

Geeignete Messzeitpunkte in der Psychotherapieevaluation: Ein Vorschlag

Edgar Geissner^a Petra Maria Ivert^b Manfred Schmitt^c

^aDepartment Psychologie, Ludwig-Maximilians-Universität München, München, Deutschland;

^bPsychotherapeutische Praxis, Rosenheim, Deutschland; ^cFachbereich Psychologie, Universität Koblenz-Landau, Landau, Deutschland

Schlüsselwörter

Therapieverlauf mit und ohne spezifische Intervention · Hoffnung auf Erfolg · Kontexteffekt · Regression zur Mitte · Mehrzeitpunktmessung

Zusammenfassung

Untersucht man Therapieverläufe nicht nur zwischen unmittelbarem Beginn und Beendigung, können Fortschritte bereits zwischen Anmeldung und Beginn, noch ohne Intervention, aufgezeigt werden. Diese t_0 - t_1 -Veränderungen werden anhand eigener Befunde diskutiert: Ist der Vergleich Beginn (t_1) – Ende der Therapie (t_2) bzw. Katamnese (t_3) sinnvoll, oder erfordert die angemessene Erfolgsbeurteilung nicht gerade Anmeldezeitpunkt t_0 ? Wir plädieren für t_0 . (1) Anmeldung (t_0) impliziert eine ausgeprägte, realistische Symptomschilderung, (2) positive Erwartungen zur erfolgreichen Behandlung verbessern Messwerte zu t_1 (Aufnahme), (3) Aufnahmezeitpunkt t_1 ist durch Sicherheitssignale der Klinik/der Behandler beeinflusst, Patientenratings – z.B. bei Angst – sind geringer (Kontextfaktoren), (4) Regression zur Mitte relativiert die Aussagekraft eines Einzelmesszeitpunkts: Hohe Messwerte dort fallen zu einem zweiten Zeitpunkt geringer aus. Um dies zu kontrollieren, müssen zu t_0 (in geringem Abstand) 2 Messungen durchgeführt werden, nämlich t_{01} und t_{02} . In der Praxis dient das Intervall t_{02} - t_2 zur geeigneten Erfolgsbeurteilung, aber auch t_{02} - t_3 sollte, wenn möglich, Auswertungsroutine sein. Einzelfallbeurteilungen erfordern die Berücksichtigung kritischer Differenzen, um messfehlerbereinigte Unterschiede zu sichern (Beispielerläuterung). Die Möglichkeit erweiterter Auswertung mittels Ausschluß nichtpathologischer Werte, der vereinfachten Auswertung mittels Cutoff-Bestimmung

und einige Praxishinweise komplettieren die Arbeit. Unspezifische t_0 - t_1 -Veränderungen bedeuten keine Relativierung der Ergebnisse aktiver Intervention (t_1 - t_2), sind vielmehr Bestandteil der Gesamttherapie.

© 2021 S. Karger AG, Basel

Proposal of Measurement Occasions for Unbiased Evaluation of Psychotherapy

Keywords

Course of therapy with and without specific intervention · Hope for success · Context effect · Regression towards the mean · Repeated symptom assessment

Abstract

Studies complementing the assessment of symptoms right before (t_1), right after therapy (t_2) and at follow-up (t_3) with an assessment of symptoms preceding the waiting period without intervention (t_0) have revealed substantial t_0 - t_1 changes. We discuss this phenomenon based on own data and address the following questions: does it make sense to compare symptoms at the beginning of therapy (t_1) with symptoms at the end of therapy (t_2) or at follow-up (t_3)? Or does it make more sense to use t_0 instead of t_1 ? We argue for the latter alternative based on the following reasons. (1) Symptom descriptions at t_0 are realistic. (2) Expecting therapy success mitigates symptom descriptions at t_1 . (3) Security signals emitted from the therapy context also mitigate symptoms, especially anxiety, at t_1 . (4) Regression toward the mean reduc-

es the validity of single occasion assessments. Controlling for regression requires two occasions of measurement with a short time interval at t_0 (t_{01} and t_{02}). It follows from this reasoning that therapy success should be evaluated using the $t_{02}-t_2$ and $t_{02}-t_3$ intervals. Single case evaluations require reliable critical differences. This will be illustrated using a concrete example. The validity of treatment evaluation can be increased via the elimination of non-pathological symptom scores. A simplified calculation of cut-off scores can facilitate applied treatment evaluation. Unspecific t_0-t_1 changes do not challenge therapy effects according to t_1-t_2 changes. Rather, they are part of the whole therapy process.

© 2021 S. Karger AG, Basel

Vorbemerkung

Aus einer klinischen Studie zu differentiellen Effekten motivationaler Faktoren in der Angstbehandlung [nicht-experimentelles, naturalistisches Design; Geissner und Ivert, 2019a] standen für eine Teilstichprobe von $N = 67$ nicht nur die Aufnahme-, Entlass- und Katamnese-daten zur Verfügung, sondern auch die Daten zum *Anmeldezeitpunkt* etwa 8 Wochen vor Klinikanreise, die dort nicht Gegenstand der Analyse waren. Die Angstaussprägung über diese 4 Messzeitpunkte (im folgenden MZP) hinweg, nämlich $t_0 =$ Anmeldung, $t_1 =$ Aufnahme, $t_2 =$ Entlassung und $t_3 =$ 6-Monats-Follow-up dient als Basis der hier vorgestellten Überlegungen zur geeigneten Auswertung von Psychotherapie.

Patienten, Fragebögen, Prozedere

Die Patienten (45 w., 22 m.; Alter $M = 41$ Jahre, $SD = 11$ Jahre) erfüllten die ICD-Diagnosekriterien F 40.00, F 40.01 und F 41.00 gemäß Hiller et al. [1995], waren seit durchschnittlich 8 Jahren erkrankt, hatten zurückliegend in der Regel 1 ambulante Psychotherapie (in Einzelfällen mehr) und wurden wegen Fortbestehens der Angstproblematik für einen stationären Aufenthalt angemeldet. Neben einem allgemeinen Fragenkatalog im Rahmen der Klinikroutine bearbeiteten sie 5 Angstfragebögen: Beck Anxiety Inventory BAI [Margraf und Ehlers, 2007; Geissner und Hütteroth, 2018]; Anxiety Cognitions Questionnaire, Body Sensations Questionnaire, Mobility Index-allein MI-A und Mobility Index-begleitet MI-B [alle Ehlers und Margraf, 2001]. Zur informierten Einwilligung erhielten die Patienten ein Informationsblatt mit der Bitte zur Teilnahme, "um den Verlauf der Angsterkrankung sowie den Erfolg der Therapie dokumentieren zu können" und erteilten ihr Einverständnis auf einem gesonderten Bogen.

Fragestellung

Vorab werden die Ergebnisse der 5 oben genannten Angstindikatoren dargestellt (Tab. 1). Aus diesen geht hervor, dass für die eigentliche Interventionsphase t_1-t_2 [6- bis 7-wöchiger Aufenthalt; Angstbehandlung gemäß Geissner und Kindermann, 2010; Wambach und Rief, 2012; Margraf und Schneider, 2013] Effektstärken von $d = 0,54-0,66$ festgestellt werden konnten, unter Hinzunahme der Katamnese-phase, also t_1-t_3 , Effektstärken von $d = 0,56-0,79$.

Wertung: Zwar sind für naturalistische Settings im Vergleich zu experimentellen Studien geringere Wirksamkeiten zu erwarten [Schindler und Hiller, 2010; Schindler et al., 2011; Lutz et al., 2019, S. 23; Rief, 2020], zumal die Problematik aufgrund des wenig erfolgreichen ambulanten Therapieversuchs als eher hartnäckig zu bezeichnen war. Dennoch sind die Effektstärken für t_1-t_2 bzw. t_1-t_3 als nicht übermäßig hoch einzustufen.

Dies verdeutlichen zum Vergleich auch die Effektstärken t_0-t_1 , die immerhin zwischen $d = 0,31$ und $0,46$ lagen (vgl. Tab. 1) und gemäß den Cohenschen Interpretationsempfehlungen als "mittelhoch" gelten. Diese Effekte resultierten ohne spezifische Intervention.

Hier schließt sich die Fragestellung unseres Diskussionsbeitrags an, was geeignete Markierpunkte sind, die der Psychotherapie gerecht werden und die den Erfolg durchgeführter Bemühungen gut dokumentieren. Müsste nicht der *Anmeldezeitpunkt* für eine Psychotherapie als besser geeigneter Anker für die Beobachtung der Problematik über die Zeit hinweg fungieren? Angesichts des Umstands, dass die Effektstärken über den gesamten Analysezeitraum t_0-t_3 mit bis zu $d = 1,05$ ein deutlich besseres Bild ergaben (vgl. Tab. 1), könnte Data-squeezing zugunsten schöner Resultate dageengehalten werden. Aber es gibt auch gute Gründe pro t_0 , die wir im Folgenden zur Diskussion stellen wollen: erwartungsbezogene Gesichtspunkte, inhaltliche und spezifisch störungsbezogene Kontextfaktoren sowie diagnostisch-messmethodische Überlegungen.

Explizite Problemdarstellung zu t_0 (Anmeldung)

Wird eine gegebene Ausgangslage, also im vorliegenden Fall eine ausgeprägte Angstproblematik mit deutlichen Einschränkungen im alltäglichen Lebensvollzug (teilweise massive körperliche Symptome; im Privaten und im Beruflichen behinderndes Vermeidungsverhalten) zu prägnanten Beantwortungen in Fragebögen führen, die auf dieses Problemgeschehen spezifisch fokussiert sind? Eine gewisse Neigung zur schonungslosen Darstellung kann angenommen werden, auch ohne dass eine Aggravationstendenz zu unterstellen ist. Immerhin

Table 1. Angstausprägung im Verlauf – vor, während und nach der Therapie (VA für 5 Fragebogenmaße; $N = 67$ Patienten mit Agoraphobie/Panikstörung)

	t_0	t_1	t_2	t_3	F/p	Sig. Kontraste	$d/t_{0/1}$	$d/t_{1/2}$	$d/t_{1/3}$	$d/t_{0/3}$
BAI	35.91 (11.19)	30.89 (12.88)	20.48 (11.74)	22.69 (13.81)	33.80 0.000	a, b, c, d	0.45	0.58	0.69	1.05
ACQ	2.38 (0.67)	2.13 (0.65)	1.78 (0.61)	1.80 (0.68)	31.13 0.000	a, b, c, d	0.37	0.57	0.68	0.94
BSQ	2.93 (0.71)	2.70 (0.70)	2.22 (0.74)	2.27 (0.81)	47.33 0.000	a, b, c, d	0.31	0.66	0.79	1.00
MI-A	3.35 (1.06)	2.86 (1.08)	2.19 (0.88)	2.40 (1.15)	38.43 0.000	a, b, c, d	0.46	0.61	0.56	0.85
MI-B	2.52 (0.94)	2.18 (0.86)	1.70 (0.62)	1.80 (0.92)	24.17 0.000	a, b, c, d	0.36	0.54	0.57	0.67

VA, Varianzanalyse; BAI, Beck Anxiety Inventory; ACQ, Anxiety Cognitions Questionnaire; BSQ, Body Sensations Questionnaire; MI, Mobilitätsinventar (A, allein; B, in Begleitung); t_0 , Anmeldung; t_1 , Aufnahme; t_2 , Entlassung; t_3 , 6-Monats-Follow-up. Signifikante Kontraste nach VAs: $a = t_0/t_1$, $b = t_1/t_2$, $c = t_1/t_3$, $d = t_0/t_3$, $d = \text{Effektstärke } [(x_1 - x_2)/SD]$. Spalten t_0 bis t_3 = Mittelwerte und (in Klammern) Standardabweichungen. Range: BAI min. = 0, max. = 63; ACQ bis MI-B min. = 1, max. = 5 (BAI = Skalensummenwert; ACQ bis MI = Skalensumme/Itemanzahl).

ist ein Entschluss zur – hier stationären – Behandlung gefallen, verbunden mit dem Wunsch, den künftigen Behandlern gegenüber die Lage offen als dringend zu schildern. Wartezeiten sind sowohl ambulant als auch stationär gang und gäbe, und ein Bericht in klaren Worten verbunden mit entsprechenden Fragebogenangaben mag aus der Sicht der Betroffenen unterstreichen, wie stark die Änderungsmotivation ist.

Hoffnung auf Erfolg zwischen t_0 (Anmeldung) und t_1 (Aufnahme)

Auch positive Erwartungen, dass mit dem konkreten Beginn der Behandlung der erste Schritt in Richtung Therapieerfolg getan sei, können angenommen werden, zumindest in all jenen Fällen, in welchen Patienten eigenmotiviert sind und nicht geschickt wurden. Hier ist zu vermuten, dass in Fragebogenangaben zu t_1 eine erfolgsbezogene positive Erwartung mit eingeht – eine zwischenzeitliche Behandlungszusage war ja erfolgt – und die Werte im Vergleich zu t_0 weniger stark ausfallen könnten (siehe Tab. 1). Erwartungen in diesem Sinne stehen im Einklang mit dem Hoffnung-auf-Erfolg-Konzept der Motivationspsychologie [Rheinberg, 2018].

Komplementär zur Hoffnung auf Erfolg hat die Motivationspsychologie aber auch Furcht vor Misserfolg beschrieben bzw. ein individuell resultierendes Verhältnis aus Erfolgshoffnung und Misserfolgsschmerz. So können Skepsis (hinsichtlich des Therapieansatzes, der Person des Therapeuten, wegen früherer Misserfolge) oder eine depressiv getönte Stimmung dazu beitragen, dass Er-

folgserwartungen reduziert sind. Gleichwohl wird in der Therapieforschung die Reduktion einer Störungsausprägung ohne spezifische Therapie häufig mit Erwartungseffekten erklärt [Rief et al., 2009; Lutz et al., 2019; Rief, 2019, 2020; Bingel und Kersting, 2020].

Sicherheits-/Erleichterungssignale zu t_1 (Aufnahme)

In eine Behandlung bzw. in eine Klinik aufgenommen zu sein, ist mit einer Reihe von Sicherheitssignalen verbunden. Solche Signale können von Ärzten, Psychologen, Schwestern/Pflegern sowie der Verfügbarkeit von Medikamenten ausgehen und über selbstberuhigende Kognitionen (“Endlich bin ich in guten Händen”, “Im Zweifelsfall kann mir nichts passieren. Wenn doch, werde ich gerettet”) ein Gefühl von Erleichterung bewirken. Beruhigende Effekte von Sicherheitssignalen sind bei Angstpatienten – aber nicht nur bei diesen – zu erwarten. Sie schlagen sich in einer Reduktion der subjektiv empfundenen und berichteten Symptomausprägung zu t_1 (Aufnahme in die Klinik) im Vergleich zu t_0 (Anmeldung) nieder. Zwar geben Fragebögen häufig die Instruktion, zu antworten “wie es typisch ist” oder den Zeitrahmen “auf die letzte Woche” zu beziehen, aber es ist nicht sicher, ob sich Patienten beim Ausfüllen daran halten oder nicht doch den aktuellen Zustand angeben. Die Beurteilung der Symptomstärke kann zudem durch die Prognose geprägt sein, dass “es nun besser werden wird.” Zum Zeitpunkt t_1 wäre dann die empfundene und angegebene Stärke der Angst durch ein vorweggenommenes Behandlungsergebnis beeinflusst.

Kontexteffekte bei der konkreten Fragebogenbearbeitung zu t_1 (Aufnahme)

Des Weiteren unterliegt die Itembearbeitung auch Kontextfaktoren: So kann beispielsweise das Antwortverhalten bei einem Angstfragebogen wie dem Mobilitätsinventar [MI; Ehlers und Margraf, 2001], welches die angstbedingten situativen Einschränkungen im Alltag erfasst, durch die Aufnahme in die Klinik beeinflusst werden. Denn der Eintritt in den Klinikkontext könnte zu einer Reduktion der empfundenen Einschränkungen führen. Insbesondere Variante "MI-B" ("B" steht für "in Begleitung") dürfte für einen solchen Kontexteffekt empfänglich sein, da die empfundene Mobilität im Vergleich zum Alltag größer sein sollte, wenn im Behandlungssetting eine Begleitung – sei es durch Mitarbeiter, sei es auch nur durch andere Patienten – gewährleistet ist. Entsprechendes gilt für die Bedrohlichkeitseinschätzung von körperbezogenen Angstkomponenten [Body Sensations Questionnaire, Ehlers und Margraf, 2001] innerhalb des sicheren Klinikkontextes, oder aber ohne diesen.

Zwischenfazit: Geeigneter Anker für Eingangsmessungen

Die geschilderten Überlegungen sprechen dafür, den MZP t_0 (Anmeldung) dem MZP t_1 (Aufnahme) für die Schätzung des Therapieeffekts vorzuziehen, da zu t_1 die geschilderten Effekte symptomreduzierend wirken. Ungeachtet der Gefahr einer möglicherweise übertriebenen Schilderung der Symptome zu t_0 würde der Therapieverlauf durch Einbezug des Anmeldezeitpunkts (t_0) angemessener abgebildet, da Erwartungs- und Kontexteffekte durchaus als Bestandteile der Therapie angesehen werden können. Die Differenz von t_0 zu t_2 (Ende der Therapie) oder zu t_3 Nachbefragung würde größer.

Auf ein zusätzliches Problem wird im Folgenden eingegangen.

Messwertverlauf bei wiederholter Messung: Der Regression-zur-Mitte-Effekt

Ein in diesem Zusammenhang wichtiger weiterer Gesichtspunkt ergibt sich aus der Regression zur Mitte. Diese tritt grundsätzlich auf, wenn Messwerte keine perfekte Korrelation zwischen zwei Zeitpunkten aufweisen und zum ersten Zeitpunkt eine Stichprobe mit extremen Messwerten vorliegt [Schmitt und Gerstenberg, 2014]. In klinischen Kontexten ist dies der Regelfall, da die Anmeldung zur Behandlung ein deutliches Problemausmaß voraussetzt. Bei gleicher Variabilität der Messwerte zu zwei Zeitpunkten impliziert die ein-

geschränkte Korrelation, dass die Messwerte intraindividuell veränderlich sind und die Schwankungen sich überdies interindividuell unterscheiden. Ursache für dieses Muster sind alle unsystematischen (Messfehler) und alle systematischen Faktoren (wahre Merkmalsveränderung), die Messwerte einer Person zu einem ersten Zeitpunkt anders beeinflussen als zum zweiten. Bei klinisch relevanten Variablen wie Angst kann das bedeuten, dass der Leidensdruck bei einer Person zum ersten Zeitpunkt größer ist als zum zweiten, bei einer anderen Person hingegen nicht (im selben Ausmaß). Es kann auch bedeuten, dass eine Person die Symptome zum ersten Zeitpunkt drastischer beschreibt als eine andere Person, zum zweiten Zeitpunkt aber nicht. Es kann auch bedeuten, dass bei zwei Personen zum ersten Zeitpunkt kein Hoffnungseffekt auf die Messwerte einwirkt (skeptische Haltung nach bisher nur geringer Besserung), dieser Effekt beim zweiten Zeitpunkt aber eintritt (Überraschung durch einen "positiven Eindruck" der Therapie), jedoch bei beiden Personen in unterschiedlicher Stärke.

Ein früher Vorschlag zum Umgang mit Regression-zur-Mitte bei Extremgruppen besteht darin, ein Merkmal nicht nur zweimal, sondern dreimal zu messen [Nesselrode et al., 1980]. Dahinter steckt die Überlegung, dass die zum ersten Zeitpunkt extremen Messwerte aufgrund der Regression zur Mitte beim zweiten Zeitpunkt weniger extrem ausfallen und der Unterschied zwischen dem zweiten und dritten Zeitpunkt keinem oder einem schwächeren Regressionseffekt unterliegt.

Angewendet auf unsere Thematik würde dies erfordern, Angst bereits zu t_0 – nämlich beim Entschluss zur Therapie nebst Anmeldung – mindestens zweimal zu erheben, also zu t_{01} und einige Tage später zu t_{02} . Die fairere Beurteilung der Wirksamkeit der Intervention geschähe zum einen, wie oben besprochen, um die geschilderten Erwartungs- und Kontexteffekte zu t_1 zu vermeiden. Und zum anderen würde die 2-malige Erhebung zu t_0 den Regression-zur-Mitte-Effekt reduzieren. Der relevante Anker für die Vergleiche mit den späteren MZP wäre dann t_{02} .

Schlussfolgerungen, Vorgehen, Empfehlungen

In der Nutzung von Psychotherapiedaten zu Evaluationszwecken plädieren wir aus den genannten Überlegungen für 4-maliges bzw. unter Einbezug der – im Grunde generell sinnvollen Nachbefragung – für 5-maliges Messen.

Die Regression zur Mitte würde durch 2-maliges, in relativ kurzen Abständen erfolgreiches Messen bei Anmeldung handhabbar gemacht, um eventuell zu drastische Item-Scorings abzumildern.

Die Messung zu t_1 unmittelbar bei Beginn der Behandlung bzw. Aufnahme in eine Klinik kann, wenn mit den Befunden bei Anmeldung (t_{02}) verglichen, als Aussage zu Erwartungseffekten bzw. Hoffnung auf Erfolg verwendet werden. Sie spiegelt auch, etwa bei Angstpatienten, die Kontextvariable "Sicherheitsgefühl" wider.

Der relevante Bezugsrahmen zur Beurteilung der eigenen therapeutischen Bemühungen ist der Vergleich zwischen t_{02} und dem Therapieende t_2 bzw. zur Feststellung der Erfolgserhaltung (Katamnese) t_3 .

Je gezielter auf die behandelte Störung bezogen (in unserem Fall 5 Indikatoren der Angst), desto besser ist der spezifische Verlauf darstellbar, wobei jedoch meist 1 bis 2 störungsbezogene Verfahren genügen dürften. Skalenübergreifende Globalwerte hingegen (z.B. über alle SCL-/BSI-Skalen hinweg) generieren wegen vieler im Einzelnen irrelevanter Fragen Rauschen und trüben das Ergebnisbild.

Therapieevaluationen werden sehr häufig anhand aggregierter Daten vorgenommen. Patienten werden dort nach Störungsgruppen zusammengefasst und dokumentiert, wobei dies meist für größere Zeiteinheiten, z.B. im Zusammenhang mit der Erstellung von Jahresberichten in Form von Mittelwertverläufen nebst Standardabweichungen, varianzanalytischen Vergleichen und Effektstärkeangaben, dargestellt ist.

Andererseits gilt die Empfehlung zur Anzahl therapieangemessener Messzeitpunkte bei der *Einzelfallauswertung* ebenfalls. Dort sollte zur Erfolgsbeurteilung die kritische Differenz bzw. der Reliable Change Index herangezogen werden, da der Messfehler, geschätzt über die Reliabilität des Fragebogens, eine stärkere Rolle spielt. Fisseni [2004, S. 73] und Schmitt und Gerstenberg [2014] geben Hinweise zur kritischen Differenz, Lutz et al. [2019, S. 21] zum Reliable Change Index. Die Differenz im Therapieverlauf muss umso größer sein, je geringer die Reliabilität ist, damit eine tatsächliche Reduktion der Problematik auch als solche zu werten ist. Für die Schätzung der Reliabilität gibt es verschiedene Methoden [Schmitt und Gerstenberg, 2014]. Die eigentlich naheliegende Retest-Korrelation sollte man aber nur verwenden, wenn im betrachteten Zeitraum wahre Merkmalsveränderungen ausgeschlossen sind. Andernfalls sind Unreliabilität und wahre Veränderungen konfundiert, Reliabilität würde unterschätzt, was gerade in klinischen Kontexten relevant wäre. Muss mit wahren Veränderungen zwischen den MZP gerechnet werden, sind die interne Konsistenz (α) oder die Split-Half-Korrelation als Reliabilitätsschätzer zu verwenden.

Am Beispiel des BAI [Geissner und Hütteroth, 2018] ist dies wie folgt zu veranschaulichen: Wir benötigen die Reliabilität, die hier mit $\alpha = 0,92$ recht gut ist, die Standardabweichung (12 Punkte), errechnen den Standardmessfehler (4,8 Punkte) und legen eine Irrtumswahr-

scheinlichkeit fest (z.B. 5%, daher Multiplikation mit dem z-Wert von 1,96). Man erhält so eine kritische Differenz von 9 Punkten. Diese muss mindestens erreicht sein, erst ab da oder höher dürfte von einer messfehlerbereinigten, tatsächlich interpretierbaren Angstreduktion ausgegangen werden. Bezogen auf die BAI-Angaben in Tabelle 1 – hier einmal angenommen, es läge das Messergebnis für einen Einzelfall vor – würde die kritische Differenz für t_1/t_2 mit 10 Punkten gerade überschritten, für t_1/t_3 jedoch nicht. Deutlich ausserhalb der kritischen Differenz – und somit ein klarer Befund – wäre jedoch erst der Einbezug des t_0 -Wertes als Ausgangspunkt. Dieses Beispiel mag demonstrieren, dass Vorher-Nachher-Unterschiede auf Einzelfallebene relativ deutlich sein müssen, um sie als Erfolge werten zu können.

Ergänzender Gesichtspunkt Ausschluss nichtpathologischer Werte

Zur fairen Beurteilung der Wirksamkeit durchgeführter Therapien wird über die Wahl angemessener MZP hinaus auch der Ausschluss nichtpathologischer Werte bei der Eingangsmessung vorgeschlagen. Dies gilt vor allem für praxisangewandte Fragestellungen [Schindler und Hiller, 2010; Hiller und Schindler, 2011; Schindler et al., 2011]. Ausgesprochen niedrige Symptomausprägungen zu Anfang lassen sich durch Therapie nicht weiter reduzieren. Aus welchen Gründen manche Patienten bei Eingangsuntersuchungen auf einem bestimmten Fragebogen niedrig scoren, auch wenn sie in Exploration und Anamnese deutlichen Leidensdruck angeben, bleibt dabei offen. Diese Fälle kommen in der klinischen Praxis aber vereinzelt vor. Hier wäre der Verlauf der entsprechenden Fragebogenwerte ein ungünstiger Indikator für die Auswertung therapeutischer Bemühungen. Geissner und Ivert [2019b] konnten zum Beispiel zeigen, dass die Effektstärken substantiell höher ausfielen, wenn jene Patienten aus der Auswertung herausgenommen wurden, die gemäß Fragebogenmanual lediglich Werte wie die nichtklinische Vergleichsgruppe erzielten, so dass im Verlauf keine oder nur eine unwesentliche Reduktion möglich war. Lagen Vorher-Nachher-Effektstärken zunächst bei $d = 0,56-0,80$, so stiegen sie nach Ausschluss nichtpathologischer Eingangswerte auf $d = 0,83-0,98$.

Cut-off-Werte

Die nichtpathologischen Werte von Messinstrumenten müssen dem Anwender bekannt sein, oft (aber nicht immer) ersichtlich aus Manualangaben zu Kontrollgruppen bzw. aus Normtabellen. Dies ist nicht trivial, denn nichtklinische Vergleichspersonen können auf

Messinstrumenten zu bestimmten Störungen ebenfalls Werte oberhalb eines numerischen Minimums erzielen, z.B. bei Angst, Stress, Ärger, Beschwerden. Es wäre daher nicht korrekt, der Einfachheit halber z.B. "Null" als Vergleichswert anzusetzen. Dies gilt auch für die Verwendung von Cut-off-Werten [Lutz et al., 2019, S.20]. Sie stellen eine vereinfachte Möglichkeit der Erfolgsbeurteilung dar und sollen hier ebenfalls kurz erwähnt werden. Liegen Referenzangaben sowohl zu klinischen als auch zu nicht-klinischen Gruppen vor, so kann über die Mittelwerte beider Gruppen wiederum ein Mittelwert gebildet werden. Dieser Wert dient dann als Grenze bei der Beurteilung eines jeden Einzelfalls: Ist überhaupt eine klinische Ausprägung zu t_0 gegeben? Und wenn ja, wurde sodann nach Therapieende eine Reduktion in den nichtklinischen Bereich zu erzielt?

Spanne t_0 - t_1 als Wartekontrollbedingung ungünstig

Die Wartephase vor Beginn der aktiven Intervention ist, wie oben dargestellt, veränderungsrelevant. Rief et al. [2009] sprachen hier zunächst von Placeboeffekten, später präzisierend von "erwartungsfokussierten Effekten" [Rief, 2019, 2020]. Wird eine gezielte Intervention zugesagt, wenn auch zeitlich etwas verzögert, wird dies das Antwortverhalten in Fragebögen verändern. Aussagen für die t_1 - t_2 -Differenz stellen dort nur Nettoeffekte der aktiven Intervention dar – im Vergleich zu Erwartungs- und Kontexteffekten für die t_0 - t_1 -Differenz. Dies würde klinischer Anwendung im Feld bzw. in naturalistischen Untersuchungsdesigns jedoch weniger gerecht, der Erfolg therapeutischer Maßnahmen würde unterschätzt.

Resümee und Praxis

Zeigt die Messung von Therapieverläufen über mehrere MZP bereits Reduktionen für t_0 - t_1 an, bedeutet dies also – so unser Statement – keine Minderung des Wertes der aktiven Intervention, keine Schwäche. Die faire Erfolgsbeurteilung schließt die Verwendung der Werte bei Anmeldung und den unspezifischen Effekt zwischen Anmeldung und Therapiebeginn als legitimen Teil des Gesamtverlaufs mit ein. Für das "Routine Outcome Monitoring" [Lutz et al., 2019, S.37, 68–70] sollte ein Set an störungsspezifischen Messinstrumenten von A wie Angst bis Z wie Zwang vorliegen, welches die typischen bzw. häufig vorkommenden Störungsbilder quantitativ beschreibt. Lutz et al. [2019, S. 70] schildern auch den Gesichtspunkt finanzieller Erlöse: Die Durchführung und Auswertung stellt demzufolge mit knapp 150 € für 5 MZP eine nicht zu vernachlässigende abrechenbare Leistung der Psychotherapie dar, die auch jüngst hinsichtlich der

nochmaligen Ausweitung der genehmigungsfähigen Anzahl an MZP aktualisiert wurde [KBV, 2020].

Dass allzuvielen Messungen jedoch nicht automatisch qualitätserhöhend sind, darauf verweist andererseits eine Übersichtsarbeit aus der sozialpsychologischen Entscheidungsforschung, in der argumentiert wird, dass ganze Serien von subjektiven Schätzungen die Bearbeitungsgüte auf lange Sicht nicht verbessern, sondern im Gegenteil reduzieren, und zwar unabhängig von den Faktoren Ermüdung oder mangelnde Anstrengungsbereitschaft [Minson und Umphres, 2020].

Statement of Ethics

Die Studie von Geissner und Ivert [2019a] und auch das Vorgehen bei der hier berichteten Teilerhebung t_0 - t_1 wurde vorab im Gremium "Wissenschaftskonferenz" der Psychosomatischen Klinik Roseneck vorgestellt, diskutiert und genehmigt. Alle Patient*innen erhielten eine ausführliche schriftliche und – bei Aufnahme – nochmals mündliche Information über die Studie und erklärten auf einem gesonderten Einwilligungsfeld die Bereitschaft zur Teilnahme. Sie erhielten vor allem die Zusicherung der jederzeitigen Rücktrittsmöglichkeit von der Studie ohne Nachteile für den weiteren Klinikaufenthalt.

Conflict of Interest Statement

Keine Interessenskonflikte.

Funding Sources

Die Autoren erhielten keine finanzielle Unterstützung.

Author Contributions

Alle 3 Autoren haben zu etwa gleichen Teilen an dieser Arbeit mitgewirkt.

Literatur

- Bingel U, Kersting A. Placebo – die Macht der Erwartung. *Nervenarzt*. 2020 Aug;91(8):665–6.
- Ehlers A, Margraf J. AKV-Fragebogen zu körperbezogenen Ängsten, Kognitionen und Vermeidung (Manual). Göttingen, Beltz Test Gesellschaft bei Hogrefe, 2001.
- Fisseni HJ. *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe; 2004.
- Geissner E, Hütteroth A. Beck Anxiety Inventory deutsch – ein reliables, valides und praxisgeeignetes Instrument zur Messung klinischer Angst. *Psychosomatik*. *Psychother Med Psychol (Stuttg)*. 2018;68(03/04):118–25.
- Geissner E, Ivert P. Motivationale Faktoren in der Gruppenpsychotherapie am Beispiel Angst – so wichtig wie das Interventionsprogramm selbst [auf Research Gate als Fulltext frei verfügbar]. *Psychotherapie*. 2019a;24:252–68.

- Geissner E, Ivert P. Angsterkrankungen – Exposition, Gruppen- und Einzeltherapie. Handout zum Workshop. 37. Kongress der Fachgruppe Klinische Psychologie und Psychotherapie der Deutschen Gesellschaft für Psychologie. Erlangen, 2019b.
- Geissner E, Kindermann P. Angstbewältigungstherapie – ein Gruppentherapieprogramm bei Panikstörung/Agoraphobie (Manual). Prien am Chiemsee: Klinik Rosenneck; 2010.
- Hiller W, Mombour W, Zaudig M. IDCL – Internationale Diagnosen Checklisten für ICD 10 und DSM IV. Bern: Huber; 1995.
- Hiller W, Schindler A. Response und Remission in der Psychotherapieforschung. Psychosomatik. *Psychother Med Psychol (Stuttg)*. 2011;61(03/04):170–6.
- KBV – Kassenärztliche Bundesvereinigung. *Praxisnachrichten* 23.04.2020. www.kbv.de/html/1150_45852.php
- Lutz W, Neu R, Rubel JA. *Evaluation und Effekterfassung in der Psychotherapie*. Göttingen: Hogrefe; 2019.
- Margraf J, Ehlers A. *Beck Angst Inventar – BAI (Manual)*. Frankfurt: Harcourt; 2007.
- Margraf J, Schneider S. *Panik – Angstanfälle und ihre Behandlung*. Berlin: Springer; 2013.
- Minson JA, Umphres C. Confidence in context: perceived accuracy of quantitative estimates decreases with repeated trials. *Psychol Sci*. 2020 Aug;31(8):927–43.
- Nesselroade JR, Stigler SM, Baltes PB. Regression toward the mean and the study of change. *Psychol Bull*. 1980;88(3):622–37.
- Rheinberg F. *Motivation*. Stuttgart: Kohlhammer; 2008.
- Rief W. Die Rolle von Placebo- und Nocebo-mechanismen bei depressiven Erkrankungen und ihrer Therapie. *Nervenarzt*. 2020 Aug; 91(8):675–83.
- Rief W. Von der verfahrensorientierten zur kompetenzorientierten Psychotherapie-Qualifikation. *Psychotherapeutenjournal*. 2019;3: 261–8.
- Rief W, Nestoriuc Y, Weiss S, Welzel E, Barsky AJ, Hofmann SG. Metaanalysis of the placebo response in antidepressant trials. *J Affect Disord*. 2009;118(1-3):1-8.
- Schindler A, Hiller W. Therapieeffekte und Responderaten bei unipolar depressiven Patienten einer verhaltenstherapeutischen Hochschulambulanz. *Z Klin Psychol Psychother*. 2010;31(2):107–15.
- Schindler AC, Hiller W, Witthöft M. Benchmarking of cognitive-behavioral therapy for depression in efficacy and effectiveness studies – how do exclusion criteria affect treatment outcome? *Psychother Res*. 2011 Nov; 21(6):644–57.
- Schmitt M, Gerstenberg F. *Psychologische Diagnostik*. Weinheim: Beltz; 2014.
- Wambach K, Rief W. Effekte kognitiver Verhaltenstherapie bei Panikstörung: mehr Exposition bringt stabilere Therapieerfolge. *Verhaltenstherapie*. 2012;22(2):95–105.