Proposal of Measurement Occasions for Unbiased Evaluation of Psychotherapy

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

^aDepartment Psychologie, Ludwig-Maximilians-Universität München, Munich, Germany; ^bPsychotherapeutische Praxis, Rosenheim, Germany; ^cFachbereich Psychologie, Universität Koblenz-Landau, Landau, Germany

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 our 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 reduces 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

Geeignete Messzeitpunkte in der Psychotherapieevaluation: Ein Vorschlag

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 (Beispiel-erläuterung). 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

Preliminary Note

A clinical study on the differential effects of motivational factors in anxiety treatment [non-experimental, naturalistic design; Geissner and Ivert, 2019a] provided data at admission, discharge, and at follow-up for a partial sample of $n = 67$, as well as data at *registration* about 8 weeks before going to the hospital, which were not analyzed there. The level of anxiety across these 4 measurement points (MPs) – namely t_0 = registration, t_1 = admission, t_2 = discharge, and t_3 = 6-month follow-up – is the basis for what we present here on the appropriate evaluation of psychotherapy.

Patients, Questionnaires, Procedures

The patients (45 women, 22 men; age $M = 41$ years, $SD = 11$ years) met ICD diagnostic criteria F 40.00, F 40.01, and F 41.00 according to Hiller et al. [1995]; had been ill for an average of 8 years; had usually 1 outpatient course of psychotherapy (more in some cases); and were referred for inpatient treatment because the anxiety problem had persisted. In addition to answering a general catalog of questions as part of the clinical routine, they filled out 5 anxiety questionnaires: Beck Anxiety Inventory BAI [Margraf and Ehlers, 2007; Geissner and Hütteroth, 2018]; Anxiety Cognitions Questionnaire, Body Sensations Questionnaire, Mobility Index-alone (alone) MI-A, and Mobility Index-begleitet (accompanied) MI-B [all Ehlers and Margraf, 2001; original English version Chambless et al., 1984]. For informed consent, the patients received information requesting their participation “to be able to document the course of the anxiety disorder as well as the success of the therapy,” and they gave their consent on a separate form.

Research Question

We first present the results for the 5 anxiety indicators listed above (Table 1). These show that for the intervention phase t_1 – t_2 [6- to 7-week hospitalization; anxiety treatment according to Geissner and Kindermann, 2010; Wambach and Rief, 2012; Margraf and Schneider, 2013],

the effect sizes were $d = 0.54$ – 0.66 , and by adding the follow-up phase, t_1 – t_3 , the effect sizes were $d = 0.56$ – 0.79 .

Evaluation: Naturalistic settings are less effective than experimental studies [Schindler and Hiller, 2010; Schindler et al., 2011; Lutz et al., 2019, p. 23; Rief, 2020], especially since the problem was rather persistent as shown by the previous unsuccessful attempt at outpatient therapy. Nevertheless, the effect sizes for t_1 – t_2 and t_1 – t_3 are not considered abnormally high.

For comparison, this is also illustrated by the effect sizes at t_0 – t_1 , which were still between $d = 0.31$ and 0.46 (see Table 1) and are considered “medium-high” according to Cohen’s classification. These effects occurred without any specific intervention.

This is related to the question posed in our discussion paper: What measurement points (MP) are appropriate for evaluating the psychotherapy and effectively document the success of the efforts undertaken? Would the *point of registration* for psychotherapy not be a more suitable anchor for observing the problem over time? Given that the effect sizes over the entire period of analysis t_0 – t_3 yielded a much better picture, with up to $d = 1.05$ (see Table 1), data squeezing could be suspected in favor of good results. But there are also good reasons for choosing t_0 , which we would like to submit for discussion: aspects related to expectation, content-related and disorder-specific contextual factors, as well as considerations of diagnosis and measurement methodology.

Explicit Problem Description at t_0 (Registration)

Will an initial condition such as, in the present case, a severe anxiety problem causing clear disability in everyday life (sometimes marked physical symptoms, or avoidance behavior that interferes with private and professional life) lead to concise responses in questionnaires that are specifically focused on this problem? A tendency toward blunt description can be assumed, even without imputing exaggeration. After all, the person has decided to go for treatment – in this case inpatient treatment – and wants to convince the future practitioner that the situation is urgent. Waiting times are common for both outpatient and inpatient treatment, and a clear report combined with the relevant questionnaire data may underscore how strong the affected person’s motivation for change is.

Hope for Success between t_0 (Registration) and t_1 (Admission)

We can also assume positive expectations on the part of the patient that the first step toward therapeutic success will have been taken with the actual start of treatment, at

Table 1. Mean anxiety levels before, during, and after therapy (VA for 5 questionnaire measures; $n = 67$ patients with agoraphobia/panic disorder)

	t_0	t_1	t_2	t_3	F/p	Significant contrasts	$d/t_{0/1}$	$d/t_{1/2}$	$d/t_{1/3}$	$d/t_{0/3}$
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, analysis of variance; BAI, Beck Anxiety Inventory; ACQ, Anxiety Cognitions Questionnaire; BSQ, Body Sensations Questionnaire; MI, Mobility Inventory (A, alone; B, accompanied); t_0 , registration; t_1 , admission; t_2 , discharge; t_3 , 6-month follow-up. Significant contrasts according to VAs: a = t_0/t_1 , b = t_1/t_2 , c = t_1/t_3 , d = t_0/t_3 , d = effect size $[(x_1 - x_2)/SD]$. Columns t_0 to t_3 = means and (in brackets) standard deviations. Range: BAI min. = 0, max. = 63; ACQ to MI-B min. = 1, max. = 5 (BAI = scale sum score; ACQ to MI = scale sum/number of items).

least in cases where patients are self-motivated and were not sent for treatment by someone else. Here, it can be supposed that questionnaire data at t_1 include a positive expectation of success – since a commitment to interim treatment has been made – and the anxiety scores could be lower than at t_0 (see Table 1). Expectations of this sort are in line with the hope-for-success concept of motivational psychology [Rheinberg, 2018].

Alongside hope for success, motivational psychology has also described fear of failure or an individual balance between hope for success and fear of failure. Thus, skepticism (with regard to the therapeutic approach or the personality of the therapist, because of previous failures) or a depressed mood could contribute to reduced expectations of success. In any event, in therapy research, reduction in the severity of a disorder without specific therapy is often explained by expectation effects [Rief et al., 2009; Lutz et al., 2019; Rief, 2019, 2020; Bingel and Kersting, 2020].

Security/Relief Signals at t_1 (Admission)

Being admitted for treatment or to a hospital is associated with a number of security signals. Such signals could come from doctors, psychologists, nurses/caregivers, as well as the availability of medication; and self-reassuring cognitions (“Finally I’m in good hands,” “In case something goes wrong, nothing can happen to me. If it does, I will be rescued”), could cause a feeling of relief. The soothing effects of security signals are to be ex-

pected in anxiety patients, but not only in them. These effects are reflected in a reduction of subjectively perceived and reported symptom severity at t_1 (admission to the hospital) compared with t_0 (registration). Although questionnaires often instruct patients to answer “as is usually the case” or to refer to the time frame “during the past week,” it is not certain whether patients adhere to this instruction when filling out the form or rather do not give a true picture of their current situation. Assessment of symptom severity may also be influenced by the prognosis that “it will get better now.” At time t_1 , the perceived and reported intensity of anxiety would then be influenced by an anticipated treatment outcome.

Contextual Effects in the Specific Filling Out of Questionnaires at t_1 (Admission)

Responses to items are also subject to contextual factors: For example, responses to an anxiety questionnaire such as the Mobility Inventory [MI; Ehlers and Margraf, 2001], which records anxiety-related situational limitations in daily life, are affected by admission to the hospital. Entering the hospital setting in itself could lead to a reduction in perceived impairment. Variant “MI-B” (“accompanied”, in German: “b” for “begleitet”) is especially susceptible to such a contextual effect, since perceived mobility may be greater compared to that in everyday life if staff or other patients provide accompaniment in the treatment setting. The same applies to threat assessment

of body-related anxiety components [Body Sensations Questionnaire, Ehlers and Margraf, 2001], within the safe clinical context vs. outside it.

Preliminary Conclusion: A Suitable Anchor for Initial Measurements

The above considerations favor MP t_0 (registration) over MP t_1 (admission) for evaluating therapeutic effect, since at t_1 , the above-described effects reduce symptoms. Notwithstanding the risk of an exaggerated description of symptoms at t_0 , the course of therapy would be more appropriately determined by including the time of registration (t_0), since expectations and contextual effects might well be viewed as components of therapy. The difference between t_0 and t_2 (end of therapy) or between t_0 and t_3 (follow-up) would be greater than if measurement started at t_1 .

Regression toward the Mean in Repeated Measures

Another important issue in this context results from the regression toward the mean. This effect occurs whenever the measured variable does not show a perfect correlation between two time points and extreme scores are overrepresented at the first time point [Schmitt and Gerstenberg, 2014]. This is the norm in clinical contexts, since registration for treatment implies that the problem is quite severe. If the variability of the scores is the same at two time points, the limited correlation implies that the scores vary intraindividually and also that the intraindividual differences between the scores vary. This pattern occurs because of all the unsystematic factors (measurement errors) and systematic factors (actual change) that affect a person's scores differently at the first point than at the second. In the case of clinically relevant variables such as anxiety, this may mean that the level of suffering is greater in one person at the first point than at the second, but not in another person (or not to the same extent). It may also mean that one person describes symptoms more dramatically at the first point than another person, but not at the second point. Two people may also experience no effect of hopefulness on the scores at the first point (having a skeptical attitude after only slight improvement), but this effect does occur at the second point (a feeling of surprise due to a "positive impression" of the therapy), yet the two people experience it to different degrees.

An early suggestion for dealing with regression toward the mean for extreme groups is to measure a variable not just twice, but three times [Nesselroade et al., 1980]. The reasoning behind this proposal is that the scores that were extreme at the first point are less extreme at the second

point due to regression toward the mean, and that the difference between the second and third points is subject to no regression effect or a weaker one.

Applied to our topic, this proposal would require anxiety to be assessed at least twice at t_0 – at the time of the decision to undergo therapy along with registration, thus at t_{01} , and a few days later, at t_{02} . The effectiveness of the intervention would be more fairly judged at t_1 , as discussed above, to avoid the above-described expectation and contextual effects. On the other hand, two assessments at t_0 would reduce the effect of regression toward the mean. The relevant anchor for comparisons with the later MPs would then be t_{02} .

Conclusions, Procedures, Recommendations

In the use of psychotherapy data for evaluation purposes, we therefore advocate 4 measurement points, or 5 points including the follow-up assessment (which is generally useful).

Regression toward the mean would be made manageable by measuring the symptom twice at relatively short intervals during registration, in order to reduce the effect of excessive item scores.

The measurement at t_1 , immediately at the start of treatment or upon hospital admission, when compared with the findings at registration (t_{02}), can be used to estimate expectation effects and/or hope of success. It also reflects the context variable "feeling of security," for example in anxiety patients.

The relevant frame of reference for assessing the effectiveness of therapeutic efforts is the comparison between t_{02} and the end of therapy t_2 or, to determine whether the results have been maintained, t_3 (follow-up).

The more specifically symptom scales are targeted to the individually relevant components of a disorder, the more appropriately treatment effects will be evaluated. Global scores, by contrast (e.g., across all SCL/BSI scales), blur the treatment effect because they include irrelevant items or subscales.

Therapeutic evaluations are very often performed using aggregated data. In order to report the effectiveness of a hospital or therapeutic unit, individual changes are summarized not only within and across disorder groups but also across treatment periods (e.g. for annual reports). This aggregation increases the statistical power of therapy evaluation.

On the other hand, our recommendation on the number of therapy-appropriate measurement points also applies to the *evaluation of individual cases*. To account for measurement error contained in individual scores and its dependability on the reliability of the used measure, the Critical Difference or the Reliable Change Index should

be used for this purpose. Fisseni [2004, p. 73] and Schmitt and Gerstenberg [2014] provide information on the critical difference, and Lutz et al. [2019, p. 21] on the Reliable Change Index. The lower the reliability, the larger the difference that must be achieved in therapy, if we are to ascertain that an actual reduction in the patient's problem has occurred. There are a variety of methods for assessing reliability [Schmitt and Gerstenberg, 2014]. For obvious reasons the retest correlation should only be used if actual changes are excluded in the period under consideration. Otherwise, unreliability and actual changes would be confused and reliability would be underestimated, which would be particularly relevant in clinical contexts. If actual changes between the MPs are to be expected, then the internal consistency (α) or the split-half correlation of the scale should be used to estimate its reliability.

Taking the example of the BAI [Geissner and Hütteroth, 2018], this can be illustrated as follows: We need the reliability, which is quite good here with $\alpha = 0.92$ and the standard deviation (12 points), to calculate the standard error of measurement (4.8 points), and to establish a significance level (e.g., 5%, therefore multiplication by the z-score of 1.96). This gives a critical difference of 9 points. At least this must be achieved; only from this level or higher can a reduction in anxiety be assumed, after adjustment for measurement errors. With reference to the BAI data in Table 1 – assuming a given measurement for an individual case – the critical difference for t_1/t_2 would be just barely exceeded with 10 points, but not for t_1/t_3 . Only inclusion of the t_0 score as a starting point would be clearly outside the critical difference, and thus provide clear results. This example may serve to demonstrate that before-and-after differences must be relatively pronounced on a case-by-case basis in order to judge them as successes.

Exclusion of Non-Pathological Scores

For fair assessment of the efficacy of therapies, aside from the choice of appropriate MPs, it has been suggested to exclude non-pathological scores from the initial measurement. This is especially relevant for questions that arise in clinical practice [Schindler and Hiller, 2010; Hiller and Schindler, 2011; Schindler et al., 2011]. Extremely low symptom severity at the beginning cannot be further reduced by therapy. It remains unknown why some patients score low on a given questionnaire during intake screening, even if they say during exploration and anamnesis that they are under definite psychological stress. These cases do occur occasionally in clinical practice. Here the course of questionnaire scores would be a poor indicator by which to evaluate therapeutic efforts. Geissner and Ivert [2019b], for example, were able to show that effect sizes were sub-

stantially higher when patients were excluded from the evaluation who, following the questionnaire manual, only achieved scores similar to those of the non-clinical comparison group, so that no reduction or only an insignificant one was possible. When before-and-after effect sizes were initially $d = 0.56$ – 0.80 , they increased to $d = 0.83$ – 0.98 after excluding non-pathological initial scores.

Cut-Off Scores

Practitioners should be familiar with the range of non-pathological scores of the measures they use. Often (but not always), they are supplied by manual information about control groups or from norm tables. This is not trivial, because nonclinical comparison subjects may also score above a numerical minimum on measurement instruments for certain disorders, e.g., anxiety, stress, anger, other complaints. It would therefore not be correct to use “zero” as a comparison score for the sake of simplicity. This also applies to the use of cut-off scores [Lutz et al., 2019, p.20]. These are a simplified way of assessing results and should also be briefly mentioned here. If reference data are available for both clinical and non-clinical groups, a mean can be calculated from the means of the two groups. This score then serves as a limit in the evaluation of each case: Is there any clinically relevant elevation at all at t_0 ? And if there is, was a reduction into the non-clinical range achieved after the end of therapy?

Time Span between t_0 – t_1 Unfavorable as a Wait-Control Condition

As discussed above, the waiting period before the start of active intervention is relevant to the extent of change. Rief et al. [2009] initially spoke of placebo effects in this situation, and later more precisely of “expectation-focused effects” [Rief, 2019, 2020]. If a targeted intervention is promised, even after a delay, this will change the patient's response behavior in filling out a questionnaire. Assertions about the t_1 – t_2 difference would then only represent net effects of the active intervention, in comparison to expectation and contextual effects for the t_0 – t_1 difference. This would do less justice to clinical application in the field or in naturalistic study designs, however, and the success of therapeutic measures would be underestimated.

Summary and Practice

If repeated symptom assessment before, during, and after therapy reveals a reduction of symptoms for t_0 – t_1 , this does not imply, in our view, a weakness or reduction

in the value of the active intervention. The fair assessment of success includes use of the scores at registration and the unspecific effect between registration and the start of therapy, as a legitimate part of the total process. For “Routine Outcome Monitoring” [Lutz et al., 2019, pp. 37, 68–70], there should be a set of disorder-specific measuring instruments, anxiety, depression, obsession-compulsion disorder, etc. which quantitatively describe typical or common disorders. Lutz et al. [2019, p. 70] also point out financial considerations: Administration and evaluation are billable services in psychotherapy, with nearly €150 for 5 MPs, which has recently been updated in view of the further expansion of the number of MPs that can be approved [KBV, 2020].

On the other hand, a review of social-psychological research on decision-making points out that a larger number of measurements does not automatically increase the quality of an evaluation, arguing that a multitude of subjective estimates do not improve the quality of decision-making in the long run, but on the contrary reduce it, independently of the factors of fatigue or lack of effort [Minson and Umphres, 2020].

References

- Bingel U, Kersting A. Placebo – die Macht der Erwartung. *Nervenarzt*. 2020 Aug;91(8):665–6.
- Chambless DL, Caputo GC, Bright P, Gallagher R. Assessment of fear of fear in agoraphobics: the body sensations questionnaire and the agoraphobic cognitions questionnaire. *J Consult Clin Psychol*. 1984 Dec;52(6):1090–7.
- 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 Roseneck; 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.

Statement of Ethics

The study by Geissner and Ivert [2019a] and also the procedure for the partial survey reported here, t_0 – t_1 , were presented, discussed, and approved in advance by the “Science Conference” committee of the Psychosomatische Klinik Roseneck (Roseneck Psychosomatic Hospital). All patients received detailed written and – upon admission – also verbal information about the study and declared their willingness to participate on a separate consent form. They were particularly assured that they could withdraw from the study at any time without any unfavorable consequences respecting their further stay in the hospital.

Conflict of Interest Statement

No conflicts of interest.

Funding Sources

The authors received no financial support.

Author Contributions

All 3 authors contributed approximately equally to this work.