten Pole derselben Skala sind, sondern eigene Konstrukte mit teilweise eigenen signifikanten Prädiktoren. Zuletzt fällt auf, dass die Problemaktualisierung generell kein signifikanter Prädiktor ist. Prinzipiell zeigen die Prädiktoren dabei den erwarteten Zusammenhang, dass eine stärkere frühe Veränderung in GCMs mit einem besseren Therapieergebnis assoziiert ist.

Unterschiede in den signifikanten Prädiktoren zwischen der dichotomen und der kontinuierlichen Erfassung des Therapieergebnisses können dabei verschieden ausgelegt werden. Die therapeuteneingeschätzten frühen Veränderungen der intrapersonalen Erfahrungen zeigen sich etwa bei der dichotomen Symptombelastung als signifikanter Prädiktor, aber nicht bei der kontinuierlichen Symptombelastung. Klinisch praktisch könnte das bedeuten, dass gerade bei der leichter interpretierbaren Darstellung von

(dichotomen/kategorialen) Ergebnissen die frühen Veränderungen der intrapersonalen Erfahrungen aus Therapeutensicht einen wertvollen Beitrag zur Einschätzung des späteren Therapieergebnisses leisten kann im Vergleich zu kontinuierlich betrachteten Ergebnissen. Die Wahrscheinlichkeit einer Verbesserung durch die Therapie könnte auch leichter für Patienten zu erklären und zu interpretieren sein als das reine kontinuierliche Therapieergebnis. Dieser Befund kann allerdings auch statistisch/methodisch betrachtet werden. Da durch die dichotome Darstellung kontinuierlicher Ergebnisse Informationen verloren gehen, kann durch eine dichotome Darstellung nicht deutlich gemacht werden, ob eine kleine, mittlere oder starke Veränderung beim Patienten stattgefunden hat, was unter Umständen zusätzlich ein wichtiger Prädiktor für den Therapieerfolg darstellen könnte. Dass der Prädiktor frühe Veränderungen in den intrapersonalen Erfahrungen aus Therapeutensicht also für die dichotome Symptombelastung signifikant ist, aber nicht für die kontinuierliche, bedeutet also letztlich nur, dass diese mit einer reliable Veränderung assoziiert ist, aber das Ausmaß der Veränderung unklar bleibt.

Die Befunde replizieren dabei teilweise die Ergebnisse von Flückiger et al. (2013) und Mander et al. (2013) zu GCMs und erweitern dabei den Forschungsbereich um die frühen Veränderungen in GCMs. Mander et al. (2013) zeigte ebenfalls, dass die Patienteneinschätzung stärker mit dem Therapieergebnis assoziiert ist als die Therapeuteneinschätzung. Zudem fanden Mander et al. (2013) und Flückiger et al. (2013) ebenfalls, dass Problemaktualisierung nicht mit dem Therapieergebnis assoziiert ist. Darüber hinaus wurde auch die Interaktion der frühen Veränderungen in den GCMs und der Anzahl Sitzungen mit dem Therapieergebnis betrachtet. Hier zeigte sich, anders als bei Flückiger et al. (2013), ein signifikanter Zusammenhang zwischen der Interaktion der frühen Veränderung der intrapersonalen Erfahrungen aus Patientensicht mit der Anzahl Sitzungen für die kontinuierliche Symptombelastung. Die Therapeuteneinschätzung der frühen Veränderungen der intrapersonalen Erfahrungen, die bei Flückiger et al. (2013) keine signifikanten

Ergebnisse zeigte, war außerdem mit der dichotomen Symptombelastung signifikant assoziiert.

Der Artikel von Gmeinwieser et al. (2019a) weist einige Stärken und Schwächen auf. Die große Stichprobe von $N = 911$ Patienten und damit die hohe Testpower ist dabei eine klare Stärke des Artikels. Von den in diese Dissertation eingegangenen Artikeln weist dieser damit die größte Stichprobe auf. Zudem wurden Patientendaten aus der therapeutischen Routineversorgung genutzt. Der Artikel hat damit eine hohe externe Validität für Psychotherapieforschung in Deutschland. Ein wichtiger Punkt ist ebenfalls die Art der Prädiktor- und Ergebnisdefinition. Ein Ziel dieses Artikels war es, die Ergebnisse für die therapeutische Praxis leichter interpretierbar zu machen, ohne dabei zu starken Informationsverlust zu erleiden. Daher standen nicht die Baseline-Erhebungen der GCMs zu einem einzelnen Zeitpunkt im Vordergrund, sondern die frühe Veränderung über die Zeit. Wenn Therapeuten in der Praxis Patienten behandeln, werden in der Regel nicht die Werte z.B. der interpersonalen Erfahrungen zwischen Patienten herangezogen. Stattdessen ist die Veränderung über die Zeit für einen einzelnen Patienten ein wichtiger Wegweiser für das therapeutische Handeln. Mit den Ergebnissen aus dieser Studie kann diese frühe Veränderung in einen empirischen Kontext gerückt werden. Es konnte gezeigt werden, dass die frühe Veränderung einiger GCMs mit dem Therapieergebnis assoziiert ist. Auch die Definition des Kriteriums unterstützt eine praxisnahe Interpretation der Ergebnisse dieses Artikels. Zwar wurde zunächst das kontinuierliche Therapieergebnis untersucht, allerdings wurde auch das in der Praxis leichter interpretierbare Konzept der reliablen Verbesserung nach Jacobson und Truax (1991) genutzt, um die Ergebnisse besser interpretierbar zu machen. Folglich wurde sowohl bei der Prädiktorwahl, als auch bei der Kriteriumswahl stark darauf geachtet, die empirischen Ergebnisse mit dem therapeutischen Handeln vereinbar zu machen.

Auf der anderen Seite weist der Artikel auch einige Schwächen auf. Vor allem die schwache Generalisierbarkeit für Kontexte außerhalb Deutschlands ist dabei anzuführen. Mit

ca. 45 Sitzungen im Durchschnitt und eine Behandlungsdauer von ca. 1 ½ Jahren sind die Therapien in dieser Studie deutlich länger als in vergleichbaren Studien zu GCMs ($M = 30$ Sitzungen in Flückiger et al. [2013] und 6-10 Wochen stationäre Behandlung in der Studie von Mander et al. [2013]). Auch wenn das Therapieergebnis praxisnah interpretiert werden kann, kann beispielsweise kritisiert werden, dass innerhalb der Gruppen weiterhin starke Heterogenität vorliegen kann. Beispielsweise wurde nicht berücksichtigt, ob Patienten, die sich nicht reliabel verbessert haben, zum Präzeitpunkt überhaupt die Möglichkeit dazu gehabt hätten. Bei einem RCI der Symptombelastung von $RCI = 0,43$ müssen Patienten auch mindestens einen Wert von 0,43 der generellen Symptombelastung zum Präzeitpunkt vorliegen haben, um sich entsprechend dem RCI verbessern zu können. Das bedeutet, dass alle Patienten mit einer initialen Symptombelastung von weniger als 0,43 automatisch als nicht verbessert gelten. Schauenburg und Strack (1998) schlugen hierzu vor, verschiedene RCIs je nach Präwert anzuwenden, um diesem Problem vorzubeugen. Bei einer durchschnittlichen Symptombelastung der Stichprobe von $M = 0,99$ mit einer Standardabweichung von $SD = 0,62$ haben einige Patienten nicht die Möglichkeit, sich um 0,43 Punkte zu verbessern. Schauenburg und Strack (1998) schlugen daher einen RCI von 0,16 für weniger belastete Patienten vor. Die Anwendung verschiedener RCIs hätte die Stichprobe noch etwas genauer einteilen können und hätte mehr Patienten die Möglichkeit geboten, sich reliabel zu verbessern. Die Anwendung des RCI im Artikel von Gmeinwieser et al. (2019a) ist daher als konservativ zu bewerten. Unter Umständen übersieht diese Methode einige Patienten, die sich im weniger belasteten Bereich reliabel verbessern. Für die Schwächen der Studie durch die Nutzung der Berner Stundenbögen und allgemeiner Therapieergebnismaße (Lebenszufriedenheit, Symptombelastung) im Vergleich zu spezifischen Maßen soll an dieser Stelle auf den Abschnitt, „Zusammenfassende Betrachtung“, verwiesen werden.

Gmeinwieser et al. (2019b)

Im Artikel von Gmeinwieser et al. (2019b) wurde untersucht, ob die GCM-Outcome Assoziation von der motivationalen Inkongruenz moderiert wird. Auf der Basis von Grosse Holtforth et al. (2008) wurde dabei insbesondere die Assoziation zwischen den frühen intrapersonalen Erfahrungen und dem Therapieergebnis fokussiert. Explorativ wurden aber auch die frühen interpersonalen Erfahrungen und die frühe Problemaktualisierung analysiert. Als Therapieergebnismaß wurde die Symptombelastung zum Postmessungszeitpunkt herangezogen.

Die Ergebnisse replizieren einige Befunde, die sich bereits bei Gmeinwieser et al. (2019a) zeigten. Die Patienteneinschätzung der frühen inter- und intrapersonalen Erfahrungen korreliert signifikant mit dem Therapieergebnis wohingegen keine Assoziation zwischen den frühen therapeutengeschätzten GCMs und dem Therapieergebnis besteht. Die frühe Problemaktualisierung zeigt in keiner der beiden Perspektiven einen signifikanten Zusammenhang zum Therapieergebnis. Die Interaktionsanalyse bestätigt die Annahme, dass die motivationale Inkongruenz den Einfluss der frühen intrapersonalen Erfahrungen auf das Therapieergebnis moderiert. Die Johnson-Neyman Technik zeigte dabei, dass Patienten ab einem Inkongruenzwert von 2,43 und einem stärkeren Fokus auf die frühen intrapersonalen Erfahrungen bessere Therapieergebnisse zeigen. Unterhalb dieses Inkongruenzwerts findet sich kein Zusammenhang. Für einen Patienten mit einer hohen Diskrepanz zwischen inneren Zielen und deren Realisierung (d.h. einer hohen Inkongruenz) würde das bedeuten, dass gerade eine Therapie mit dem Fokus auf intrapersonale Erfahrungen, d.h. mit Strategien zur Problembewältigung und motivationalen Klärung, gewinnbringend sein könnte. Ein Patient mit Agoraphobie, aber dem Wunsch, am öffentlichen Leben teilzunehmen, könnte also deshalb von der Exposition profitieren, weil gerade die Erfolgs- bzw. Bewältigungserfahrung, die mit der Exposition einhergeht (Verringerung der Angst, Habituation), die Belastung für den Patienten verringert. Die frühen interpersonalen Erfahrungen aus Patientenperspektive

dagegen sind unabhängig vom Inkongruenzniveau mit dem Therapieergebnis assoziiert. Diese Assoziation ist allerdings im Regressionsmodell zusammen mit dem Prävwert der Symptombelastung und der Anzahl Sitzungen als Kovariaten, sowie der Inkongruenz nicht länger signifikant. Auch die Assoziation der frühen patienteneingeschätzten intrapersonalen Erfahrungen ist im Regressionsmodell nicht länger signifikant. Dies kann aber mit der signifikanten Interaktion der patienteneingeschätzten intrapersonalen Erfahrungen mit der Inkongruenz erklärt werden. Patienten, die eine geringe intrapersonale Erfahrung und eine hohe Inkongruenz angeben, zeigen im Vergleich zum Patientendurchschnitt die höhere Symptombelastung. Ist die intrapersonale Erfahrung dagegen hoch eingeschätzt, zeigen Patienten mit hoher Inkongruenz die geringere Symptombelastung. Diese disordinale Interaktion kann eine Erklärung für den nicht länger signifikanten Haupteffekt der patienteneingeschätzten intrapersonalen Erfahrung auf die Symptombelastung sein.

Auch die Studie von Gmeinwieser et al. (2019b) weist einige weitere Stärken und Schwächen auf. Eine Stärke des Artikels ist wiederum die große Stichprobe von $N = 524$ Patienten, sowie die externe Validität der Stichprobe. Auch die theoretische Einbettung in der Konsistenztheorie von Grawe (2004) kann als Stärke gesehen werden. Diese Einbettung zusammen mit der Studie von Grosse Holtforth et al. (2008) bildet eine nachvollziehbare theoretische und empirische Grundlage für die abgeleiteten und operationalisierten Hypothesen. Die initiale Symptombelastung in das Regressionsmodell mit aufzunehmen ist eine weitere Stärke der Studie. Am nicht länger signifikanten Haupteffekt der intra- und interpersonalen Erfahrungen aus Patientenperspektive wird die Wichtigkeit der initialen Symptombelastung für die Vorhersage der Symptombelastung zum Therapieende deutlich.

Der signifikante Interaktionseffekt der frühen intrapersonalen Erfahrungen kann allerdings auch kritisiert werden. Dieser weist genau die Schwächen auf, die von Jacobson und Truax (1991) angesprochen wurden. Die gefundene Interaktion von den patienteneingeschätzten intrapersonalen Erfahrungen und der Inkongruenz zeigt ein

unstandardisiertes Regressionsgewicht von $b = -0,061$ ($p = 0,039$), was einem standardisierten Beta von $\beta = -0,092$ und damit einem kleinen Effekt entspricht (Cohen, 1988). Wie genau sich dieser Effekt auf die Praxis individueller Patienten auswirkt, kann nur schwer erschlossen werden. Es kommt hinzu, dass in diesem Artikel die frühen GCM-Werte herangezogen und als Kriterium lediglich das kontinuierliche Therapieergebnis genutzt wurde. Ein Transfer des Effektes auf die Praxis wird damit weiter erschwert, da die patienteneingeschätzten GCMs stets im Vergleich zu anderen Patienten betrachten werden müssen und sich ein kleiner Effekt auf das Therapieergebnis bei einem einzelnen Patienten unter Umständen nicht spürbar auswirkt.

Gmeinwieser et al. (in press)

Die Studie untersuchte, ob die GCMs interpersonale Erfahrungen, intrapersonale Erfahrungen und Problemaktualisierung das Risiko vorhersagen, die Therapie abubrechen. Dabei wurde die Veränderung der Symptombelastung ebenfalls in das Modell aufgenommen. Als statistische Methode wurde die SA genutzt, um die longitudinale Datenstruktur und unterschiedliche Therapielängen verschiedener Patienten zu berücksichtigen. Wie auch in der Studie von Gmeinwieser et al. (2019a) wurden auch in dieser Studie einige Kovariaten berücksichtigt. Die Auswahl der Kovariaten folgte dabei einem Literaturreview von Roseborough, McLeod und Wright (2016) und einer Meta-Analyse von Swift und Greenberg (2012). Zusätzlich zu den GCMs und der Symptomschwere wurden demnach auch Patienten- und Therapeutengeschlecht, Patientenalter und die Diagnosekategorien Persönlichkeitsstörung, Angststörung (ohne Zwangsstörung), Zwangsstörung und die unipolare affektive Störung in das Modell mit aufgenommen. Keine der untersuchten Kovariaten zeigte sich signifikant in der SA abgesehen vom Patientenalter. Es zeigte sich, dass höheres Patientenalter mit einem geringeren Risiko assoziiert ist, die Therapie abubrechen. Unter den GCMs zeigten sich die interpersonalen Erfahrungen aus Patienten-

und aus Therapeutensicht als signifikant. In beiden Fällen sprechen höhere d.h. bessere interpersonale Erfahrungen für ein geringeres Risiko für Therapieabbruch. Überraschenderweise war die Veränderung der Symptombelastung kein signifikanter Prädiktor. Dies kann mehrere Gründe haben. Zum einen wurde die Veränderung der Symptombelastung nicht als reliable Veränderung über die Zeit gemessen und damit nicht als sogenannte sudden gains oder early gains, zu denen es eine Reihe Befunde gibt, die für eine Assoziation mit dem Therapieergebnis sprechen (z.B. Aderka, Nickerson, Bøe & Hofmann, 2012). Zum anderen kann aufgrund dieses Befundes nicht geschlossen werden, dass die Veränderung der Symptombelastung auch kein Prädiktor für das Beenden der Therapie in beidseitigem Einvernehmen ist. Dadurch, dass Therapieabbruch und das erfolgreiche Beenden der Therapie mit einer unterschiedlich langen Therapiedauer einhergehen, müsste für diese Fragestellung ein weiteres SA-Modell mit dem Endpunkt der erfolgreichen Beendigung der Therapie bestimmt werden. Der Einfluss der Veränderung der Symptombelastung auf das erfolgreiche Beenden der Therapie könnte nur anhand eines solchen Modells bestimmt werden und kann nicht als Umkehrschluss aus der hier berichteten Analyse gezogen werden. Unter den signifikanten GCMs fällt auf, dass die Therapeutenperspektive der interpersonalen Erfahrungen der stärkere Prädiktor für Abbruch darstellt. In Tabelle 4 ist dargestellt, welche Konventionen für Odds und für Hazard Ratios in Bezug zum Cohens d gelten. Diese Konventionen wurden von Borenstein, Hedges und Higgins (2009) abgeleitet. Das Hazard Ratio der interpersonalen Erfahrungen aus Therapeutensicht mit $HR = 0,589$ ist daher als klein und das Hazard Ratio aus Patientensicht mit $HR = 0,746$ als weniger als ein kleiner Effekt zu beurteilen.

Tabelle 4. Umrechnung von Cohens d in Odds und Hazard Ratios

Effektstärke	Cohens d	Odds/Hazard Ratio	
		Negativer Effekt	Positiver Effekt
Klein	0,2	0,7	1,43
Mittel	0,5	0,4	2,48
Groß	0,8	0,23	4,27

Auch in der Studie von Gmeinwieser et al. (in press) können weitere Stärken und Kritikpunkte genannt werden. Eine Stärke der Studie ist, dass Patienten, die die Therapie abgebrochen haben, Teil der Stichprobe sind und nicht ausgeschlossen wurde, wie in den beiden vorangegangenen Publikationen (Gmeinwieser et al. 2019a, 2019b). Des Weiteren wurden nicht nur die frühen GCMs oder die frühe Veränderung in GCMs, sondern das komplette in der Therapie erfasste ROM im Modell berücksichtigt. Insgesamt hat die SA also die vorliegenden longitudinalen Daten stärker umfassend abgebildet. Auf der anderen Seite ist die SA dadurch aber auch stärker fehleranfällig. Das liegt nicht an der Natur der SA, sondern an den Daten. Da der Fokus auf den genauen Tagen liegt, zu dem ein Ereignis, wie das Risiko für Therapieabbruch, geschätzt wird, müssen die Datumsangaben zu den jeweiligen Fragebögen übereinstimmen und einen realistischen Zeitraum abbilden. Welcher Zeitraum dabei allerdings als noch realistisch angesehen wird, kann auch mit dieser Arbeit nicht abschließen geklärt werden. Es wurden einige Ausschlusskriterien definiert, die auf den ersten Blick als stimmig und sinnvoll erscheinen, aber letztlich arbiträre Grenzen ziehen. Zum Beispiel wurden Patienten ausgeschlossen, die länger als 3 ½ Jahre Therapie in Anspruch genommen hatten. Selbst für die CBT in Deutschland, die ohnehin schon länger ist als in vergleichbaren Ländern, ist dieser Zeitraum sehr lang. Bei einem in dieser Studie berichteten Durchschnitt von $M = 433$ Behandlungstagen mit einer Standardabweichung von $SD = 228$

Tagen, ist ein Zeitraum von 3 ½ Jahren Psychotherapie bereits mehr als das Dreifache der Standardabweichung über dem Mittelwert der Behandlungstage und damit bereits sehr selten. Kriterien wie dieses oder anderen in dieser Studie angewendeten dienten daher stärker der statistischen Ausreißerkontrolle als theoretisch hergeleiteten Grenzen. Die berücksichtigten Kovariaten dagegen entstammten vergangener empirischer Forschung von etwa Roseborough et al. (2016) und Swift und Greenberg (2012). Überraschend ist dennoch, dass neben dem Patientenalter keine der anderen Kovariaten signifikant wurde. Nachfolgende Forschung sollte diesen Aspekt stärker beleuchten.

Zusammenfassende Betrachtung

Die in diese Dissertation eingegangenen Publikationen bereichern das Forschungsfeld um GCMs in einigen Aspekten, allerdings sind auch einige übergreifende Limitationen der Artikel zu nennen. Der folgende Abschnitt befasst sich mit Stärken, Schwächen und allgemeinen Diskussionspunkten.

Prinzipiell orientiert sich jeder der drei Artikel primär an einer vergangenen Publikation aus dem Bereich GCMs und erweitert diese um wesentliche Aspekte. Gmeinwieser et al. (2019a) erweitert primär die Studie von Flückiger et al. (2013). Anstelle von mittleren GCMs über die Zeit wurden frühe Veränderungen betrachtet. Die frühe Veränderung der Symptombelastung wurde nicht über ein einzelnes Item direkt erfasst, sondern indirekt mithilfe der SCL-90R (Franke, 2002) und der SCL-14 (Harfst et al., 2002) und damit etablierten, reliablen und validen Instrumenten zur Erfassung der Symptombelastung. Anstelle eines Kriteriums, das sich aus mehreren Skalen verschiedener Instrumente zusammensetzte, wurden bei Gmeinwieser et al. (2019a) die Symptombelastung und die Lebenszufriedenheit als Kriterien betrachtet. Darüber hinaus wurde nicht nur die kontinuierliche Post-Messung analysiert, sondern auch die reliable Verbesserung von

Patienten mittels logistischer Regression. Diese Aspekte führen letztlich dazu, dass die Ergebnisse nah am psychotherapeutischen Alltag liegen und interpretiert werden können.

Gmeinwieser et al. (2019b) erweitert primär die Studie von Grosse Holtforth et al. (2008), in der zwei Therapieansätze betrachtet wurden. In einer Gruppe von Patienten wurde ein rein bewältigungsorientierter Ansatz verfolgt und in der anderen Gruppe ein ganzheitlicherer Ansatz mit bewältigungs- und klärungsorientierten Methoden. Anstatt jedoch unterschiedliche Therapieansätze zu vergleichen und eine differentielle Vorhersage von Inkongruenz für den bewältigungs- und klärungsorientierten Therapieansatz anzunehmen, wurde bei Gmeinwieser et al. (2019b) die Interaktion zwischen Inkongruenz und der Skala der intrapersonalen Erfahrungen in Bezug zum Therapieergebnis untersucht. Zusätzlich wurden auch die GCMs interpersonale Erfahrungen und Problemaktualisierung auf die Interaktion mit motivationaler Inkongruenz untersucht. Eine weitere Stärke der Studie von Gmeinwieser et al. (2019b) ist die Große Stichprobe von 524 Patienten im Vergleich zu den 46 Patienten pro Gruppe in der Studie von Grosse Holtforth et al (2008).

Die Grundlage der Studie von Gmeinwieser et al. (in press) war eine Studie von Kegel und Flückiger (2015). Kegel und Flückiger (2015) betrachteten, inwiefern sich die Steigungen von Drop-Out und regulärem Therapieende hinsichtlich verschiedener GCMs unterschieden. Als GCMs wurden Selbstwert, Bewältigungserfahrung, motivationale Klärung und Allianz betrachtet. Gmeinwieser et al. (in press) betrachtete dagegen interpersonale Erfahrungen (inklusive Allianz), intrapersonale Erfahrungen (inklusive Bewältigungs- und Klärungserfahrungen) und Problemaktualisierung. Die Abbruchdefinition unterschied sich in den Artikeln von Kegel und Flückiger (2015) und Gmeinwieser et al. (in press). In Kegel und Flückiger (2015) galt eine Therapie als abgebrochen, wenn Patienten nicht die vollständige Postmessungsbatterie ausfüllten. Im Artikel von Gmeinwieser et al. (in press) wurde von Therapeuten beurteilt, ob eine Therapie als abgebrochen galt. Beide Operationalisierungen sind nach Swift und Greenberg (2012) gängige Formen, Therapieabbruch zu quantifizieren.

Die unterschiedlichen Definitionen können allerdings auch eine Erklärung für die in beiden Artikeln deutlich unterschiedlichen Abbruchrate sein (57% in Kegel und Flückiger [2015] gegenüber 21% in Gmeinwieser et al [in press]). Ein weiterer Unterschied in beiden Studien war die statistische Analyse. Anstelle hierarchischer linearer Modelle wurden die GCMs in Gmeinwieser et al. (in press) mittels SA betrachtet. Mithilfe der SA war es möglich, zeitlich variable und zeitlich stabile Prädiktoren in einem Modell zu betrachten. Dadurch konnten auch deutlich mehr mögliche Kovariaten betrachtet werden als bei Kegel und Flückiger (2015).

Interpersonale Erfahrungen, intrapersonale Erfahrungen und Problemaktualisierung wurden in allen Publikationen gemeinsam untersucht. Über alle Publikationen hinweg erschienen die (frühen) interpersonalen Erfahrungen aus Patientenperspektive einer der konsistentesten Prädiktoren zu sein. Die frühe Veränderung der patienteneingeschätzten interpersonalen Erfahrungen sagt kontinuierliches wie auch dichotomes Therapieoutcome vorher, abgesehen von der kontinuierlichen Symptombelastung (Gmeinwieser et al., 2019a), sie werden nicht von der motivationalen Inkongruenz moderiert (Gmeinwieser et al., 2019b) und zeigen sich zudem als Prädiktor für das Drop-out-Risiko über den Gesamtverlauf der Therapie (Gmeinwieser et al., in press). Die Therapeutenperspektive der frühen Veränderungen in den interpersonalen Erfahrungen zeigt dagegen lediglich einen Zusammenhang mit der dichotomen Symptombelastung, sowie in Interaktion mit der Anzahl Sitzungen signifikante Zusammenhänge mit der kontinuierlichen und der dichotomen Lebenszufriedenheit (Gmeinwieser et al., 2019a). Zudem zeigt die Therapeuteneinschätzung der interpersonalen Erfahrungen einen Zusammenhang mit dem Risiko für Drop-out (Gmeinwieser et al., in press). Die intrapersonalen Erfahrungen zeigen ebenfalls im Vergleich weniger konsistente Ergebnisse. Die Therapeutenperspektive der frühen (Veränderung in den) intrapersonalen Erfahrungen hängt lediglich mit der dichotomen Symptombelastung zusammen (Gmeinwieser et al., 2019a) und zeigt keine Interaktion mit der motivationalen

Inkongruenz (Gmeinwieser et al., 2019b) oder einen Zusammenhang zum Drop-out-Risiko (Gmeinwieser et al., in press). Die patienteneingeschätzten (frühen Veränderungen der) intrapersonalen Erfahrungen zeigt dagegen in Interaktion mit der Anzahl Sitzungen (Gmeinwieser et al., 2019a) und in Interaktion mit der Inkongruenz (Gmeinwieser et al., 2019b) signifikante Zusammenhänge mit Symptombelastung und zeigt auch ohne Interaktion einen Zusammenhang mit der Lebenszufriedenheit (Gmeinwieser et al., 2019a). Allerdings hängen die patienteneingeschätzten intrapersonalen Erfahrungen nicht mit dem Risiko für Drop-out zusammen (Gmeinwieser et al., in press). Die Problemaktualisierung ist dagegen perspektivenunabhängig der schwächste Prädiktor und zeigt in keiner der drei Studien einen signifikanten Zusammenhang mit dem Therapieergebnis (Gmeinwieser et al., 2019a, 2019b, in press).

Die Publikationen weisen darüber hinaus auch einige übergreifende Stärken und Schwächen auf. Eine Stärke der Studien ist, dass diese überwiegend praxisrelevante Ergebnisse zeigen. Patienten können auf Basis ihrer frühen GCM-Veränderungen als mehr oder weniger wahrscheinlich eingeschätzt werden, zum Ende der Therapie als reliabel verbessert zu gelten (Gmeinwieser et al., 2019a), ab einem Inkongruenzcutoff von 2,43 sollten Therapeuten stärker intrapersonale Interventionen fokussieren (Gmeinwieser et al., 2019b) und um Therapieabbruch vorzubeugen sollten während des gesamten Therapieverlaufs vor allem interpersonale Erfahrungen fokussiert werden (Gmeinwieser et al., in press).

Eine weitere Stärke der Studien ist die Stichprobe. Dabei lässt sich nicht nur die Stichprobengröße von mehr als 500 Patienten über die Studien als positiv bewerten, sondern auch die Nähe der Stichprobe am (deutschen) therapeutischen Alltag. Alle in den Studien berichteten Ergebnisse entstammen der Routineversorgung von Patienten an einer psychotherapeutischen Ambulanz und können daher auch auf ambulante Patienten in Deutschland generalisiert werden.

Auf der anderen Seite zeigen die hier berichteten Studien auch einige Schwächen und Limitationen. Die von Grawe (1997) als wichtiger GCM deklarierter Ressourcenaktivierung wurde in keiner der Studien untersucht, obwohl Studien zum Zusammenhang mit dem Therapieergebnis vorliegen (z.B. Flückiger & Grosse Holtforth, 2008; Gassmann & Grawe, 2006; Stangier et al., 2010). Prinzipiell wurde die Ressourcenaktivierung mit den Berner Stundenbögen auch erfasst und hätte als weiterer Prädiktor in die Modelle aufgenommen werden können. Der Grund dafür, die Ressourcenaktivierung nicht mit aufzunehmen war allerdings, dass diese nicht als Sekundärskala des Berner Stundenbogens von Flückiger et al. (2013) beschrieben wurde. Auch Schwartz, Hilbert, Schlegl, Diedrich und Voderholzer (2018) argumentieren dafür, die Ressourcenaktivierung, aufgrund der großen Überschneidung zur Skala Therapiebeziehung (enthalten in den interpersonalen Erfahrungen), nicht zu analysieren. Es kommt hinzu, dass die Ressourcenaktivierung lediglich mit dem TSTB erfasst wird und nicht mit dem PSTB und damit keine Möglichkeit einer vergleichbaren Perspektive gegeben ist.

Die Stichprobe und das Setting können in einigen Aspekten als Stärke der Studien herangezogen werden, allerdings schränken diese auch die Aussagekraft in einigen Aspekten ein. Beispielsweise ist die Dauer der Therapien in allen Studien mit mehr als 40 Sitzungen pro Patient deutlich höher als im deutschsprachigen, aber vor allem auch im nicht-deutschsprachigen Ausland, wie etwa der Schweiz oder den USA. Dies ermöglicht zwar eine Generalisierung auf deutsche ambulante Patienten, schränkt die Aussagekraft aber in Bezug zu Stichproben aus anderen Ländern ein. Die Patienten wurden außerdem beinahe ausschließlich von Psychotherapeuten in Ausbildung behandelt. Die Generalisierung auf approbierte oder in anderen Ländern lizenzierte Psychotherapeuten ist daher unter Vorbehalt zu betrachten.

Ein weiterer Punkt sind die fehlenden Werte und der Therapieabbruch. In den Studien von Gmeinwieser et al. (2019a, 2019b) wurde kein Therapieabbruch berücksichtigt. Bei einer

Abbruchrate von etwa 21% (Gmeinwieser et al., in press) ist die Generalisierbarkeit auf Patienten, die die Therapie einseitig beenden daher ebenfalls eingeschränkt. Gmeinwieser et al. (2019a) berichtet zwar, dass sich aufgrund von fehlenden Werten ausgeschlossene Patienten nicht signifikant von Patienten unterscheiden, deren Daten vollständig vorliegen, dennoch kann ein Therapieabbruch, wie in Gmeinwieser et al. (in press) bereits argumentiert, aufgrund der unterschiedlichen Therapielängen generell mit anderen Variablen assoziiert sein als eine erwartungsgemäße Beendigung der Therapie

Die Nutzung allgemeiner Maße zur Therapieergebnismessung ist ebenfalls eine Limitation der Studien. Lebenszufriedenheit (Gmeinwieser et al., 2019a) und Symptombelastung (Gmeinwieser et al., 2019a, 2019b; als Prädiktor auch in Gmeinwieser et al., in press) lassen sich zwar durch ihre breite Problemerkennung sinnvoll auf heterogene Stichproben mit unterschiedlichen Diagnosen anwenden, allerdings müssen sich unterschiedliche Störungen nicht zwangsläufig auffällig auf diesen Maßen zeigen. Depressivität wird beispielsweise mit der SCL-90R auf 13 Items erfasst. Ein Patient mit einer isolierten Depression, der auf jeder dieser Items die zweithöchste Ausprägung angibt, aber auf den anderen Items keine Auffälligkeit zeigt, würde einen Gesamtwert von 0,43 erreichen, was, laut Schauenburg und Strack (1998), noch im klinisch unauffälligen Bereich liegt. Zudem würde sich dieser Patient, bei einem RCI von genau 0,43, laut den Autoren, auch nur dann reliabel verbessern, wenn in keinem der Items nach der Therapie noch eine Auffälligkeit gezeigt würde. An dieser Stelle sind spezifische Maße sinnvoll, wie etwa das Beck Depression Inventory (Beck, Ward, Mendelson, Mock & Erbaugh, 1961), durch das kleinere Veränderungen schneller und leichter ersichtlicher wären, da das Instrument auf die kontinuierliche Messung von Depression ausgelegt ist. Aufgrund der gemischten Stichprobe war es allerdings nicht möglich, für jedes Störungsbild spezifische Messinstrumente heranzuziehen.

Die statistische Fundierung der interpersonalen und der intrapersonalen Erfahrungen kann ebenfalls kritisch betrachtet werden. Dass die Faktoren als Sekundärfaktoren aus den Berner Stundenbögen extrahiert wurden und dabei Primärskalen umfassen, die für sich genommen bereits genutzt werden könnten (z.B. die Skalen Therapiebeziehung, motivationale Klärung, Bewältigungserfahrung), erschwert letztlich den Bezug der Ergebnisse zu anderen Studien, die die Primärskalen untersuchten. Zudem fehlen der faktorenanalytische Vergleich von Primär- und Sekundärskalen, sowie generelle Fit-Indizes der Sekundärskalen. Eine Möglichkeit, die Faktoren stärker zu etablieren wäre etwa eine konfirmatorische Faktorenanalyse, die die Skaleneinteilung bestätigt und so eine statistische Grundlage bietet. Zudem sollte die theoretische Grundlage weiter ausgebaut werden, z.B. mit einem Modell, das beschreibt, zu welchem Zeitpunkt inter- oder intrapersonale Erfahrungen aktiv sind und, wenn in der Therapie fokussiert, in welchem Ausmaß oder auf welcher Dimension zu einer therapeutischen Veränderung führen könnten. Angesichts der kleinen Effekte, welche die in diese Arbeit eingegangenen Studien zu interpersonale und intrapersonalen Erfahrungen (Gmeinwieser et al., 2019a, 2019b, in press) zeigen, ist allerdings fraglich, ob es gerechtfertigt ist, speziell für diese Faktoren eine fundiertere theoretische Grundlage zu formulieren. Die ursprünglichen Konzepte von Grawe (1998), Bewältigungserfahrungen, motivationale Klärung, Ressourcenaktivierung und Problemaktualisierung könnten ebenso herangezogen werden. Dadurch würde auch die Vergleichbarkeit zu Artikeln gegeben sein, die diese Konzepte bereits betrachtet haben (z.B. Call et al., 2018; D'Errico et al., 2019; Gassmann & Grawe, 2006; Kegel & Flückiger, 2016; Mander et al., 2013, 2015; Sander et al., 2012; Schwartz et al., 2018; Vogel et al., 2016).

Zukünftige Forschung und Ausblick

Zukünftige Forschung sollte an den zuvor genannten Punkten ansetzen: Die Rolle der Ressourcenaktivierung scheint vielversprechend und sollte näher beleuchtet werden. Eine

Möglichkeit, diese und andere GCMs mit einem aktualisierten Instrument zu erfassen, ist dabei der von Mander et al. (2015) entwickelte Individual Therapy Process Questionnaire, der darüber hinaus auch GCMs wie die emotionale Bindung oder die Patientenangst, sich zu öffnen, erfasst. Settings sollten stärker an Therapien im Ausland angelehnt sein, um die Vergleichbarkeit und Generalisierbarkeit der Befunde zu erhöhen. Das bedeutet auch, dass GCMs häufiger erfasst werden sollten, z.B. jede oder jede zweite Sitzung, um gerade den möglicherweise wechselseitigen Einfluss zwischen GCMs und Symptombelastung oder letztlich den Zusammenhang zwischen GCMs und dem Therapieergebnis näher betrachten zu können. Laut Cuijpers et al. (2018) stellt dies eine gute Möglichkeit dar, herauszufinden, was Psychotherapie genau wirksam macht. Eine Implikation, die über den Erkenntnisgewinn von Therapieprozessen hinausgeht, ist zudem die Anpassung der Therapie an die spezifischen Bedürfnisse von Patienten. Ähnlich wie in Gmeinwieser et al. (2019b) für Patienten mit hoher Inkongruenz gezeigt, die ein besseres Therapieergebnis bei stärkerem Fokus auf intrapersonale Erfahrungen zeigten, könnte so Therapie auf die individuellen Bedürfnisse und Probleme von Patienten abgestimmt und durch Verlaufskontrollen sogar im Therapieprozess angepasst werden.

Nicht zuletzt sollte auch darüber nachgedacht werden, einen laborexperimentellen Ansatz mit den GCMs zu verfolgen. Ähnlich wie in der Studie von Grosse Holtforth et al. (2008) könnten Patienten randomisierten Therapieformen zugewiesen werden, die bestimmte GCMs mehr oder weniger fokussieren. Zwar zeigt jede der Studien von Gmeinwieser et al. (2019a, 2019b, in press) einen längsschnittlichen Zusammenhang der GCMs auf das Therapieergebnis und liefert daher Hinweise auf eine kausale Beziehung, allerdings könnten randomisiert kontrollierte Studien diese kausale Beziehung empirisch und auf Goldstandard-Niveau festigen. Ob in diesen zukünftigen Studien die Konzepte der inter- und intrapersonalen Erfahrung den ursprünglich formulierten GCMs nach Grawe (1997) vorgezogen werden sollten, ist allerdings fraglich.

In Gmeinwieser et al. (2019a) wird berichtet, dass sich 42% der Patienten im Bereich Symptombelastung und 46% der Patienten im Bereich Lebenszufriedenheit bis zum Therapieende reliabel verbessern. Im Umkehrschluss bedeutet das, dass sich über 50% der Patienten während der Therapie nicht reliabel verbessert haben. Vorausgegangene Studien berichten in vergleichbaren Settings ähnliche Werte und zeigen damit, dass es sich hierbei nicht um einen Einzelfund handelt (z.B. Jacobi, Uhlmann & Hoyer, 2011; Nelson & Hiller, 2013). Diese Zahlen zeigen, warum die Forschung zu GCMs so wichtig ist. Die dahinterliegende Frage der GCMs ist letztlich, welche Interventionen und Techniken für welchen Patienten wirksam sind. Wie sollte eine Therapie gestaltet sein, damit Patienten mit ihren individuellen Voraussetzungen maximal profitieren können? Die Forschung zur Allianz lieferte bereits Hinweise, die von der GCM-Forschung und nicht zuletzt den hier verwendeten Artikeln weiter ausgebaut wurde. GCMs sind differentiell für Patienten wirksam und können schulenübergreifend angewandt werden. Weitere Forschung ist aber unbedingt nötig, damit sich die Psychotherapie stärker den individuellen Bedürfnissen von Patienten anpassen kann.

Literaturverzeichnis zum Rahmentext

- Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(1), 93–101. <https://doi.org/10.1037/a0026455>
- 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>
- Altmann, U., Wittmann, W. W., Steffanowski, A., Heymann, F. v., Bruckmayer, E., Pfaffinger, I., . . . Strauß, B. M. (2014). Abbrüche antragspflichtiger ambulanter Psychotherapien: Häufigkeit, Risikofaktoren, Outcome [Non-completion of insurance-funded outpatient psychotherapy: frequency, risk factors, outcome]. *Psychiatrische Praxis, 41*(6), 305–312. <https://doi.org/10.1055/s-0034-1370172>
- Altmann, U., Thielemann, D., Zimmermann, A., Steffanowski, A., Bruckmeier, E., Pfaffinger, I., . . . Strauß, B. (2018). Outpatient Psychotherapy Improves Symptoms and Reduces Health Care Costs in Regularly and Prematurely Terminated Therapies. *Frontiers in Psychology, 9*, 748. <https://doi.org/10.3389/fpsyg.2018.00748>
- Barber, J. P., Connolly, M. B., Crits-Christoph, P., Gladis, L., & Siqueland, L. (2000). Alliance predicts patients' outcome beyond in-treatment change in symptoms. *Journal of Consulting and Clinical Psychology, 68*(6), 1027–1032. <https://doi.org/10.1037//0022-006X.68.6.1027>
- Bassler, M., Potratz, B., & Krauthauser, H. (1995). Der "Helping Alliance Questionnaire" (HAQ) von Luborsky. *Psychotherapeut. (40)*, 23–32.
- Beck, A. T. (1970). Cognitive therapy: Nature and relation to behavior therapy. *Behavior therapy, 1*(2), 184–200. [https://doi.org/10.1016/S0005-7894\(70\)80030-2](https://doi.org/10.1016/S0005-7894(70)80030-2)

- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. (1961).). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561–571. <https://doi.org/10.1001/archpsyc.1961.01710120031004>
- Bergman, H., Kornør, H., Nikolakopoulou, A., Hanssen-Bauer, K., Soares-Weiser, K., Tollefsen, T. K., & Bjørndal, A. (2018). Client feedback in psychological therapy for children and adolescents with mental health problems. *The Cochrane Database of Systematic Reviews*, 8. <https://doi.org/10.1002/14651858.CD011729.pub2>.
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy*, 16(3), 252–260. <https://doi.org/10.1037/h0085885>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis* (Reprinted.). Chichester: Wiley.
- 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>
- Call, M., D'Errico, L., Bents, H., & Mander, J. (2018). Allgemeine Wirkfaktoren in ambulanter Ausbildungstherapie. *Psychotherapeut*, 63(6), 473–482. <https://doi.org/10.1007/s00278-018-0310-4>
- Carroll, L. (2000). *Alice's Adventures in Wonderland*. Chicago: VolumeOne Publishing. Retrieved from https://www.adobe.com/be_en/active-use/pdf/Alice_in_Wonderland.pdf (Originalwerk publiziert 1865)
- Cinkaya, F., Schindler, A., & Hiller, W. (2011). Wenn Therapien vorzeitig scheitern. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 40(4), 224–234. <https://doi.org/10.1026/1616-3443/a000121>

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale: Lawrence Erlbaum Associates.
- Crits-Christoph, P. (1997). Limitations of the Dodo Bird Verdict and the Role of Clinical Trials in Psychotherapy Research: Comment on Wampold et al. (1997). *Psychological bulletin*, *122*(3), 216–220.
- D’Errico, L., Call, M., Blanck, P., Vonderlin, E., Bents, H., & Mander, J. (2019). Associations between mindfulness and general change mechanisms in individual therapy: Secondary results of a randomised controlled trial. *Counselling and Psychotherapy Research*, *1*(2), 113. <https://doi.org/10.1002/capr.12233>
- DeRubeis, R. J., Brotman, M. A., & Gibbons, C. J. (2005). A Conceptual and Methodological Analysis of the Nonspecifics Argument. *Clinical Psychology: Science and Practice*, *12*(2), 174–183. <https://doi.org/10.1093/clipsy/bpi022>
- DeRubeis, R. J., & Crits-Christoph, P. (1998). Empirically Supported Individual and Group Psychological Treatments for Adult Mental Disorders. *Journal of Consulting and Clinical Psychology*, *66*(1), 37–52.
- Elkins, D. N. (2009). The Medical Model in Psychotherapy. *Journal of Humanistic Psychology*, *49*(1), 66–84. <https://doi.org/10.1177/0022167807307901>
- Ellis, A. (1957). Outcome of employing three techniques of psychotherapy. *Journal of Clinical Psychology*, *13*(4), 344–350. [https://doi.org/10.1002/1097-4679\(195710\)13:4<344::AID-JCLP2270130407>3.0.CO;2-9](https://doi.org/10.1002/1097-4679(195710)13:4<344::AID-JCLP2270130407>3.0.CO;2-9)
- Falkenström, F., Granström, F., & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology*, *60*(3), 317–328. <https://doi.org/10.1037/a0032258>

- Falkenström, F., Granström, F., & Holmqvist, R. (2014). Working alliance predicts psychotherapy outcome even while controlling for prior symptom improvement. *Psychotherapy Research*, 24(2), 146–159. <https://doi.org/10.1080/10503307.2013.847985>
- Feeley, M., DeRubeis, R. J., & Gelfand, L. A. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 67(4), 578–582. <https://doi.org/10.1037/0022-006X.67.4.578>
- Flückiger, C., Del Re, 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>
- Flückiger, C., & Grosse Holtforth, M. (2008). Focusing the therapist's attention on the patient's strengths: A preliminary study to foster a mechanism of change in outpatient psychotherapy. *Journal of Clinical Psychology*, 64(7), 876–890. <https://doi.org/10.1002/jclp.20493>
- Flückiger, C., Grosse Holtforth, M., Znoj, H. J., Caspar, F., & Wampold, B. E. (2013). Is the relation between early post-session reports and treatment outcome an epiphenomenon of intake distress and early response? A multi-predictor analysis in outpatient psychotherapy. *Psychotherapy Research*, 23(1), 1–13. <https://doi.org/10.1080/10503307.2012.693773>
- Flückiger, C., Meyer, A., Wampold, B. E., Gassmann, D., Messerli-Bürgy, N., & Munsch, S. (2011). Predicting Premature Termination Within a Randomized Controlled Trial for Binge-Eating Patients. *Behavior therapy*, 42(4), 716–725.

- Flückiger, C., Regli, D., Zwahlen, D., Hostettler, S., & Caspar, F. (2010). Der Berner Patienten- und Therapeutenstundenbogen 2000. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 39(2), 71–79. <https://doi.org/10.1026/1616-3443/a000015>
- 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>
- Franke, G. H. (2002). *SCL-90-R: Symptom Checkliste von L.R. Derogatis (deutsche Version)*. Goettingen: Beltz.
- Gassmann, D., & Grawe, K. (2006). General change mechanisms: The relation between problem activation and resource activation in successful and unsuccessful therapeutic interactions. *Clinical Psychology & Psychotherapy*, 13(1), 1–11. <https://doi.org/10.1002/cpp.442>
- Gmeinwieser, S., Hagmayer, Y., Pieh, C., & Probst, T. (2019b). General change mechanisms in the early treatment phase and their associations with the outcome of cognitive behavioural therapy in patients with different levels of motivational incongruence. *Clinical Psychology & Psychotherapy*, 26(5), 550–561. <https://doi.org/10.1002/cpp.2381>
- Gmeinwieser, S., Kuhlencord, M., Ruhl, U., Hagmayer, Y., & Probst, T. (2019a). Early developments in general change mechanisms predict reliable improvement in addition to early symptom trajectories in cognitive behavioral therapy. *Psychotherapy Research*. Advance online publication. <https://doi.org/10.1080/10503307.2019.1609709>

- Gmeinwieser, S., Schneider, K. S., Bardo, M., Brockmeyer, T., Hagmayer, Y. (in press). Risk for psychotherapy drop-out in survival analysis: The influence of general change mechanisms and symptom severity. *Journal of Counseling Psychology*. <https://doi.org/10.1037/cou0000418>
- Grawe, K. (1997). Research-Informed Psychotherapy. *Psychotherapy Research*, 7(1), 1–19. <https://doi.org/10.1080/10503309712331331843>
- Grawe, K. (1998). *Psychologische Therapie*. Göttingen: Hogrefe.
- Grawe, K. (1999). Gründe und Vorschläge für eine Allgemeine Psychotherapie. *Psychotherapeut*, 44(6), 350–359. <https://doi.org/10.1007/s002780050190>
- Grawe, K. (2004). *Neuropsychotherapie*. Göttingen: Hogrefe.
- Greenson, R. R. (1965). The Working Alliance and the Transference Neurosis. *The Psychoanalytic Quarterly*, 34(2), 155–181. <https://doi.org/10.1080/21674086.1965.11926343>
- Grosse Holtforth, M., Grawe, K., Fries, A., & Znoj, H. (2008). Inkonsistenz als differenzielles Indikationskriterium in der Psychotherapie. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 37(2), 103–111. <https://doi.org/10.1026/1616-3443.37.2.103>
- Harfst, T., Koch, U., Kurtz von Aschoff, C., Nutzinger, D. O., Rüdell, H., & Schulz, H. (2002). Entwicklung und Validierung einer Kurzform der Symptom Checklist-90-R. *DRV-Schriften*, 33, 71-73.
- Henrich, G., & Herschbach, P. (2000). Questions on Life Satisfaction (FLZM) - A Short Questionnaire for Assessing Subjective Quality of Life. *European Journal of Psychological Assessment*. 16(3), 150–159.

- 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 and Research*, 36(5), 427–440. <https://doi.org/10.1007/s10608-012-9476-1>.
- Horvath, A. O., Del Re, A. C., Fluckiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy*, 48(1), 9–16. <https://doi.org/10.1037/a0022186>
- Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, 36(2), 223–233.
- Horvath, A. O., & Symonds, D. (1991). Relation Between Working Alliance and Outcome in Psychotherapy: A Meta-Analysis. *Journal of Counseling Psychology*, 38(2), 139–149.
- Issakidis, C., & Andrews, G. (2004). Pretreatment attrition and dropout in an outpatient clinic for anxiety disorders. *Acta Psychiatrica Scandinavica*, 109(6), 426–433. <https://doi.org/10.1111/j.1600-0047.2004.00264.x>
- Jacobi, F., Uhlmann, S., & Hoyer, J. (2011). Wie häufig ist therapeutischer Misserfolg in der ambulanten Psychotherapie? *Zeitschrift für Klinische Psychologie und Psychotherapie*, 40(4), 246–256. <https://doi.org/10.1026/1616-3443/a000122>
- Jacobson, N. S., Follette, W. C., & Revenstorf, 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>
- 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>

- Knox, S., Adrians, N., Everson, E., Hess, S., Hill, C., & Crook-Lyon, R. (2011). Clients' perspectives on therapy termination. *Psychotherapy Research, 21*(2), 154–167. <https://doi.org/10.1080/10503307.2010.534509>
- Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research, 17*(1), 1–14. <https://doi.org/10.1080/10503300601032506>
- Lambert, M. J. (2013). The Efficacy and Effectiveness of Psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed.). New York: Wiley.
- Leichsenring, F., & Rabung, S. (2008). Effectiveness of Long-term Psychodynamic Psychotherapy. *Journal of the American Medical Association, 300*(13), 1551–1565.
- Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative Studies of Psychotherapies: Is It True That "Everyone Has Won and All Must Have Prizes"? *Archives of General Psychiatry, 32*, 995–1008.
- Lutz, W., Jong, K. de, & 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., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., . . . Barlow, D. H. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology, 82*(2), 287–297. <https://doi.org/10.1037/a0035535>.
- 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), 7819. <https://doi.org/10.1038/s41598-018-25953-0>
- Mander, J., Schlarb, A., Teufel, M., Keller, F., Hautzinger, M., Zipfel, S., . . . Sammet, I. (2015). The individual therapy process questionnaire: Development and validation of a revised measure to evaluate general change mechanisms in psychotherapy. *Clinical Psychology & Psychotherapy*, 22(4), 328–345. <https://doi.org/10.1002/cpp.1892>
- Mander, J. V., Wittorf, A., Schlarb, A., Hautzinger, M., Zipfel, S., & Sammet, I. (2013). Change mechanisms in psychotherapy: Multiperspective assessment and relation to outcome. *Psychotherapy Research*, 23(1), 105–116. <https://doi.org/10.1080/10503307.2012.744111>
- Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68(3), 438–450. <https://doi.org/10.1037//0022-006X.68.3.438>
- Navarro, D. (2015). Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.5). Adelaide, Australia: University of Adelaide
- Nelson, N. E., & Hiller, W. (2013). Therapeutischer Misserfolg in der ambulanten Psychotherapie. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 42(4), 217–229. <https://doi.org/10.1026/1616-3443/a000223>
- Ogles, B. M., Lunnen, K. M., & Bonesteel, K. (2001). Clinical significance. *Clinical psychology review*, 21(3), 421–446. [https://doi.org/10.1016/S0272-7358\(99\)00058-6](https://doi.org/10.1016/S0272-7358(99)00058-6)
- Pavlov, I. P. (1928). *Conditioned Reflexes: an Investigation of the Physiological Activity of the Cerebral Cortex*. Oxford: Oxford University Press.

- Piselli, A., Halgin, R. P., & Macewan, G. H. (2011). What went wrong? Therapists' reflections on their role in premature termination. *Psychotherapy Research, 21*(4), 400–415. <https://doi.org/10.1080/10503307.2011.573819>
- Probst, T., Jakob, M., Kaufmann, Y. M., Müller-Neng, J. M. B., Bohus, M., & Weck, F. (2018). Patients' and therapists' experiences of general change mechanisms during bug-in-the-eye and delayed video-based supervised cognitive-behavioral therapy. A randomized controlled trial. *Journal of Clinical Psychology, 74*(4), 509–522. <https://doi.org/10.1002/jclp.22519>
- Probst, T., Lambert, M. J., Loew, T. H., Dahlbender, R. W., Göllner, R., & Tritt, K. (2013). Feedback on patient progress and clinical support tools for therapists: Improved outcome for patients at risk of treatment failure in psychosomatic in-patient therapy under the conditions of routine practice. *Journal of Psychosomatic Research, 75*(3), 255–261. <https://doi.org/10.1016/j.jpsychores.2013.07.003>
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Revelle, W. (2018). Psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych>, Version = 1.8.4.
- Richard Sterba's (1934) "The fate of the ego in analytic therapy". Panel report (1994). *Journal of the American Psychoanalytic Association, 42*(3), 863–873. <https://doi.org/10.1177/000306519404200310>
- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*(2), 95–103. <https://doi.org/10.1037/h0045357>

- Roseborough, D. J., McLeod, J. T., & Wright, F. I. (2016). Attrition in Psychotherapy. *Research on Social Work Practice, 26*(7), 803–815. <https://doi.org/10.1177/1049731515569073>
- Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry, 6*, 412–415.
- Rubel, J. A., Rosenbaum, D., & Lutz, W. (2017). Patients' in-session experiences and symptom change: Session-to-session effects on a within- and between-patient level. *Behaviour Research and Therapy, 90*, 58–66. <https://doi.org/10.1016/j.brat.2016.12.007>
- Saatsi, S., Hardy, G. E., & Cahill, J. (2007). Predictors of outcome and completion status in cognitive therapy for depression. *Psychotherapy Research, 17*(2), 185–195. <https://doi.org/10.1080/10503300600779420>
- Sander, L., Zobel, I., Dykieriek, P., & Schramm, E. (2012). Allgemeine Wirkfaktoren bei der Behandlung chronisch depressiver Patienten – Vergleich zweier Psychotherapieansätze. *Verhaltenstherapie, 22*(4), 228–235. <https://doi.org/10.1159/000345644>
- Schauenburg, H., & Strack, M. (1998). Die Symptom Checklist-90-R (SCL-90-R) zur Darstellung von statistisch und klinischsignifikanten Psychotherapieergebnissen. *PPmP. Psychotherapie, Psychosomatik, medizinische Psychologie, 48*(7), 257–264.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research, 99*(6), 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Schulte, D. (1996). *Therapieplanung*. Göttingen: Hogrefe.

- Schwartz, C., Hilbert, S., Schlegl, S., Diedrich, A., & Voderholzer, U. (2018). Common change factors and mediation of the alliance-outcome link during treatment of depression. *Journal of Consulting and Clinical Psychology, 86*(7), 584–592. <https://doi.org/10.1037/ccp0000302>
- Sharf, J., Primavera, L. H., & Diener, M. J. (2010). Dropout and therapeutic alliance: A meta-analysis of adult individual psychotherapy. *Psychotherapy (Chicago, Ill.), 47*(4), 637–645. <https://doi.org/10.1037/a0021175>
- 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>
- Simon, G. E., & Ludman, E. J. (2010). Predictors of early dropout from psychotherapy for depression in community practice. *Psychiatric Services, 61*(7), 684–689. <https://doi.org/10.1176/ps.2010.61.7.684>
- Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts, Inc.
- Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60*(3), 402–408. <https://doi.org/10.1037/0022-006X.60.3.402>
- Stangier, U., Consbruch, K. von, Schramm, E., & Heidenreich, T. (2010). Common factors of cognitive therapy and interpersonal psychotherapy in the treatment of social phobia. *Anxiety, Stress, and Coping, 23*(3), 289–301. <https://doi.org/10.1080/10615800903180239>

- Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(4), 547–559. <https://doi.org/10.1037/a0028226>
- Task Force on Promotion and Dissemination of Psychological Procedures (1995). Training in and dissemination of empirically-validated psychological treatment: Report and recommendations. *The Clinical Psychologist, 48*, 2–23.
- Tang, T. Z., DeRubeis, R. J., Beberman, R., & Pham, T. (2005). Cognitive changes, critical sessions, and sudden gains in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 73*(1), 168–172. <https://doi.org/10.1037/0022-006X.73.1.168>
- Universität Bern (n.d.). Klinische Psychologie und Psychotherapie: Weiterbildung. Retrieved from: <https://blog.apastyle.org/apastyle/2016/11/writing-website-in-text-citations-and-references.html> (Zugriff am 26.07.2019)
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. (4th ed.). New York: Springer
- Vogel, E., Blanck, P., Bents, H., & Mander, J. (2016). Wirkfaktoren in der Gruppentherapie: Entwicklung und Validierung eines Fragebogens [Change Factors in Group Therapy: Development and Validation of a Questionnaire]. *Psychotherapie, Psychosomatik, medizinische Psychologie, 66*(5), 170–179. <https://doi.org/10.1055/s-0042-104495>
- 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>

- Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice*, 24(2), 190–195. <https://doi.org/10.1037/0735-7028.24.2.190>
- World Health Organization (1990). *Composite international diagnostic interview (CIDI)*. Geneva: Author
- Zilcha-Mano, S. (2017). Is the alliance really therapeutic? Revisiting this question in light of recent methodological advances. *The American Psychologist*, 72(4), 311–325. <https://doi.org/10.1037/a0040435>
- Zilcha-Mano, S., Dinger, U., McCarthy, K. S., & Barber, J. P. (2014). Does alliance predict symptoms throughout treatment, or is it the other way around? *Journal of Consulting and Clinical Psychology*, 82(6), 931–935. <https://doi.org/10.1037/a0035141>.

Appendix

Appendix A: R-Skript

Appendix A ist im R-Skript-Format geschrieben. Das bedeutet, dass Kommentare vorweg mit einer Raute gekennzeichnet sind und R-Code ohne Raute. Das R-Skript kann in R kopiert und komplett ausgeführt werden.

```
#Simulation von zwei korrelierten Datensätzen (Prä-Post-Vergleich für zwei Gruppen, die sich im Postwert unterscheiden)
```

```
#Packages
```

```
#für mvrnorm
```

```
library("MASS")
```

```
#für describe
```

```
library("psych")
```

```
#für Cohens d
```

```
library("lsr")
```

```
#Sample simulieren
```

```
#Varianz Kovarianzmatrix
```

```
#Varianz = 20
```

```
#Kovarianz = 10, entspricht Korrelation von  $r = 0.5$ 
```

```
corMat <- matrix(c(20, 10,  
                  10, 20), nrow = 2)
```

```
#Simulation des Datensatzes mit Seed zur besseren Replizierbarkeit
```

```
#Seeds sind optional, die Simulation funktioniert auch ohne diese
```

```

set.seed(210790)

contr <- mvrnorm(n = 100, mu = c(50, 55), Sigma = corMat)

set.seed(010386)

expe <- mvrnorm(n = 100, mu = c(50, 59), Sigma = corMat)

#umwandeln in data.frame zur leichteren Bearbeitung später

contr <- data.frame(Prae = contr[,1], Post = contr[,2])

expe <- data.frame(Prae = expe[,1], Post = expe[,2])

#Prüfen der simulierten Daten

#Mittelwertsvergleich Prae-Post der beiden Stichproben

t.test(contr[,2], contr[,1], paired = T)

t.test(expe[,2], expe[,1], paired = T)

#Mittelwertsvergleich zum Postzeitpunkt

t.test(expe[,2], contr[,2])

#Korrelation

cor.test(contr[,1], contr[,2])

cor.test(expe[,1], expe[,2])

#Effektgröße Prä-Post Control Group

cohensD(contr[,2], contr[,1], method = "paired")

#Effektgröße Prä-Post Experimental Group

cohensD(expe[,2], expe[,1], method = "paired")

```

```
#Effektgröße Post Experimental Group vs. Control Group
```

```
cohensD(expe[,2], contr[,2], method = "pooled")
```

```
#Grafische und deskriptive Beschreibung
```

```
#Übersicht über den Datensatz
```

```
par(mfrow = c(2,2))
```

```
hist(contr[,1], main = "Prae-Wert Control Group")
```

```
describe(contr[,1])
```

```
hist(contr[,2], main = "Post-Wert Control Group")
```

```
describe(contr[,2])
```

```
hist(expe[,1], main = "Prae-Wert Experimental Group")
```

```
describe(expe[,1])
```

```
hist(expe[,2], main = "Post-Wert Experimental Group")
```

```
describe(expe[,2])
```

```
par(mfrow = c(1,1))
```

```
#Wie haben sich einzelne Personen verändert?
```

```
#Festlegung von RCI und Cut-off zur gesunden Stichprobe
```

```
SE = sd(contr[,1]) * sqrt(1 - cor(contr[,1], contr[,2]))
```

```
Sdiff = sqrt(2*(SE^2))
```

```
RCI = Sdiff * 1.96
```

```
rm(SE, Sdiff)
```

```
norm = mean(contr[,1]) + 2*sd(contr[,1])
```

```
#Klassifizierung der einzelnen Personen nach Jacobson und Truax (1991)
```

```
#zunächst Berechnung des Differenzwertes für jede Person, der mit dem RCI verglichen wird
```

```
contr$diff <- contr[,2] - contr[,1]
```

```
contr$kat <- NA
```

```
#Kategorienbildung
```

```
contr$kat[contr$diff >= RCI & contr$Post < norm] <- "Reliable Verbesserung"
```

```
contr$kat[contr$diff >= RCI & contr$Prae < norm & contr$Post > norm] <- "Klinisch  
signifikante Verbesserung"
```

```
contr$kat[contr$diff < RCI] <- "Keine Veränderung"
```

```
contr$kat[contr$diff <= -RCI] <- "Reliable Verschlechterung"
```

```
#Umwandlung in Faktor zur leichteren Darstellung
```

```
contr$katFak <- factor(contr$kat, levels = c("Reliable Verschlechterung",
```

```
      "Keine Veränderung",
```

```
      "Reliable Verbesserung",
```

```
      "Klinisch signifikante Verbesserung"))
```

```
table(contr$katFak)
```

```
#Klassifizierung der einzelnen Personen nach Jacobson und Truax (1991)
```

```
#zunächst Berechnung des Differenzwertes für jede Person, der mit dem RCI verglichen wird
```

```
expe$diff <- expe[,2] - expe[,1]
```

```
expe$kat <- NA
```

```
#Kategorienbildung
```

```
expe$kat[expe$diff >= RCI & expe$Post < norm] <- "Reliable Verbesserung"
```

```
expe$kat[expe$diff >= RCI & expe$Prae < norm & expe$Post > norm] <- "Klinisch  
signifikante Verbesserung"
```

```
expe$kat[expe$diff < RCI] <- "Keine Veränderung"
```

```
expe$kat[expe$diff <= -RCI] <- "Reliable Verschlechterung"
```

```
#Umwandlung in Faktor zur leichteren Darstellung
```

```
expe$katFak <- factor(expe$kat, levels = c("Reliable Verschlechterung",
```

```
      "Keine Veränderung",
```

```
      "Reliable Verbesserung",
```

```
      "Klinisch signifikante Verbesserung"))
```

```
table(expe$katFak)
```

Appendix B: Lebenslauf

Lebenslauf

Name Sebastian Gmeinwieser
Geburtstag 21.07.1990
Geburtsort Fritzlär,
Adresse Groner Landstraße 55B, App.67

Akademische Ausbildung und Qualifikation

Seit 11/2016 Promotion im Fach Psychologie, Georg-August-Universität Göttingen, Georg-Elias-Müller Institut für Psychologie, Abteilung für Klinische Psychologie und Psychotherapie, Therapie- und Beratungszentrum
09/2016 Master of Science in Psychologie
09/2014 Bachelor of Science in Psychologie
10/2011-09/2016 Studium der Psychologie and der Georg-August-Universität Göttingen, Georg-Elias-Müller Institut für Psychologie
06/2010 Allgemeine Hochschulreife
07/2001-06/2010 Ursulinenschule Fritzlär

Berufliche Weiterbildung

Seit 2016 Postgraduale Ausbildung Psychologische Psychotherapie (Verhaltenstherapie), Georg-August-Universität Göttingen, Georg-Elias-Müller Institut für Psychologie, Therapie- und Beratungszentrum

Wissenschaftliche Tätigkeit

Seit 2016 Wissenschaftlicher Mitarbeiter, Georg-August-Universität Göttingen, Georg-Elias-Müller Institut für Psychologie, Abteilung für Klinische Psychologie und Psychotherapie, Therapie- und Beratungszentrum

Sonstige Berufserfahrung

10/2013-09/2016 Wissenschaftliche Hilfskraft, Georg-August-Universität Göttingen, Georg-Elias-Müller Institut für Psychologie, Abteilung für Klinische

- Psychologie und Psychotherapie, Therapie und Beratungszentrum, im Bereich Diagnostik
- 10/2012-09/2013 Wissenschaftliche Hilfskraft, Georg-August-Universität Göttingen, Georg-Elias-Müller Institut für Psychologie, Tutorium für Quantitative Methoden I & II
- 07/2010-12/2010 Grundwehrdienst im 3. Führungsunterstützungsbattailon 362 und der Stabskompanie der Luftbeweglichen Brigade 1

Publikationen

- Gmeinwieser, S.**, Hagmayer, Y., Pieh, C., & Probst, T. (2019). General change mechanisms in the early treatment phase and their associations with the outcome of cognitive behavioural therapy in patients with different levels of motivational incongruence. *Clinical Psychology & Psychotherapy*. Advance online publication. <https://doi.org/10.1002/cpp.2381>
- Gmeinwieser, S.**, Kuhlencord, M., Ruhl, U., Hagmayer, Y., & Probst, T. (2019). Early developments in general change mechanisms predict reliable improvement in addition to early symptom trajectories in cognitive behavioral therapy. *Psychotherapy Research*. Advance online publication. <https://doi.org/10.1080/10503307.2019.1609709>
- Gmeinwieser, S.**, Schneider, K. S., Bardo, M., Brockmeyer, T., Hagmayer, Y. (in press). Risk for psychotherapy drop-out in survival analysis: The influence of general change mechanisms and symptom severity. *Journal of Counseling Psychology*.

Poster und Kongressbeiträge

- Gmeinwieser, S.** & Hagmayer, Y. (2019, Juli). *Does Feedback in Routine Psychotherapy Improve Outcomes? Preliminary Results of a Doctoral Thesis*. Poster präsentiert auf dem 9th World Congress of Behavioral & Cognitive Therapies, Berlin, GER.
- Gmeinwieser, S.**, Kuhlencord, M., Ruhl, U., Hagmayer, Y. & Probst, T. (2018, März). *Therapieerfolg und Therapieabbruch im ambulanten psychotherapeutischen Setting*. Referat präsentiert auf dem 30. DGVT-Kongress für Klinische Psychologie, Psychotherapie und Beratung, Berlin, GER.
- Mitglied des Planungskomitees des Symposiums *Einflussfaktoren auf den Therapieerfolg* auf dem 30. DGVT-Kongress für Klinische Psychologie, Psychotherapie und Beratung, Berlin, GER. (März, 2018)