CLINICAL PSYCHOLOGY IN EUROPE

The Official Academic Journal of the European Association of Clinical Psychology and Psychological Treatment

PsychOpen GOLD

PsychOpen GOLD

Contents

Editorial

Research Articles

Scientific Update and Overview

Body Exposure, its Forms of Delivery and Potentially Associated Working Mechanisms: How to Move the Field Forward

Andrea S. Hartmann, Eva Naumann, Silja Vocks, Jennifer Svaldi, Jessica Werthmann

Different hypothesized working mechanisms of body exposure in eating disorders shape the way of its delivery. Roads to find best (individualized) practice are outlined.

Has the Time Come to Stop Using the 'Standardised Mean Difference'?

Pim Cuijpers

The 'standardised mean difference' has important limitations which are reviewed in this position paper. Possible alternatives for clinical psychology are discussed.

Editorial

(How) Can Clinical Psychology Contribute to Increasing Vaccination Rates in Europe?

Tania M. Lincoln[1], Winfried Rief[2]

[1] *Clinical Psychology and Psychotherapy, Institute of Psychology, University of Hamburg, Hamburg, Germany.*
[2] *Division of Clinical Psychology and Psychotherapy, Department of Psychology, Philipps-University of Marburg, Marburg, Germany.*

Corresponding Author: Tania M. Lincoln, Universität Hamburg, Institut für Psychologie, Klinische Psychologie und Psychotherapie, Von-Melle-Park 5, 20146 Hamburg, Germany. E-mail: tania.lincoln@uni-hamburg.de

The speed in which several vaccines for COVID-19 were developed, approved and rolled out in Europe is amazing. Unfortunately, though, what started out as a story of success is presently being spoilt by the high rate of vaccine hesitancy and refusal. In those European countries that have already offered a vaccine to all adults the rate of vaccine uptake is levelling off well below 70%. While this seemed sufficient at the beginning of the pandemic, the newer virus variants with higher infectiousness require much higher participation rates for vaccination campaigns, and experts now estimate the necessary threshold to be about 90%.

To better understand vaccine hesitancy, numerous researchers have diligently been studying its putative predictors. Sociodemographic variables they found to be consistently associated with hesitancy or refusal were younger age, female gender, lower income, lower education, unemployment, and migrant status (Freeman et al., 2020; Neumann-Böhme et al., 2020; Sallam, 2021; Wake, 2021). Also, people with more extreme political views (Peretti-Watel et al., 2020), higher social media consumption (Allington et al., 2021; Ebrahimi et al., 2021; Murphy et al., 2021), and more frequent use of messenger services and online video platforms (Holzmann-Littig et al., 2021) seem to be less likely to accept a COVID-19 vaccine.

In general, the decision to participate mainly depends on three factors: "benefit" (What kind of benefit do I expect if I participate?), "harm" (i.e. Are these vaccines producing negative effects on my body?), and feasibility (How difficult is it to participate?). In regard to harm expectations, research has identified attitudes indicative of a general mistrust of the government and its institutions (Ebrahimi et al., 2021; Freeman

et al., 2020; Lincoln et al., 2021; Murphy et al., 2021), conspiracy beliefs (Allington et al., 2021; Freeman et al., 2020; Murphy et al., 2021), and specific concerns related to vaccine safety and efficacy (Freeman et al., 2020; Neumann-Böhme et al., 2020) to predict vaccine willingness. Related to benefit expectations, it has been shown that the lower people perceive the risk of getting COVID-19, the less willing they are to get vaccinated (Allington et al., 2021; Bono et al., 2021; Ebrahimi et al., 2021).

Interventions to combat vaccine hesitancy have included increasing the incentives (e.g. paying people for vaccination or reducing the options of social participation for re-fusers), using role models, combatting misinformation in social media, providing patient choice, and providing low-threshold vaccination in socially deprived areas. Although some of these interventions align with the known socio-demographic predictors, sur-prisingly, none of them seem to focus directly at changing the beliefs driving vaccine hesitancy or refusal. Is this, perhaps, because changing peoples' beliefs is seen as partic-ularly difficult? Maybe. But not for clinical psychologists! Changing beliefs to motivate adaptive behavior is what we do every day. We do not intend to imply that vaccine hesitancy is indicative of a mental health problem. We do think, however, that the beliefs driving vaccine hesitancy are maladaptive, both from an empirical perspective, as they are not well backed up by evidence, and from a functional perspective, as they are putting people at risk. Classifying these attitudes as maladaptive provides us with a unique opportunity to bring in our expertise to the aim of finally moving out of this pandemic. To make this a bit more concrete, what do you think about the following five-step basic cognitive-behavioral intervention for dealing with vaccine hesitant fellow citizens?

1. Specifying the problem in order to set a realistic aim

According to the German COSMO study (COSMO-Konsortium, 2021), only about half of those not yet vaccinated report to be definite refusers, the other half are either merely unsure or even basically willing. A clear understanding of where someone stands is key to setting a realistic goal and finding the most appropriate intervention. Is it a problem of attitude (e.g. not wanting to get a vaccine) or of behavior (being willing in principle, but not having put this willingness into practice yet)? And if it is a problem of attitude, how pronounced is it? Is someone an absolute refuser and is actively spreading misinformation? In this case, a realistic next step could be to sow some seeds of doubt. Or has someone merely got a couple of specific concerns that are preventing him or her from getting a vaccine? In this case, the aim could be to overcome these doubts and develop a principle vaccination willingness. Finally, for those who are already willing in principle, the aim should be to overcome the practical barriers and actually get the vaccine. Prochaska and Di Clemente's (1986) motivational model for change provides a helpful framework for defining what the problem is, where the patient is, and what would be a reasonable next step in this case.

2. **Assessing the reasons and delineating the appropriate intervention**

If the problem is one of attitude, the main concerns need to be explored. Are they related to a profound and generalized mistrust of the government and its institutions or even to conspiracy beliefs or does someone have very specific vaccine safety-concerns? Are we dealing with a fear of side- or long-term effects or with phobic concerns related to the prick of the needle? Is someone clearly underestimating the risk of the pandemic or have they perhaps not understood the differences in probabilities between developing serious side effects, versus getting infected with serious consequences?

In addition, we need to understand how these attitudes are maintained even if contradicting information is provided. Do people use "cognitive immunization" (Rief & Joormann, 2019) strategies to block any effects of corrective information? In this case, it will not be sufficient to provide new information, but also to challenge or circumvent these cognitive immunization strategies.

If the problem is one of behaviour, we need to find out what the specific barriers are (Is it, for example, a lack of time? A lack of knowledge about where to go? Or is someone worried about not having the documents that may be required?). Once we have understood the reasons, it will be clearer which type of intervention is the most promising.

3. **Building rapport to motivate behaviour change**

As any therapist will know, simply telling someone that their beliefs are wrong or unfounded is unlikely to be helpful. It tends to motivate people to defend their beliefs, or even to leave the conversation, never to come back. People have reasons for their beliefs and, in any case, no one can be 100% certain of what is right and wrong. Staying open minded oneself and also expressing authentic understanding for the other persons' beliefs will create an open space in which a change of perspective is more likely to occur. Even in the case of extreme mistrust or conspiracy beliefs, which may be more difficult to empathize with, it can help to express understanding for the frustration with political decisions and the restrictions.

4. **Challenging beliefs**

If we are dealing with maladaptive beliefs, we can now follow the basic pattern of cognitive therapy. We begin by narrowing down the beliefs to one core statement. We then assess how certain the person is of the truth of this belief (maybe using a scale from 1 to 100 percent). We can now move on to the empirical dispute by creating a list of evidence for and against the belief. Mind to use guided discovery and let the other person come up with the pros and cons herself. But if you can't stop yourself from adding to the list, then be sure to add to both sides of evidence. You might also want to discuss the quality of each piece of evidence to help the other person decide which weight to give it. For example, while severe adverse

effects are indeed a matter of concern, mild side effects are generally fleeting in nature and easy to cope with.

Depending on the type of belief, it may also make sense to use functional dispute. This is done by looking at the pros and cons of sticking with the belief versus changing it. Knowing your stats will come in helpful here as this process may well involve weighing up the risk of severe side effects with the risk of a serious COVID-19 infection (Rief, 2021).

Established and powerful "cognitive immunization" strategies, need to be addressed specifically. These strategies can be categorized into two types: devaluating the source of information (e.g. you cannot believe those people who are in favor of vaccination anyway), and devaluating the content of expectation-violating information (e.g. information violating my beliefs is only based on accidental events). They are also evident on a behavioral level by avoidance of any conflicting information and continuing to live in an information bubble. Again, we need to validate cognitive immunization strategies, because they provide stability in our lives, but they are also the basis of continuous misjudgments, and therefore, searching for alternative information channels is crucial to prevent tunnel perspectives.

This process can be concluded by re-assessing the conviction in the belief to see where you stand now. Don't worry if it hasn't changed much, after all its not about winning an argument, but about helping someone to take different aspects into account. According to Prochaska and Di Clemente's (1986) model, it's just about moving to the next level of change, not about completely convincing a person in one blow.

5. **Overcoming behavioral barriers**

If the willingness to get a vaccine is there, in principle, but the problem is one of getting organized, we can use a simplified version of the good old problem-solving scheme. Starting out by repeating the aim (i.e. to "get a vaccine as soon as possible") and defining the barriers ("haven't found the time yet" or "can't find vaccine booklet" etc.), continue with a brainstorming of different options of where, when and how the aim can be achieved, motivate the person to think about the disadvantages and advantages of each option, then encourage to select an option and outline a concrete plan that includes the specific steps involved along with a time and place. If possible, support the person practically (e.g. find the closest doctor offering vaccines) and don't forget to follow up by asking whether the plan was put into practice. If not, elucidate the reasons, select a new option and take it from there.

Does this sound straightforward enough to give it a go? If so, we are curious to hear whether it worked and are awaiting your "case-reports" in the next issue.

References

Allington, D., Duffy, B., Wessely, S., Dhavan, N., & Rubin, J. (2021). Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychological Medicine, 51*(10), 1763-1769. https://doi.org/10.1017/S003329172000224X

Bono, S. A., Faria de Moura Villela, E., Siau, C. S., Chen, W. S., Pengpid, S., Hasan, M. T., Sessou, P., Ditekemena, J. D., Amodan, B. O., Hosseinipour, M. C., Dolo, H., Siewe Fodjo, J. N., Low, W. Y., & Colebunders, R. (2021). Factors affecting COVID-19 vaccine acceptance: An international survey among low- and middle-income countries. *Vaccines, 9*(5), Article 515. https://doi.org/10.3390/vaccines9050515

COSMO-Konsortium. (2021, September 20). *Zusammenfassung und Empfehlungen Wellen 48 bis 51.* https://projekte.uni-erfurt.de/cosmo2020/web/summary/48-51/

Ebrahimi, O. V., Johnson, M. S., Ebling, S., Amundsen, O. M., Halsøy, Ø., Hoffart, A., Skjerdingstad, N., & Johnson, S. U. (2021). Risk, trust, and flawed assumptions: Vaccine hesitancy during the COVID-19 Pandemic. *Frontiers in Public Health, 9*, Article 700213. https://doi.org/10.3389/fpubh.2021.700213

Freeman, D., Loe, B. S., Chadwick, A., Vaccari, C., Waite, F., Rosebrock, L., Jenner, L., Petit, A., Lewandowsky, S., Vanderslott, S., Innocenti, S., Larkin, M., Giubilini, A., Yu, L.-M., McShane, H., Pollard, A. J., & Lambe, S. (2020). COVID-19 vaccine hesitancy in the UK: The Oxford coronavirus explanations, attitudes, and narratives survey (Oceans) II. *Psychological Medicine.* Advance online publication. https://doi.org/10.1017/S0033291720005188

Holzmann-Littig, C., Braunisch, M. C., Kranke, P., Popp, M., Seeber, C., Fichtner, F., Littig, B., Carbajo-Lozoya, J., Allwang, C., Frank, T., Meerpohl, J. J., Haller, B., & Schmaderer, C. (2021). COVID-19 vaccination acceptance and hesitancy among healthcare workers in Germany. *Vaccines, 9*(7), Article 777. https://doi.org/10.3390/vaccines9070777

Lincoln, T. M., Schlier, B., Strakeljahn, F., Gaudiano, B., So, S., Kingston, J., Morris, E., & Ellett, L. (2021). *A machine learning approach to identify psychological factors driving vaccine hesitancy in high income countries.* Research Square. https://doi.org/10.21203/rs.3.rs-787170/v1

Murphy, J., Vallières, F., Bentall, R. P., Shevlin, M., McBride, O., Hartman, T. K., McKay, R., Bennett, K., Mason, L., Gibson-Miller, J., Levita, L., Martinez, A. P., Stocks, T. V. A., Karatzias, T., & Hyland, P. (2021). Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom. *Nature Communications, 12*(1), Article 29. https://doi.org/10.1038/s41467-020-20226-9

PsychOpen GOLD

Neumann-Böhme, S., Varghese, N. E., Sabat, I., Barros, P. P., Brouwer, W., van Exel, J., Schreyögg, J., & Stargardt, T. (2020). Once we have it, will we use it? A European survey on willingness to be vaccinated against COVID-19. *The European Journal of Health Economics, 21*(7), 977-982. https://doi.org/10.1007/s10198-020-01208-6

Peretti-Watel, P., Seror, V., Cortaredona, S., Launay, O., Raude, J., Verger, P., Fressard, L., Beck, F., Legleye, S., L'Haridon, O., Léger, D., & Ward, J. K. (2020). A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. *The Lancet Infectious Diseases, 20*(7), 769-770. https://doi.org/10.1016/S1473-3099(20)30426-6

Prochaska, J. O., & Diclemente, C. C. (1986). Toward a comprehensive model of change. In W. R. Miller & N. Heather (Eds.), *Treating addictive behaviors: Processes of change* (pp. 3–27). Springer US. https://doi.org/10.1007/978-1-4613-2191-0_1

Rief, W. (2021). Fear of adverse effects and COVID-19 vaccine hesitancy: Recommendations of the treatment expectation expert group. *JAMA Health Forum, 2*(4), Article e210804. https://doi.org/10.1001/jamahealthforum.2021.0804

Rief, W., & Joormann, J. (2019). Revisiting the cognitive model of depression: The role of expectations. *Clinical Psychology in Europe, 1*(1), Article e32605. https://doi.org/10.32872/cpe.v1i1.32605

Sallam, M. (2021). COVID-19 vaccine hesitancy worldwide: A concise systematic review of vaccine acceptance rates. *Vaccines, 9*(2), Article 160. https://doi.org/10.3390/vaccines9020160

Wake, A. D. (2021). The willingness to receive COVID-19 vaccine and its associated factors: "Vaccination refusal could prolong the war of this pandemic" – A systematic review. *Risk Management and Healthcare Policy, 14*, 2609-2623. https://doi.org/10.2147/RMHP.S311074

PsychOpen GOLD

Research Articles

Explaining the Efficacy of an Internet-Based Behavioral Activation Intervention for Major Depression: A Mechanistic Study of a Randomized-Controlled Trial

Zhongfang Fu[1], Huibert Burger[2], Retha Arjadi[3,4], Maaike H. Nauta[4],

Claudi L. H. Bockting[1,5] (iD)

[1] *Department of Psychiatry, Amsterdam University Medical Centers, location AMC, University of Amsterdam, Amsterdam, The Netherlands.* **[2]** *Department of General Practice and Elderly Care Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands.* **[3]** *Faculty of Psychology, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia.* **[4]** *Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, Groningen, The Netherlands.* **[5]** *Centre for Urban Mental Health, University of Amsterdam, Amsterdam, The Netherlands.*

Corresponding Author: Claudi L. H. Bockting, Meibergdreef 5, 1105 AZ Amsterdam, The Netherlands. E-mail: c.l.bockting@amsterdamumc.nl

Supplementary Materials: Materials [see Index of Supplementary Materials]

Abstract

Background: Behavioral activation is an effective treatment for depression that is theorized to facilitate structured increases in enjoyable activities that increase opportunities for contact with positive reinforcement; to date, however, only few mechanistic studies focused on a standalone intervention.

Method: Interventions using internet-based behavioral activation or psychoeducation were compared based on data from a randomized-controlled trial of 313 patients with major depressive disorder. Activation level and depression were measured fortnightly (baseline, Weeks 2, 4, 6, 8, 10), using the Patient Health Questionnaire-9 and the Behavioral Activation for Depression Scale-Short Form, respectively. Analysis was performed to determine if a change in activation level mediated treatment efficacy.

Results: Latent growth modeling showed that internet-based behavioral activation treatment significantly reduced depressive symptoms from baseline to the end of treatment (standardized coefficient = −.13, p = .017) by increasing the rate of growth in the activation level (mediated effect

estimate = −.17, 95% CI [−.27, −.07]. Results from mixed effects and simplex models showed that it took 4 weeks before mediation occurred (i.e., a significant change in activation that led to a reduction in depressive symptoms).

Conclusion: Activation level likely mediated the therapeutic effect of behavioral activation on depression in our intervention. This finding may be of significant value to clinicians and depressed individuals who should anticipate a 4-week window before seeing a prominent change in activation level and a 6-week window before depressive symptomatology reduces. Future research must consolidate our findings on how behavioral activation works and when mediation occurs.

Keywords

psychological interventions, working mechanisms, behavioral activation, depression, internet-based intervention, lay counselors

Highlights

- Activation level mediates depression outcomes in an 8-week internet-based behavioral intervention.
- Internet-based behavioral activation appeared to work by changing the level of activation at Week 4 and reducing depressive symptoms over the next 2 weeks.
- Internet-based treatment requires patience and perseverance from clinicians and patients.

Background

Depression is a prevalent and disabling mental health condition characterized by sadness and lack of interest (American Psychiatry Association, 2015). Behavioral activation is well-established as an effective treatment (Cuijpers, Van Straten, & Warmerdam, 2007; Stein, Carl, Cuijpers, Karyotaki, & Smits, 2021) and as a standalone therapy in relevant clinical guidelines (National Collaborating Centre for Mental Health [UK], 2010). It is also considered a cost-effective therapy that can be delivered easily and disseminated in a range of formats (Arjadi et al., 2018; Carlbring et al., 2013). However, more research is needed to clarify uncertainties about how behavioral activation exerts its clinical effects (Janssen et al., 2020).

Rooted in behavioral frameworks, the theory underpinning behavioral activation conceptualizes depression as the result of low levels of (response-contingent) positive reinforcement: the consequences of environmental interaction that increase the likelihood of a given behavior (Ferster, 1973, 1981; Lazarus, 1972; Lewinsohn, 1974). The theory posits that a lack of this positive reinforcement can result in decreased behavioral activation or withdrawal from the environment, which precipitates depression (Manos, Kanter, & Busch, 2010). Therefore, actively engaging in behavioral activation can help to break the negative cycle of depression by promoting meaningful and adaptive engagement in life (Martell, Dimidjian, & Herman-Dunn, 2013). This strong theoretical basis allows for

PsychOpen GOLD

changes in levels of activation and avoidance (i.e., the activation level) to be evaluated as the hypothesized mediator of change in depressive symptoms during treatment (Curry & Meyer, 2016). However, two research gaps remain. First, contrasting starkly with research into cognitive processes, there is limited empirical evidence of activation level as a potential mediator (Lemmens, Müller, Arntz, & Huibers, 2016; Moreno-Peral et al., 2020). Second, mediators have rarely been examined in randomized-controlled trials (RCTs) of behavioral activation as a standalone treatment (Janssen et al., 2020). Further study is needed to correct this lack of mechanistic research into mediation processes.

Most research into behavioral activation has investigated it as a component of cognitive behavior therapy (e.g., van Luenen, Kraaij, Spinhoven, Wilderjans, & Garnefski, 2019), for which the underlying theoretical assumption differs, suggesting instead that behavioral change helps to improve symptoms through cognitive restructuring. To date, ten studies have examined activation level for the treatment of depression (Dimidjian et al., 2017; Forand et al., 2018; Gaynor & Harris, 2008; Nasrin, Rimes, Reinecke, Rinck, & Barnhofer, 2017; Richards et al., 2017; Rovner et al., 2014; Santos et al., 2019; Silverstein et al., 2018; van Luenen et al., 2019; Weidberg, González-Roz, García-Fernández, & Secades-Villa, 2021). Among these, four investigated a standalone behavioral activation intervention, producing inconsistent results, and none assessed both depression and activation during treatment, precluding mediation analyses. The inconsistent findings likely result from clinical heterogeneity and a failure to meet specific methodological requirements, such as using an RCT design, examining variables of interest longitudinally to assess temporal ordering, and being sufficiently large to ensure robust statistical analyses (Curran et al., 2010; Kazdin, 2007; Lemmens et al., 2016). Studies assessing the activation level as a mediator of depression treatment have not complied with all these requirements (Janssen et al., 2020), with some adopting small samples (e.g., <40 per trial arm) (Gaynor & Harris, 2008) and others using too few repeat observations (e.g., <3) (Richards et al., 2017; Weidberg et al., 2021) or no control group (e.g., Santos et al., 2019). Thus, adequately powered trials of standalone behavioral activation interventions for depression are needed to clarify the extent to which the activation level mediates treatment outcomes.

Our group has previously conducted an RCT for an internet-based intervention involving a large sample of patients with major depressive disorder treated by behavioral activation under the guidance of lay counselors (intervention) compared with psychoeducation (controls) (Arjadi et al., 2018). In that study, we concluded that, after 10 weeks, patients in the intervention group reported significantly fewer depressive symptoms (effect size, 0.24) and had a 50% higher chance of remission than those in the control group. Crucially, this study complied fully with the methodological requirements of mechanistic research into mediation processes. In the present study, we therefore aimed to use data from that study to demonstrate that the activation level mediates the relationship between treatment with behavioral activation and improved depression. This was

considered achievable if we could demonstrate two criteria (Kazdin, 2007; MacKinnon, 2008). First, that the treatment condition correlated with changes in the activation level, which in turn, correlated with changes in depressive symptoms and was conditional on treatment allocation (Criterion 1). Second, that the change in activation level produced the change in depressive symptoms, and not vice versa (i.e., temporal ordering; Criterion 2).

Materials and Method

Design

This study reports on a post-hoc analysis of an earlier two-group RCT of an internet-based behavioral activation program for patients with major depressive disorders (N = 313). Details of the original RCT are reported elsewhere (Arjadi et al., 2018). All assessments were completed on the Qualtrics survey platform and administered at baseline and every 2 weeks thereafter up to the main post-treatment evaluation at Week 10 (endpoint), with follow-up at 12 and 24 weeks after baseline. For the purposes of the current study, depression and activation level were examined fortnightly at baseline and at Weeks 2, 4, 6, 8, and 10.

Participants and Randomization

In total, 313 participants were included and randomized into the treatment (n = 159) and control (n = 154) groups (see Arjadi et al., 2018, for a detailed flowchart). The baseline characteristics we comparable in each group, as presented in Table 1, indicating successful randomization. Participants were recruited via online self-referral. Eligible participants were aged ≥16 years, scored ≥10 on the Patient Health Questionnaire-9 (PHQ-9), and had a principal diagnosis of major depressive disorder or persistent depressive disorder defined according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Diagnosis was by semi-structured diagnostic interview (SCID-5) (First et al., 2015). Participants with current substance use disorder, current or previous manic or hypomanic episodes, psychotic disorder, or acute suicidality were excluded, as were those receiving psychological interventions.

Eligible participants were allocated (1:1) by a research assistant in a random permuted block design stratified by sex and depression severity (score 10–14 or ≥15 on the PHQ-9) via a web-based program. Current depressive episodes and post-traumatic stress disorder were assessed by clinical diagnostic interview conducted by trained clinical interviewers who were required to hold at least a bachelor's degree in psychology.

Table 1

Descriptive Statistics of Baseline Demographic Characteristics

Demographic information	GAF (*n* = 159)[a]	PE (*n* = 154)[a]
Age (*M, SD*)	24.5 (4.9)	24.5 (5.2)
Sex		
Female	128	125
Male	31	29
Current PTSD		
Yes	22	30
No	137	124
Education		
Above bachelor	89	81
Others	70	73
Living area		
Urban	93	96
Others	67	58
Socioeconomic class		
Low	32	27
Middle	98	100
High	29	27
Ethnicity		
Java	69	64
Tionghoa	30	18
Sunda	21	22
Others	39	40

[a]Note that all patients were in a depressive episode. Abbreviations: GAF = Guided Act-and-Feel-Indonesia; PE = Psychoeducation; PTSD = post-traumatic stress disorder; *SD* = standard deviation.

Treatments

Intervention Group: Guided Act-and-Feel-Indonesia (GAF-ID)

Participants in the intervention group received an internet-based behavioral activation intervention (the GAF-ID) supported by lay counselors. The intervention program was adapted from an online intervention for behavioral activation based on Lewinsohn's (1974) theory of depression. The original program was published in Dutch (Doe en Voel; Bockting & Van Valen, 2015) and was translated to Bahasa Indonesian. The GAF-ID program was delivered using an online platform in eight structured modules delivered weekly. Each module was expected to be completed online in 30–45 minutes. The intervention group was guided and supported by lay counselors who were supervised by

a licensed clinical psychologist. A detailed description of the guidance and support is available elsewhere (Arjadi et al., 2018).

Control Group: Online Psychoeducation

Participants in the control group were given access to another online platform from which they could find basic psychoeducation on depression and brief tips on coping with depression in general. This information was distilled from the psychoeducation module of the GAF-ID program, but no guidance or support was provided.

Measures

Demographic information was collected at baseline, including age, gender, ethnicity, education (above bachelor/other), living area (urban/other), and socioeconomic class. The latter was determined by monthly expenditure in Indonesian rupiah (IDR): low, <1 million; middle, 1–5 million; and high, >5 million. In addition, the PHQ-9 and Behavioral Activation for Depression Scale-Short Form (BADS-SF) were completed fortnightly.

Patient Health Questionnaire-9 Item Version

The PHQ-9 is a 9-item self-report questionnaire in which participants rate how they felt during the previous two weeks (e.g., "Feeling tired or having little energy"). Each question is scored 0 to 3 (0 = *not at all*, 1 = *several days*, 2 = *more than half the days*, and 3 = *nearly every day*). Sum scores range from 0 to 27, with higher scores representing higher levels of depression. The PHQ-9 has acceptable validity and reliability (Carroll et al., 2020), and the Cronbach's alphas in the current study ranged from .78 to .87 at the different assessments.

Behavioral Activation for Depression Scale-Short Form

The BADS-SF is a 9-item self-report questionnaire that measures changes in activation and avoidance in the previous week (e.g., "There were certain things I needed to do that I didn't do"). Each question is scored 0 to 6 (0 = *not at all*, 6 = *completely*). Items 1, 6, 7, and 8 are reverse-coded. Sum scores can range from 0 to 54, with higher scores representing higher activation. The validity and reliability of BADS-SF have been established (Manos, Kanter, & Luo, 2011), and the Cronbach's alphas in the current study ranged from .78 to .88 at different assessments.

Data Analysis

Mixed Effects Model to Compare Mean Depression and Activation Levels

Mixed effects models were used to inspect how treatment influenced activation level and depression at each time point. Baseline and follow-up measures were treated as response variables. Missing values were imputed by multiple imputation, including treatment

allocation and all PHQ-9 and BADS-SF assessments in the predictor matrix. Given that the functional form of the mean responses during treatment can be difficult to anticipate, time was specified as a class effect in an unstructured manner. The contrasts between treatment groups at each time point were obtained by comparing the least squares means of the variables of interest. Mixed effect analyses were conducted using the *nlme* R package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2020), and for multiple imputations, we used the *mice* R package (van Buuren & Groothuis-Oudshoorn, 2011).

Mediation Analyses Using Latent Growth and Simplex Mediation Models

Mediation analyses were based on latent growth models to address criterion 1 (MacKinnon, Cheong, & Pirlott, 2012) and simplex mediation models to address criterion 2 (Goldsmith et al., 2018) in a structural equation model framework.

We refer to the path estimating the relationship between treatment allocation (T) and activation level (M) as the *a* path and refer to the path between activation level and depression (Y) as the *b* path. The direct effect from treatment allocation to depression is noted as the *c* path, after accounting for M as *c′*. The product of *a* × *b* coefficients method was used to indicate the indirect effect (Goldsmith et al., 2018). Coefficients were provided based on a completely standardized solution, and the confidence intervals of *a* × *b* were estimated by bootstrapping (1,000 times). A mediated effect was deemed statistically significant if the 95% confidence interval (95% CI) did not cross zero.

Latent growth model analyses were performed in three steps to model the relationship between treatment and the growth trajectories of activation and depression (Cheong, MacKinnon, & Khoo, 2003). First, to investigate the shape of the growth trajectories for depression and activation, unconditional growth models were built. Second, to examine if the growth rates of depression and activation differed by treatment condition, two conditional models were constructed with the treatment conditions. Third, to assess the indirect effect of treatment allocation on the outcome, via the mediator (activation level), we combined the two conditional growth models into a parallel process growth model. In this, the path coefficients (*a*, *b*, *c*, and *c′*) of the mediation model were estimated and the contributions of baseline characteristics as covariates were examined (e.g., sex, ethnicity, urban/rural, socioeconomic status, post-traumatic stress disorder, and education level).

A simplex mediation model was then adopted to determine if there was temporal ordering. This was achieved by evaluating whether a prior activation level was associated with the level of depression at a subsequent measurement. We specified models as either a lagged *b* path (activation affects depression at adjacent time points) or a contemporaneous *b* path (activation affects depression at the same time point). We added treatment allocation as a time-invariant antecedent variable to predict depression and activation level at each time point. Autoregressive and cross-lagged effects were constrained to be equal over time (Goldsmith et al., 2018). To assess the timing of the

potential mediation process, *a* paths were freely estimated. In addition, to evaluate the extent to which prior depression influenced the subsequent activation level, we reversed the position of depression and activation level in a supplementary analysis (see Supplementary Materials).

The time-specific indirect effect was estimated using a series of product terms to indicate the possible timing of the putative mediator taking effect. Figure 1 shows an example simplex model with lagged *b* paths: for the third time point, depression Y_3 (i.e., Week 4 depression), one indirect effect of treatment could be T→ M_2 →Y_3. Calculation was performed as $a_2 \times b_{23}$, where the subscripts indicated direction (e.g., the coefficient a_2 was the effect to activation at Point 2, and b_{23} was the effect from activation at Point 2 to depression at Point 3, and all *b* paths were considered equal). A significant result could suggest a lagged mediation effect from Week 2 activation (M_2) to Week 4 depression (Y_3). The overall indirect effect in the model for Y_3 was the sum of all time-specific indirect effects estimated by the products of the parameters that estimated the paths between T and Y_3 and passed through the mediator. Coefficient *a* at baseline (i.e., a_1) was fixed at zero because treatment had not been implemented at this time.

Figure 1

Example Diagram of Simplex Models for Mediation With Contemporaneous b Paths (Right Side) and Lagged b Paths (Left Side) With Depression at Third Timepoint (Week 6) as Outcome
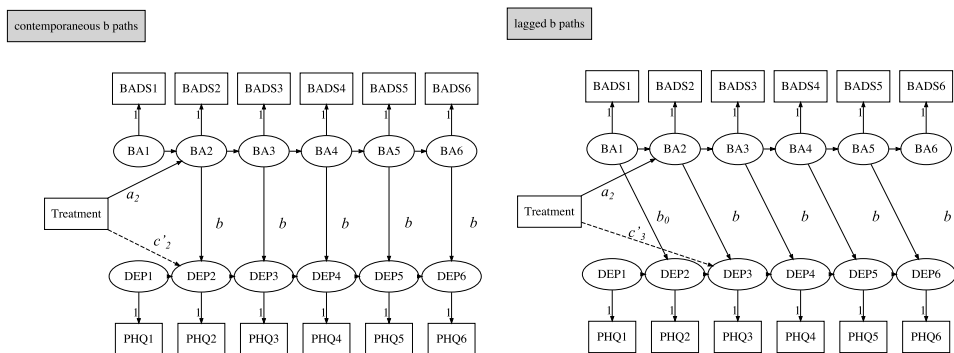
Note. Abbreviations: a_2 = parameter estimated coefficient from treatment to Week 4 behavioral activation; *b* = parameter estimated coefficient from mediator to outcome; b_0 = parameter estimated coefficient from baseline mediator to Week 2 depression; BA, behavioral activation; BADS(1, 2, 3, 4, 5, 6) = Behavioral activation of depression scale-Short form (baseline and 2, 4, 6, 8, 10 weeks, respectively); c'_2, c'_3 = parameter estimated coefficient from treatment to Week 4, 6 depression after controlled for intermediate behavioral activation; PHQ(1, 2, 3, 4, 5, 6) = Patient Health Questionnaire-9 items (baseline and 2, 4, 6, 8, 10 weeks, respectively).

Data were assumed to be missing at random or completely at random (Graham, 2009), so we used a full-information maximum likelihood estimation in the structural equation

modeling analysis. Participants who had at least one measurement for depression were retained in the model and analysis performed on an intention-to-treat basis. Model fit was assessed by the comparative fit index (CFI), Tucker–Lewis index (TLI), root mean squared error of approximation (RMSEA), and standardized root mean square residual (SRMR). We used established guidelines of acceptable fit, requiring that the CFI and TLI should exceed 0.90–0.95, that the RMSEA should not exceed 0.06–0.10, and that the SRMR should not exceed 0.08. All structural equation modeling analyses were performed in Mplus 8.3 (Muthén & Muthén, 2019).

Results

A full overview of the levels of activation and depression at each measurement is presented in Table 2.

Table 2

Means and Standard Deviations of PHQ-9 and BAD-SF for Each Group at Each Assessment

Measure	GAF-ID			PE		
	Missing	Means	*SD*	Missing	Means	*SD*
Depression (PHQ-9)						
Week 0 (Baseline)	0	17.92	5.39	0	18.01	5.05
Week 2	21	12.04	6.05	2	12.81	5.97
Week 4	33	10.53	6.04	10	11.33	6.01
Week 6	31	9.79	5.80	8	11.18	5.85
Week 8	43	9.07	6.22	11	10.48	6.12
Week 10 (Endpoint)	39	8.50	5.75	9	10.83	6.21
Behavioral activation (BADS-SF)						
Week 0 (Baseline)	0	16.67	6.72	0	16.38	6.29
Week 2	21	19.59	6.75	2	18.68	6.64
Week 4	33	23.22	7.32	10	19.93	6.87
Week 6	31	24.11	7.94	8	20.57	7.61
Week 8	43	24.93	8.06	11	22.22	7.72
Week 10 (Endpoint)	39	24.12	7.37	9	20.73	7.45

Note. Abbreviations: BADS-SF = Behavioral Activation for Depression Scale – Short Form; GAF = Guided Act-and-Feel-Indonesia; PE = Psychoeducation; PHQ-9 = Patient Health Questionnaire-9; *SD* = standard deviation.

Each fortnightly assessment was completed by at least 83% of the sample, but 17.5% of all data points were missing in the GAF-ID group versus 4.3% in the control group. Participants in both groups had at least 4 data points (83.6% for the GAF-ID group and 95.4% for the control group). The main reasons for dropout at Week 10 were "no time"

(18 in the GAF-ID group) and "no improvement" (12 in GAF-ID group and 6 in the control group).

Mixed Effects Model: Differences of Depression and Activation Level

Treatment allocation had significant effects on depression ($p < .001$) and activation ($p < .001$) across all included time points. As shown in Table 3, the mean differences in activation and depression increased over time between the treatment and control groups, reaching statistical significance from Week 4 (Assessment 3) for activation and Week 6 (Assessment 4) for depression.

Table 3

Means Difference of Depression and Activation Between Treatment and Control Groups Over Time (Unstructured Time Model)

Time point	LSMD	SE	95% CI	p value
Behavioral Activation (BADS-SF)				
Week 0 (Baseline)	0.30	0.74	[−0.77, 1.36]	.688
Week 2	0.70	0.77	[−0.46, 1.87]	.360
Week 4	3.47	0.94	[1.72, 5.21]	**< .001**
Week 6	3.41	1.01	[1.39, 5.42]	**.002**
Week 8	2.86	0.96	[1.05, 4.63]	**.004**
Week 10 (Endpoint)	3.36	0.89	[1.82, 4.91]	**< .001**
Depression (PHQ-9)				
Week 0 (Baseline)	−0.08	0.59	[−0.77, 0.60]	.890
Week 2	−0.61	0.69	[−1.55, 0.33]	.379
Week 4	−0.97	0.72	[−1.97, 0.04]	.178
Week 6	−1.41	0.68	[−2.31, −0.50]	**.039**
Week 8	−1.76	0.74	[−0.68, −2.84]	**.019**
Week 10 (Endpoint)	−2.59	0.71	[−3.56, −1.61]	**< .001**

Note. Abbreviations: BADS-SF = Behavioral activation for depression scale-short form; CI = confidence interval; LSMD = least squares mean difference; PE = Psychoeducation; PHQ-9 = Patient Health Questionnaire-9; *SE* = standard error.

Latent Growth Model for Mediation

Unconditional Growth Model

Model fit indices, as shown in Table 4, were acceptable. The RMSEA for the model of depression was higher than that of activation level, suggesting that the variance in depression could be explained by a potential covariate (e.g., treatment).

Table 4

Fit Indices of Latent Growth Models

Model	CFI	TLI	RMSEA (90%CI)	SRMR
Depression (unconditional model)	0.96	0.94	0.11 [0.08, 0.14]	0.07
Treatment–Depression	0.96	0.94	0.09 [0.07, 0.12]	0.06
BA (unconditional model)	0.99	0.99	0.04 [0, 0.08]	0.04
Treatment–BA	0.99	0.99	0.04 [0, 0.07]	0.04
Treatment–BA–Depression	0.97	0.96	0.05 [0.04, 0.07]	0.05

Note. Abbreviations: BA = Behavioral activation; CFI = comparative fit index; CI = confidence interval; RMSEA = root mean squared error of approximation; SRMR = standardized root mean square residual; TLI = Tucker–Lewis index.

Conditional Growth Models: The Effect of Treatment

The fitness of both conditional models appeared acceptable (Table 4). The GAF-ID group showed a larger increase in activation (standardized coefficient = .27, p < .001) and a larger reduction in depression compared with the control group (standardized coefficient = −.13, p = .017). This confirmed that treatment was efficacious in producing a difference in trajectories between the treatment and control groups.

Parallel Process Growth Models: The Mediation Effect

Model fit of the parallel process growth model was acceptable (Figure 2). Factor loadings of the slope growth factor indicating the predicted trajectory of depression and activation are presented in Table 5.

Table 5

Growth Factor Loadings for Intercept and Slope Factors in the Parallel Latent Growth Models for Depression and Activation Level

Time point	Depression (PHQ-9)		Behavioral Activation (BADS-SF)	
	Intercept	Slope	Intercept	Slope
Week 0 (Baseline)	1	0	1	0
Week 2	1	0.65	1	0.42
Week 4	1	0.85	1	0.84
Week 6	1	0.93	1	1.00
Week 8	1	1.01	1	1.17
Week 10 (Endpoint)	1	1.00	1	1.00

Note. Abbreviations: BADS-SF = Behavioral activation for depression scale-short form; PHQ-9 = Patient Health Questionnaire-9.

Consistent with the plotted growth trajectory for depression based on data for the whole sample (see Figure 3a), there was a sharp decrease (0.65 unit) in depressive symptoms from the second week. The reduction in depression continued, reaching a trough at Week 8 that persisted to Week 10 (endpoint). A slightly different pattern was observed for the trajectory of the activation level. As shown in Figure 3b and Table 5, activation increased by 0.42 units after the second week of treatment, peaking at Week 8 before decreasing slightly at Week 10 (endpoint).

Figure 2

Parallel Process Latent Growth Model of Depression and Activation Level Conditioned on Treatment Groups
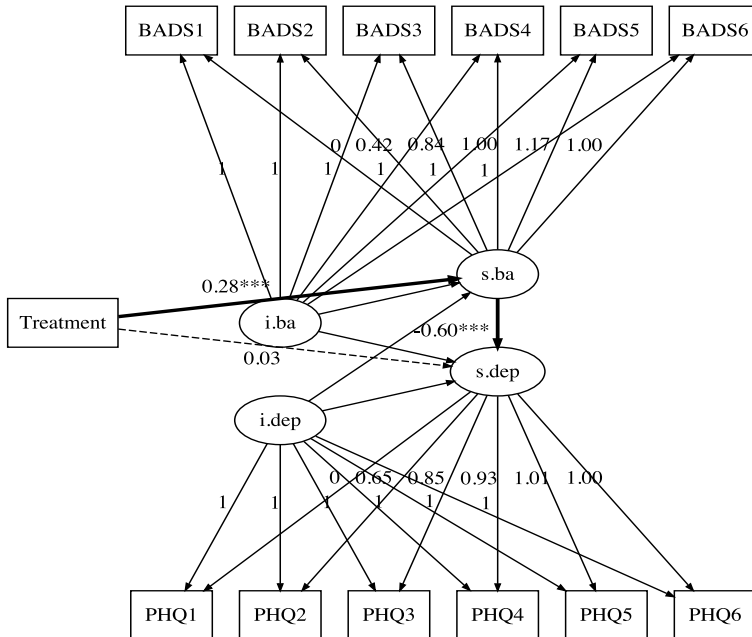
Note. Rectangles denote observed variables, and ellipses denote latent variables. Bolded arrows indicated the significant prediction from treatment to growth of activation, growth of activation to growth of depression. Dashed arrow indicated the insignificant prediction from treatment to growth of depression. Abbreviations: BADS(1, 2, 3, 4, 5, 6) = Behavioral activation of depression scale-Short form (baseline and 2, 4, 6, 8, 10 weeks, respectively); i.dep = intercept growth factor of depression; i.ba = intercept growth factor of behavioral activation; PHQ(1, 2, 3, 4, 5, 6) = Patient Health Questionnaire-9 items (baseline and 2, 4, 6, 8, 10 weeks, respectively); s.ba = slope growth factor of behavioral activation; s.dep = slope growth factor of depression.

Figure 3a

Trajectories of Depression (PHQ-9) Across Measurements in Treatment (GAF) and Control (PE) Groups
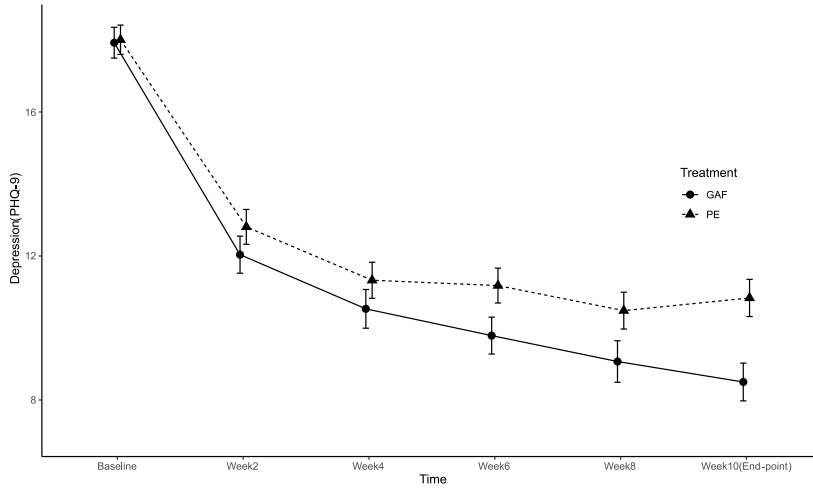
Note. GAF = Guided Act and Feel treatment; PE = Psychoeducation.

Figure 3b

Trajectories of Activation (BADS-SF) Across Measurements in Treatment (GAF) and Control (PE) Groups

Note. GAF = Guided Act and Feel treatment; PE = Psychoeducation.

PsychOpen GOLD
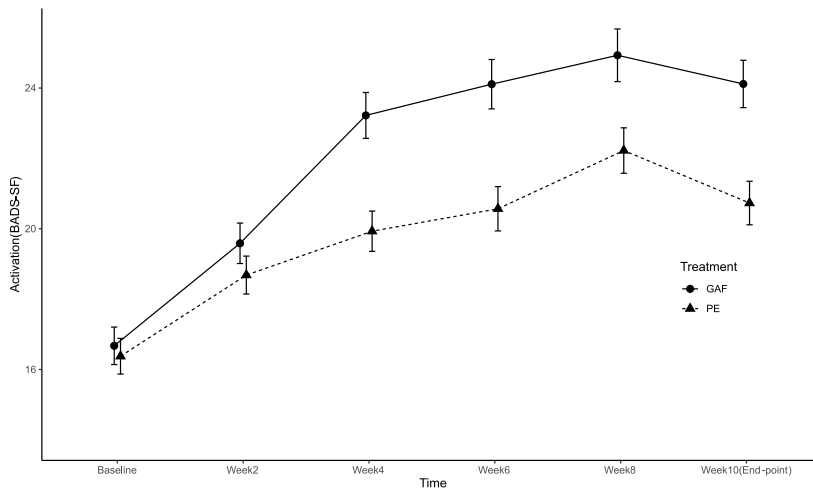
Treatment condition (GAF-ID or control) was significantly associated with the slope factor of activation level (path *a*, standardized coefficient = 0.28, *p* < .001), which in turn was associated with the slope factor of depression (path *b*, standardized coefficient = −0.60, *p* < .001). After accounting for the growth trajectory of the activation level, the prediction that treatment affected depression was no longer significant (path *c′*, standardized coefficient = 0.03, *p* = .483). Table 6 shows that the estimated mediated effect (*a* × *b* product) was standardized as −0.17, 95% CI [−0.27, −0.07], *p* = .001. After adding the baseline characteristics as covariates, model fit was similar, CFI = 0.97, TLI = 0.96, RMSEA = 0.04, 90% CI [0.03, 0.05], and SRMR = 0.05. The estimated mediated effect in this model was similar to that in the model without baseline characteristics as covariates, standardized estimate = −0.15, 95% CI [−0.25, −0.08], *p* < .001.

**Table 6**

Regression Coefficients of Mediational Parallel Process Growth Models

Model	Standard coefficient	SE	p value
Conditional Models			
Treatment–Depression	−0.13	0.06	**.017**
Treatment– BA	0.27	0.06	**< .001**
Parallel process model			
Treatment–BA (*a* path)	0.28	0.06	**< .001**
BA–Depression (*b* path)	−0.60	0.08	**< .001**
Treatment–Depression (*c′* path)	0.03	0.05	.483
a × *b* product	−0.17	0.05	**.001**

Note. Abbreviations: BA = Behavioral activation; *SE* = standard error.

Time-Specific Mediation Effect in the Simplex Models

For the simplex models with activation level as a mediator, fit indices with a contemporaneous *b* path were adequate, CFI = 0.96, TLI = 0.95, RMSEA = 0.06, 90% CI [0.05, 0.08], and SRMR = 0.07. Table 7a shows that the contemporaneous indirect effect reached significance from Week 6. Table 7b summarizes the results with only significant lagged indirect paths, showing that the paths all passed through M_3 (i.e., activation level at Week 4) to influence either contemporary depression or subsequent mediators (M_n), and ultimately, later depression. Fit indices of the simplex mediation model with the lagged *b* path were adequate, CFI = 0.95, TLI = 0.94, RMSEA = 0.07, 90% CI [0.06, 0.08], and SRMR = 0.08. As shown in Table 7b, the indirect effect reached significance from Week 6 onwards. As with the contemporaneous *b* paths, M_3 was the only mediator to be passed through during the treatment.

Table 7a

Simplex Model With Contemporaneous B Paths for Activation Level as a Mediator

Simplex for mediation with contemporaneous *b* path				95% CI	
Time-specific outcome / Significant Paths and Effect of treatment	Standardized estimate	SE	*p*	LL	UL
Week 2 Depression (Y_2)					
Total effect	−0.05	0.05	.320	−0.17	0.06
Indirect effect	−0.01	0.01	.379	−0.03	0.01
Week 4 Depression (Y_3)					
Total effect	−0.08	0.06	.189	−0.20	0.05
Indirect effect	−0.09	0.04	.035	−0.19	−0.001
$T{\rightarrow}M_3{\rightarrow}Y_3$	−0.04	0.02	.006	−0.08	−0.01
Week 6 Depression (Y_4)					
Total effect	−0.13	0.06	.028	−0.25	−0.004
Indirect effect	−0.12	0.05	.016	−0.22	−0.01
$T{\rightarrow}M_3{\rightarrow}Y_3{\rightarrow}Y_4$	−0.03	0.01	.005	−0.06	−0.01
$T{\rightarrow}M_3{\rightarrow}M_4{\rightarrow}Y_4$	−0.04	0.02	.006	−0.08	−0.02
Week 8 Depression (Y_5)					
Total effect	−0.15	0.06	.012	−0.27	−0.02
Indirect effect	−0.14	0.05	.003	−0.25	−0.04
$T{\rightarrow}M_3{\rightarrow}Y_3{\rightarrow}Y_4{\rightarrow}Y_5$	−0.02	0.01	.004	−0.04	−0.01
$T{\rightarrow}M_3{\rightarrow}M_4{\rightarrow}Y_4{\rightarrow}Y_5$	−0.03	0.01	.004	−0.05	−0.01
$T{\rightarrow}M_3{\rightarrow}M_4{\rightarrow}M_5{\rightarrow}Y_5$	−0.04	0.01	.005	−0.07	−0.01
Week 10 Depression (Endpoint, Y_6)					
Total effect	−0.22	0.06	< .001	−0.34	−0.09
Indirect effect	−0.16	0.05	.001	−0.26	−0.06
$T{\rightarrow}M_3{\rightarrow}Y_3{\rightarrow}Y_4{\rightarrow}Y_5{\rightarrow}Y_6$	−0.02	0.01	.004	−0.03	−0.01
$T{\rightarrow}M_3{\rightarrow}M_4{\rightarrow}Y_4{\rightarrow}Y_5{\rightarrow}Y_6$	−0.02	0.01	.004	−0.03	−0.01
$T{\rightarrow}M_3{\rightarrow}M_4{\rightarrow}M_5{\rightarrow}Y_5{\rightarrow}Y_6$	−0.03	0.01	.004	−0.05	−0.01
$T{\rightarrow}M_3{\rightarrow}M_4{\rightarrow}M_5{\rightarrow}M_6{\rightarrow}Y_6$	−0.04	0.01	.005	−0.06	−0.01

Note. Only significant paths are shown to save space. Abbreviations: M_3, M_4, M_5 = mediator measurements (taken at Weeks 4, 6, and 8, respectively); *SE* = standard error; T = Treatment allocation (treatment group = 1, control group = 0); Y_2, Y_3, Y_4, Y_5, Y_6 = outcome measurements (taken at Weeks 2, 4, 6, 8, and 10, respectively).

Table 7b

Simplex Model With Lagged b Paths for Activation Level as a Mediator

Simplex model for mediation with lagged b path				95% CI	
Time-specific outcome / Significant Paths and Effect of treatment	Standardized estimate	SE	p	LL	UL
Week 4 Depression (Y_3)					
Total effect	−0.08	0.06	.187	−0.21	0.05
Indirect effect	−0.05	0.05	.269	−0.15	0.03
Week 6 Depression (Y_4)					
Total effect	−0.13	0.06	.025	−0.25	−0.01
Indirect effect	−0.11	0.05	.033	−0.22	0.001
$T \rightarrow M_3 \rightarrow Y_4$	−0.04	0.02	.01	−0.07	−0.01
Week 8 Depression (Y_5)					
Total effect	−0.16	0.06	.01	−0.28	−0.02
Indirect effect	−0.15	0.05	.003	−0.26	−0.04
$T \rightarrow M_3 \rightarrow Y_4 \rightarrow Y_5$	−0.03	0.01	.008	−0.05	−0.01
$T \rightarrow M_3 \rightarrow M_4 \rightarrow Y_5$	−0.04	0.02	.01	−0.07	−0.01
Week 10 Depression (Endpoint,Y_6)					
Total effect	−0.22	0.06	< .001	−0.35	−0.09
Indirect effect	−0.15	0.05	.002	−0.26	−0.04
$T \rightarrow M_3 \rightarrow Y_4 \rightarrow Y_5 \rightarrow Y_6$	−0.02	0.01	.007	−0.04	−0.01
$T \rightarrow M_3 \rightarrow M_4 \rightarrow Y_5 \rightarrow Y_6$	−0.03	0.01	.008	−0.05	−0.01
$T \rightarrow M_3 \rightarrow M_4 \rightarrow M_5 \rightarrow Y_6$	−0.04	0.01	.009	−0.07	−0.01

Note. Only significant paths are shown to save space. Abbreviations: M_3, M_4, M_5 = mediator measurements (taken at Weeks 4, 6, and 8, respectively); *SE* = standard error; T = Treatment allocation (treatment group = 1, control group = 0); Y_2, Y_3, Y_4, Y_5, Y_6 = outcome measurements (taken at Weeks 2, 4, 6, 8, and 10, respectively).

For the simplex models with depression as a mediator, the fit indices were acceptable for both contemporary b paths, CFI = 0.97, TLI = 0.95, RMSEA = 0.06, 90% CI [0.05−0.08], SRMR = 0.06, and lagged b paths, CFI = 0.95, TLI = 0.93, RMSEA = 0.07, 90% CI [0.06−0.09], SRMR = 0.08. None of the significant indirect effect from treatment allocation to activation level at each time point passed through depression, indicating that our intervention works though the impact of activation on depression rather than the other way around. More detailed results are provided in the Supplementary Materials.

PsychOpen GOLD

Discussion

In this study of data from a large RCT, we provide evidence that activation level underpinned the clinical response to a guided internet-based intervention for depression. During the 8-week treatment period, we showed that (1) our treatment improved activation levels from Week 4 and reduced depressive symptoms from Week 6, and (2) the activation level acted as a mediator for the change in depressive symptoms.

These findings support the theory that a change in depression is contingent on a change in activation level (e.g., Lewinsohn, 1974). We first confirmed that statistically significant associations existed between treatment allocation, activation, and depression level that were not affected by controlling for baseline characteristics. We further supported this by demonstrating temporal order, evidencing that the significant increase in activation level at Week 4 preceded the significant decrease in depressive symptoms at Week 6. This was strengthened by the lack of a "reverse" effect of depression on the activation level when conditioned on treatment. Together, these findings strongly suggest that the hypothesized mediation process occurred around Week 4.

Our findings are consistent with those of similar randomized studies (e.g., Dimidjian et al., 2017; Nasrin et al., 2017; Santos et al., 2017), but conflict with those presented elsewhere. For example, Richards et al. (2017) observed no mediation effect of activation level in a large RCT comparing behavioral activation and cognitive behavioral therapy, nor did Rovner et al. (2014), when they compared behavioral activation and supportive therapy to prevent depression in older adults. There are a couple of plausible explanations for these incongruencies. First, different control conditions were used, with inactive control groups in the first two (waitlist control or usual obstetric care; similar to ours) (Dimidjian et al., 2017; Nasrin et al., 2017) and active control groups in the latter two (Richards et al., 2017; Rovner et al., 2014). Second, measurements were taken at different times, with previous studies assessing mediation either immediately (Dimidjian et al., 2017; Nasrin et al., 2017) or 4 to 6 months (Richards et al., 2017; Rovner et al., 2014) after completing the intervention. Delaying measurements in this way is less likely to capture significant changes caused by the mediator during treatment.

Two studies have used interventions for depression in which the activation level was examined as a putative mediator, and among these, our findings agree with one and disagree with another. In the research by van Luenen et al. (2019) who adopted a similar intervention timeframe (eight sessions completed in 8–10 weeks), it was concluded that the investigated mediation occurred between Weeks 3 and 5. However, this was not apparent in the research by Forand et al. (2018) in another 10-week internet-based trial of cognitive behavioral therapy for depression, who found that the change in activation from baseline to Week 3 did not predict the subsequent change in depression. This inconsistency could be attributed to the fact that Forand et al. (2018) included another potential mediator (cognitive skills) in their mediation model. If activation level were a proximal process that led to another mediation process, controlling for this specific

factor may fail to reveal the activation level as a mediator. It could also be that mediation occurred after Week 3 of the intervention; therefore, a test based on earlier change will not have captured the required period. Nevertheless, although the weight of evidence may be shifting, these inconsistencies point to a requirement for more evidence to confirm the mediational role of activation level.

Regarding missing data, more was missing in the intervention group (17.5%) than in the control group (4.3%). This was presumably because the GAF-ID intervention demanded greater effort to accomplish and because some participants could not afford the time. Alternatively, sending the fortnightly measurements via email separately to monitoring within the intervention may have led to some participants erroneously believing that they had already completed the questionnaires.

Our results help to clarify how internet-based and lay-counselor-guided behavioral activation treatments work. Clinicians can use this new knowledge to prepare patients with depression for a 4- to 6-week lag before a major change occurs in their activation level, and subsequently, their symptoms of depression improve. This may encourage depressed individuals to persevere with treatment when they encounter difficulties increasing activity levels in the first phase of treatment. Clinicians and patients alike can be reassured that persistence with therapy will reduce depressive symptoms and lead to recovery.

The present study has several strengths. First, we used data from a well-powered RCT to ensure that the effect estimates from treatment allocation to activation level and depression could be readily and precisely interpreted as causal. The sample size calculated for the RCT was ample for the current mediation analysis, for which a sample size of at least 100 with at least three repeated observations per individual was considered appropriate (Curran et al., 2010). Second, the fortnightly measures added precision and the low dropout rate (0.20%) contributed to both precision and low risk of bias. Third, we adopted latent growth and simplex mediation modeling to estimate, as precisely as possible, the association between the mediator and depression while controlling for the within-participant change. According to criteria set by Lemmens et al. (2016), our work constitutes a high-quality mediation study.

Some limitations also warrant discussion. Notably, the mediator–outcome relationship could still have been confounded by a third unmeasured variable (e.g., cognition). In addition, we only included a single mediator in our model, limiting us to identifying activation as the mediator. Other working mechanisms correlated with activation level may have mediated part its effect, such as a change in cognition that may have preceded the reduction in depressive symptomatology. Aside from using the SCID-5 to assess unipolar depressive disorder before and after treatment, measurements in the RCT relied on self-reporting every 2 weeks. Thus, the assessments of activation level may not have been objective and may have missed a more nuanced dynamic (Folke et al., 2015). Moreover, lay counselors had no role in assessment of the participants and the effect of

change in activation level on depression outcomes was also not assessed by lay counselors and fully independently conducted from these counselors. Therefore, although some bias can never be fully excluded, it is unlikely bias explained the outcomes.

Future research must seek to replicate our findings with different control groups. It should have a more temporally sensitive design (e.g., experience sampling method), more objective measures of activation, and include other variables (e.g., cognitive variables). Such research may also benefit from experimental manipulation of mediator levels (e.g., component analysis) (Emmelkamp et al., 2014) and micro-trials using experimental designs, such as RCTs with temporally sensitive designs (Brouwer et al., 2020; Slofstra et al., 2018), to reach firm (causal) conclusions (Lorenzo-Luaces, Lemmens, Keefe, Cuijpers, & Bockting, 2021).

Conclusion

This study provides evidence that a change in activation level underpinned the effects of a guided internet-based intervention using behavioral activation to treat depression. In a large-scale RCT, it took 4 and 6 weeks to change activation levels and depressive symptoms, respectively. More studies are still required to support these findings and optimize treatment strategies.

PsychOpen GOLD

Supplementary Materials

Detailed results for the mediation examination in simplex models with depression as mediator were provided in the Supplementary Materials (for access see Index of Supplementary Materials below).

Index of Supplementary Materials

Fu, Z., Burger, H., Arjadi, R., Nauta, M. H., & Bockting, C. L. H. (2021). *Supplementary materials to "Explaining the efficacy of an internet-based behavioral activation intervention for major depression: A mechanistic study of a randomized-controlled trial"* [Additional results]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.5092

References

American Psychiatry Association. (2015). *Depressive disorders: DSM-5 selections.* American Psychiatric Publishing.

Arjadi, R., Nauta, M. H., Scholte, W. F., Hollon, S. D., Chowdhary, N., Suryani, A. O., Uiterwaal, C. S. P. M., & Bockting, C. L. (2018). Internet-based behavioural activation with lay counsellor support versus online minimal psychoeducation without support for treatment of depression: A randomised controlled trial in Indonesia. *The Lancet Psychiatry, 5*(9), 707-716. https://doi.org/10.1016/S2215-0366(18)30223-2

Bockting, C. L., & Van Valen, E. (2015). *Doe en Voel.* The Netherlands.

Brouwer, M. E., Molenaar, N. M., Burger, H., Williams, A. D., Albers, C. J., Lambregtse-van den Berg, M. P., & Bockting, C. L. (2020). Tapering antidepressants while receiving digital preventive cognitive therapy during pregnancy: An experience sampling methodology trial. *Frontiers in Psychiatry, 11*, Article 1116. https://doi.org/10.3389/fpsyt.2020.574357

Carlbring, P., Hägglund, M., Luthström, A., Dahlin, M., Kadowaki, Å., Vernmark, K., & Andersson, G. (2013). Internet-based behavioral activation and acceptance-based treatment for depression: A randomized controlled trial. *Journal of Affective Disorders, 148*(2-3), 331-337. https://doi.org/10.1016/j.jad.2012.12.020

Carroll, H. A., Hook, K., Perez, O. F. R., Denckla, C., Vince, C. C., Ghebrehiwet, S., Ando, K., Touma, M., Borba, C. P. C., Fricchione, G. L., & Henderson, D. C. (2020). Establishing reliability and validity for mental health screening instruments in resource-constrained settings: Systematic review of the PHQ-9 and key recommendations. *Psychiatry Research, 291*, Article 113236. https://doi.org/10.1016/j.psychres.2020.113236

Cheong, J. W., MacKinnon, D. P., & Khoo, S. T. (2003). Investigation of mediational processes using parallel process latent growth curve modeling. *Structural Equation Modeling, 10*(2), 238-262. https://doi.org/10.1207/S15328007SEM1002_5

Cuijpers, P., Van Straten, A., & Warmerdam, L. (2007). Behavioral activation treatments of depression: A meta-analysis. *Clinical Psychology Review, 27*(3), 318-326. https://doi.org/10.1016/j.cpr.2006.11.001

PsychOpen GOLD

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development, 11*(2), 121-136. https://doi.org/10.1080/15248371003699969

Curry, J. F., & Meyer, A. E. (2016). Can less yield more? Behavioral activation for adolescent depression. *Clinical Psychology: Science and Practice, 23*(1), 62-65. https://doi.org/10.1111/cpsp.12141

Dimidjian, S., Goodman, S. H., Sherwood, N. E., Simon, G. E., Ludman, E., Gallop, R., Welch, S. S., Boggs, J. M., Metcalf, C. A., Hubley, S., Powers, J. D., & Beck, A. (2017). A pragmatic randomized clinical trial of behavioral activation for depressed pregnant women. *Journal of Consulting and Clinical Psychology, 85*(1), 26-36. https://doi.org/10.1037/ccp0000151

Emmelkamp, P. M. G., David, D., Beckers, T., Muris, P., Cuijpers, P., Lutz, W., Andersson, G., Araya, R., Banos Rivera, R. M., Barkham, M., Berking, M., Berger, T., Botella, C., Carlbring, P., Colom, F., Essau, C., Hermans, D., Hofmann, S. G., Knappe, S., . . . Vervliet, B. (2014). Advancing psychotherapy and evidence-based psychological interventions. *International Journal of Methods in Psychiatric Research, 23*(S1), 58-91. https://doi.org/10.1002/mpr.1411

Ferster, C. B. (1973). A functional analysis of depression. *The American Psychologist, 28*(10), 857-870. https://doi.org/10.1037/h0035605

Ferster, C. B. (1981). A functional analysis of behavior therapy. In L. P. Rehm (Ed.), *Behavior therapy for depression: Present status and future directions* (pp. 181–196). Academic Press.

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). *Structured clinical interview for DSM-5—Research version* (SCID-5 for DSM-5, research version; SCID-5-RV). American Psychiatric Association.

Folke, F., Hursti, T., Tungström, S., Söderberg, P., Kanter, J. W., Kuutmann, K., . . . Ekselius, L. (2015). Behavioral activation in acute inpatient psychiatry: A multiple baseline evaluation. *Journal of Behavior Therapy and Experimental Psychiatry, 46*, 170-181. https://doi.org/10.1016/j.jbtep.2014.10.006

Forand, N. R., Barnett, J. G., Strunk, D. R., Hindiyeh, M. U., Feinberg, J. E., & Keefe, J. R. (2018). Efficacy of guided iCBT for depression and mediation of change by cognitive skill acquisition. *Behavior Therapy, 49*(2), 295-307. https://doi.org/10.1016/j.beth.2017.04.004

Gaynor, S. T., & Harris, A. (2008). Single-participant assessment of treatment mediators: Strategy description and examples from a behavioral activation intervention for depressed adolescents. *Behavior Modification, 32*(3), 372-402. https://doi.org/10.1177/0145445507309028

Goldsmith, K. A., MacKinnon, D. P., Chalder, T., White, P. D., Sharpe, M., & Pickles, A. (2018). Tutorial: The practical application of longitudinal structural equation mediation models in clinical trials. *Psychological Methods, 23*(2), 191-207. https://doi.org/10.1037/met0000154

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*(1), 549-576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Janssen, N. P., Hendriks, G.-J., Baranelli, C. T., Lucassen, P., Oude Voshaar, R., Spijker, J., & Huibers, M. J. (2020). How does behavioural activation work? A systematic review of the evidence on

potential mediators. *Psychotherapy and Psychosomatics, 90*(2), 85-93. https://doi.org/10.1159/000509820

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology, 3*, 1-27. https://doi.org/10.1146/annurev.clinpsy.3.022806.091432

Lazarus, A. A. (1972). Some reactions to Costello's paper on depression. *Behavior Therapy, 3*(2), 248-250. https://doi.org/10.1016/S0005-7894(72)80085-6

Lemmens, L. H., Müller, V. N., Arntz, A., & Huibers, M. J. (2016). Mechanisms of change in psychotherapy for depression: An empirical update and evaluation of research aimed at identifying psychological mediators. *Clinical Psychology Review, 50*, 95-107. https://doi.org/10.1016/j.cpr.2016.09.004

Lewinsohn, P. M. (1974). A behavioral approach to depression. In J. C. Coyne (Ed.), *Essential papers on depression* (pp. 150–172). New York University Press.

Lorenzo-Luaces, L., Lemmens, L. H. J. M., Keefe, J. R., Cuijpers, P., & Bockting, C. L. (2021). The efficacy of cognitive behavioral therapy for emotional disorders. In A. Wenzel (Ed.), *Handbook of cognitive behavioral therapy* (Vol. 1, pp. 51–89). American Psychological Association.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis.* Routledge.

MacKinnon, D. P., Cheong, J., & Pirlott, A. G. (2012). Statistical mediation analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 313–331). American Psychological Association. https://doi.org/10.1037/13620-018https://doi.org/10.1037/13620-018

Manos, R. C., Kanter, J. W., & Busch, A. M. (2010). A critical review of assessment strategies to measure the behavioral activation model of depression. *Clinical Psychology Review, 30*(5), 547-561. https://doi.org/10.1016/j.cpr.2010.03.008

Manos, R. C., Kanter, J. W., & Luo, W. (2011). The behavioral activation for depression scale–short form: Development and validation. *Behavior Therapy, 42*(4), 726-739. https://doi.org/10.1016/j.beth.2011.04.004

Martell, C. R., Dimidjian, S., & Herman-Dunn, R. (2013). *Behavioral activation for depression: A clinician's guide.* Guilford Press.

National Collaborating Centre for Mental Health (UK). (2010). *Depression: The treatment and management of depression in adults* (Updated ed.). British Psychological Society.

Moreno-Peral, P., Bellón, J. Á., Huibers, M. J., Mestre, J. M., García-López, L. J., Taubner, S., . . . Conejo-Cerón, S. (2020). Mediators in psychological and psychoeducational interventions for the prevention of depression and anxiety: A systematic review. *Clinical Psychology Review, 76*, Article 101813. https://doi.org/10.1016/j.cpr.2020.101813

Muthén, B. O., & Muthén, L. K. (2019). Mplus (Version 8.3) [Computer software]. Los Angeles, CA, USA.

Nasrin, F., Rimes, K., Reinecke, A., Rinck, M., & Barnhofer, T. (2017). Effects of brief behavioural activation on approach and avoidance tendencies in acute depression: Preliminary findings.

Behavioural and Cognitive Psychotherapy, 45(1), 58-72.
https://doi.org/10.1017/S1352465816000394

Pinheiro, J., & Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2020). *nlme: Linear and nonlinear mixed effects models* [Computer software]. Retrieved from
https://CRAN.R-project.org/package=nlme

Richards, D. A., Rhodes, S., Ekers, D., McMillan, D., Taylor, R. S., Byford, S., Barrett, B., Finning, K., Ganguli, P., Warren, F., Farrand, P., Gilbody, S., Kuyken, W., O'Mahen, H., Watkins, E., Wright, K., Reed, N., Fletcher, E., Hollon, S. D., . . . Woodhouse, R. (2017). Cost and outcome of behavioural activation (COBRA): A randomised controlled trial of behavioural activation versus cognitive-behavioural therapy for depression. *Health Technology Assessment, 21*(46), 1-366. https://doi.org/10.3310/hta21460

Rovner, B. W., Casten, R. J., Hegel, M. T., Massof, R. W., Leiby, B. E., Ho, A. C., & Tasman, W. S. (2014). Low vision depression prevention trial in age-related macular degeneration: A randomized clinical trial. *Ophthalmology, 121*(11), 2204-2211.
https://doi.org/10.1016/j.ophtha.2014.05.002

Santos, M. M., Rae, J. R., Nagy, G. A., Manbeck, K. E., Hurtado, G. D., West, P., . . . Kanter, J. W. (2017). A client-level session-by-session evaluation of behavioral activation's mechanism of action. *Journal of Behavior Therapy and Experimental Psychiatry, 54*, 93-100.
https://doi.org/10.1016/j.jbtep.2016.07.003

Santos, M. M., Ullman, J., Leonard, R. C., Puspitasari, A. J., Cook, J., & Riemann, B. C. (2019). Behavioral activation as a mechanism of change in residential treatment for mood problems: A growth curve model analysis. *Behavior Therapy, 50*(6), 1087-1097.
https://doi.org/10.1016/j.beth.2019.05.004

Silverstein, M., Cabral, H., Hegel, M., Diaz-Linhart, Y., Beardslee, W., Kistin, C. J., & Feinberg, E. (2018). Problem-solving education to prevent depression among low-income mothers: A path mediation analysis in a randomized clinical trial. *JAMA Network Open, 1*(2), Article e180334.
https://doi.org/10.1001/jamanetworkopen.2018.0334

Slofstra, C., Nauta, M. H., Bringmann, L. F., Klein, N. S., Albers, C. J., Batalas, N., . . . Bockting, C. L. (2018). Individual negative affective trajectories can be detected during different depressive relapse prevention strategies. *Psychotherapy and Psychosomatics, 87*(4), 243-245.
https://doi.org/10.1159/000489044

Stein, A. T., Carl, E., Cuijpers, P., Karyotaki, E., & Smits, J. A. J. (2021). Looking beyond depression: A meta-analysis of the effect of behavioral activation on depression, anxiety, and activation. *Psychological Medicine, 51*(9), 1491-1504. https://doi.org/10.1017/S0033291720000239

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1-67. https://doi.org/10.18637/jss.v045.i03

van Luenen, S., Kraaij, V., Spinhoven, P., Wilderjans, T. F., & Garnefski, N. (2019). Exploring mediators of a guided web-based self-help intervention for people with HIV and depressive symptoms: Randomized controlled trial. *JMIR Mental Health, 6*(8), Article e12711.
https://doi.org/10.2196/12711

PsychOpen GOLD

Weidberg, S., González-Roz, A., García-Fernández, G., & Secades-Villa, R. (2021). Activation level as a mediator between behavioral activation, sex, and depression among treatment-seeking smokers. *Addictive Behaviors, 114*, Article 106715. https://doi.org/10.1016/j.addbeh.2020.106715

PsychOpen GOLD

Research Articles

Check for updates

Lifetime Trauma History and Cognitive Functioning in Major Depression and Their Role for Cognitive-Behavioral Therapy Outcome

Lena Schindler[1] (iD) , Tobias Stalder[2] (iD) , Clemens Kirschbaum[1], Franziska Plessow[3] (iD) ,

Sabine Schönfeld[1,4], Jürgen Hoyer[5] (iD) , Sebastian Trautmann[5,6] (iD) , Kerstin Weidner[7],

Susann Steudte-Schmiedgen[1,7] (iD)

[1] *Faculty of Psychology, Technische Universität Dresden, Dresden, Germany.* **[2]** *Department Erziehungswissenschaften und Psychologie, Universität Siegen, Siegen, Germany.* **[3]** *Neuroendocrine Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.* **[4]** *Department of Psychology, Lund University, Lund, Sweden.* **[5]** *Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Dresden, Germany.* **[6]** *Department of Psychology, Medical School Hamburg, Hamburg, Germany.* **[7]** *Department of Psychotherapy and Psychosomatic Medicine, Medical Faculty Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany.*

Corresponding Author: Susann Steudte-Schmiedgen, Department of Psychotherapy and Psychosomatic Medicine, Medical Faculty Carl Gustav Carus, Technische Universität Dresden, Fetscherstraße 74, 01307 Dresden, Germany. Tel.: +49 351 458 3634. E-mail: susann.schmiedgen@tu-dresden.de

Supplementary Materials: Materials [see Index of Supplementary Materials]

Abstract

Background: While cognitive-behavioral therapy (CBT) is the gold-standard psychological treatment for major depression (MD), non-response and lacking stability of treatment gains are persistent issues. Potential factors influencing treatment outcome might be lifetime trauma history and possibly associated primarily prefrontal-cortex- and hippocampus-dependent cognitive alterations.

Method: We investigated MD and healthy control participants with (MD+T+, n = 37; MD-T+, n = 39) and without lifetime trauma history (MD+T-, n = 26; MD-T-, n = 45) regarding working memory, interference susceptibility, conflict adaptation, and autobiographical memory specificity. Further, MD+T+ (n = 21) and MD+T- groups (n = 16) were re-examined after 25 CBT sessions, with MD-T- individuals (n = 34) invited in parallel in order to explore the stability of cognitive

alterations and the predictive value of lifetime trauma history, cognitive functioning, and their interaction for treatment outcome.

Results: On a cross-sectional level, MD+T+ showed the highest conflict adaptation, but MD+T- the lowest autobiographical memory specificity, while no group differences emerged for working memory and interference susceptibility. Clinical improvement did not differ between groups and cognitive functioning remained stable over CBT. Further, only a singular predictive association of forward digit span, but no other facets of baseline cognitive functioning, lifetime trauma history, or their interaction with treatment outcome emerged.

Discussion: These results indicate differential roles of lifetime trauma history and psychopathology for cognitive functioning in MD, and add to the emerging literature on considering cognitive, next to clinical remission as a relevant treatment outcome.

Keywords

Highlights

- Conflict adaptation was highest in MD with lifetime trauma history.
- Autobiographical memory specificity was lowest in MD without lifetime trauma history.
- No differential treatment response was found in MD with and without lifetime trauma history.
- There were no changes of cognitive functioning over CBT, irrespective of lifetime trauma history.
- Only singular predictive value of cognitive functioning for CBT success emerged.

Meta-analyses suggest cognitive-behavioral therapy (CBT) as the gold-standard psychological treatment for major depression (MD; e.g., Barth et al., 2013; Cuijpers et al., 2014), a condition characterized by depressed mood and loss of motivation together with behavioral alterations such as reduced activity and disturbed sleep (Diagnostic and Statistical Manual of Disorders – Fifth Edition; American Psychiatric Association, 2013). However, a substantial patient subgroup fails to achieve clinically significant symptom improvement, with non-response and dropout rates of approximately 34% and 25%, respectively (for meta-analytic data, see Cuijpers et al., 2014; Hans & Hiller, 2013). This highlights the need to enhance our understanding of factors associated with psychopathology and treatment outcome, allowing an optimization of CBT effects and reduction of dropout rates. Here, trauma history is frequently discussed, defined as exposure to actual or threatened death, serious injury, or sexual violence (American Psychiatric Association, 2013). Particularly for childhood trauma in MD, associations with poorer therapy response, longer remission time, and greater need for additional medication are relatively well-researched (for review and meta-analytic data, see Nanni et al., 2012; Nemeroff,

2016; Teicher & Samson, 2013). Notably, lifetime trauma, including childhood, adulthood, or both types of trauma, has been far less well studied, except for one study suggesting negative associations of both childhood and adulthood adversity with therapy outcome in MD (Miniati et al., 2010).

Importantly, lifetime trauma history is assumed to co-occur with neurobiological (e.g., Kolassa & Elbert, 2007; Sherin & Nemeroff, 2011) and cognitive alterations (e.g., Vasterling & Arditte Hall, 2018). However, data on this and its influence on therapy success in the context of MD and trauma is sparse. In particular, primarily prefrontal-cortex- and hippocampus-dependent functioning have received attention (McIntyre et al., 2013; Rock et al., 2014; Snyder, 2013; Snyder & Hankin, 2019). Regarding the former, of importance might be working memory (WM) as a facet of executive functioning (EF) relevant for temporal maintenance (usually assessed by the repetition of a list of numbers) and manipulation (usually assessed by the repetition of a list of numbers in a backward fashion) of content necessary for current tasks (Diamond, 2013). Accumulating evidence suggests impaired WM in patients with MD (for reviews, see Snyder, 2013; Snyder & Hankin, 2019). Further, one study reported childhood trauma to predict performance in a compound WM score of information maintenance and manipulation in both patients with MD and healthy controls (Saleh et al., 2017), but another found no WM differences with respect to information maintenance or manipulation in patients with MD with or without childhood trauma (Dannehl et al., 2017).

An EF domain considered to be even more impaired in MD (e.g., Snyder, 2013; Snyder & Hankin, 2019) is the ability to suppress irrelevant and/or interfering response tendencies while pursuing mentally represented goals (i.e., inhibitory control, Diamond, 2013). Typically, this is studied via the well-known *Simon task* (Simon, 1990), where the inhibition of a response following a task-irrelevant visual stimulus is necessary as a different response is required. The resulting additional performance costs (i.e., slower reaction times [RTs] and/or increased percentages of error [PEs]) compared to trials with matching automatic and required tendencies comprise the so-called Simon effect as a measure of interference susceptibility (Simon, 1990). After response conflicts, inhibitory control is typically increased, leading to a decreased impact of task-irrelevant information compared to trials not following conflicts. The resulting difference in the Simon effect is termed conflict adaptation (Botvinick et al., 2001). In MD, particularly this conflict adaptation according to task demands is suggested to be increased (van Steenbergen et al., 2012). Notably, previous work from our group revealed similar findings for patients with posttraumatic stress disorder (PTSD) and, albeit less clearly, trauma-exposed controls (Schindler et al., 2020; Steudte-Schmiedgen et al., 2014), encouraging research on the interaction of trauma and MD.

Of note, there is an abundance of studies suggesting not only EF, but also mainly hippocampally-driven overgeneral memory retrieval (OGM) to be a central correlate of MD (for meta-analytic data, see, e.g., Sumner et al., 2010). This increased recall of over-

general (e.g., "I am happy when meeting friends") instead of specific autobiographical memories (e.g., "I was happy on July 8 when I met friends"; Williams et al., 2007) is also prevalent in PTSD, with trauma history a potential shared mechanism (Moore & Zoellner, 2007; Ono et al., 2016; Sumner et al., 2010; Williams et al., 2007). However, previous contrasting of trauma-exposed and non-exposed individuals with MD (notably, again only focusing on childhood trauma) provided mixed results, with one study finding OGM only in trauma-exposed (Aglan et al., 2010) and another only in non-exposed individuals (Kuyken et al., 2006).

Next to these cross-sectional findings of certain alterations of EF and autobiographical memory domains, and the possible mediating role of trauma history in MD, it is plausible to assume that such alterations show significant change over psychotherapy. However, the vast majority of studies could not detect any changes of the cognitive alterations described above over psychotherapy/combined psycho- and pharmacotherapy (for WM, see, e.g., Beblo et al., 1999; Lahr et al., 2007; for inhibitory control, see, e.g., Schmid & Hammar, 2013; but Ajilchi et al., 2016; for OGM, see, e.g., Peeters et al., 2002). Thus, a current meta-analysis (Bernhardt et al., 2019) rather support the suggestions from previous reviews (e.g., Bernhardt et al., 2019; Köhler et al., 2015; Moore & Zoellner, 2007; Snyder & Hankin, 2019) of high stability of such alterations even after clinical remission, with improvements not exceeding task-specific practice effects. While previous data on cognitive markers as predictors for clinical outcome in the context of pharmacotherapy is promising (Groves et al., 2018), research on CBT is outstanding, except for initial studies suggesting a predictive value of enhanced autobiographical memory specificity (Sumner et al., 2010), but not interference susceptibility (Goodkind et al., 2016). However, while lifetime trauma history is assumed to be associated with both therapy outcome (e.g., Nemeroff, 2016; Teicher & Samson, 2013) and cognitive alterations (e.g., Vasterling & Arditte Hall, 2018) in MD, a combined investigation is still pending.

Hence, the aim of the current study was to examine (i) lifetime trauma history and (ii) facets of cognitive functioning (i.e., WM, interference susceptibility, conflict adaptation, and OGM) as well as (iii) their interaction in the context of MD symptomatology and therapy success. Due to the inconclusive literature on the interplay of lifetime trauma history and MD for cognitive functioning, our first step was to study respective baseline alterations in MD and healthy control participants with (MD+T+, n = 37; MD-T+, n = 39)[1] and without lifetime trauma history (MD+T-, n = 26; MD-T-, n = 45). Specifically, we aimed to (1) investigate whether the previously found effect of lifetime trauma history on conflict adaptation (Schindler et al., 2020; Steudte-Schmiedgen et al., 2014) is also visible in MD and (2) shed light on the conflicting evidence regarding OGM (Aglan

---

1) MD+T+ = patients with MD with lifetime trauma history; MD+T- = patients with MD without lifetime trauma history; MD-T+ = patients without MD with lifetime trauma history; MD-T- = patients without MD and without lifetime trauma history.

et al., 2010; Kuyken et al., 2006). Further, we assessed clinical and cognitive treatment outcome under consideration of lifetime trauma history by re-examining patients with MD with (MD+T+, *n* = 21) and without lifetime trauma history (MD+T-, *n* = 16) after 25 CBT sessions. In order to account for practice effects, non-traumatized healthy control individuals (MD-T-, *n* = 34) were re-invited in parallel. Here, we hypothesized (3) poorer treatment outcome for MD+T+ than for MD+T- individuals. Based on recent meta-analytic evidence (Bernhardt et al., 2019), we aimed to examine whether we could confirm the finding of (4) no changes of cognitive functioning over CBT, irrespective of lifetime trauma history, also for the tasks studied here. On a last note, we aimed to (5) exploratorily study the predictive value of cognitive functioning for CBT outcome.

Method

Participants and Procedures

Recruitment was conducted within the outpatient unit of the Institute of Clinical Psychology and Psychotherapy of the Technische Universität Dresden, as well as via flyers and local advertisements. Individuals were included in the study if they were aged between 18 and 65 years, not pregnant (women), and did not report any severe physical diseases (e.g., cancer, encephalopathy) over the past five years. Further exclusion criteria concerned hair-related and endocrine factors due to biomarker analyses reported elsewhere (e.g., glucocorticoid medication; Steudte et al., 2013; Steudte-Schmiedgen et al., 2014). The presence of MD and any other DSM-IV (American Psychiatric Association, 2007) mental disorders was assessed using the standardized Munich Composite International Diagnostic Interview (*DIA-X/M-CIDI;* Wittchen & Pfister, 1997) conducted by therapists of the outpatient unit or trained research team members and confirmed by an experienced clinical psychologist. Twenty-eight participants from the MD groups showed psychiatric comorbidities within the last 12 months (one: *n* = 15, two: *n* = 8, three or more: *n* = 5). Those encompassed specific (*n* = 12) or social phobia (*n* = 13), somatoform disorders (*n* = 6), panic disorder with or without agoraphobia (*n* = 8), generalized anxiety (*n* = 3), obsessive-compulsive (*n* = 2), adjustment (*n* = 2), or eating disorders (*n* = 1).

An assignment to the MD groups was based on a current primary 12-month MD diagnosis and no 12-month diagnosis of substance abuse or dependence (except for nicotine) or any lifetime diagnoses of psychosis, severe depressive disorder with psychotic symptoms, or bipolar disorder. Notably, individuals meeting the lifetime diagnostic criteria for PTSD were also excluded from the study, in order to allow insights into the role of lifetime trauma exposure *per se* for cognitive functioning in MD. Participants were included in the control group if they did not report any lifetime mental disorders according to the *DIA-X/M-CIDI* stem questions and the *Mini International Neuropsychiatric Interview* (*M.I.N.I.;* Sheehan et al., 1998). Participants were further classified as exposed or

non-exposed to lifetime trauma based on the *Posttraumatic Stress Diagnostic Scale* (*PDS;* Ehlers, Steil, Winter, & Foa, 1996). For an allocation to the T+ groups, both the "objective" A1 ("actual or threatened death or serious injury or a threat to the physical integrity of oneself or others") and the "subjective" A2 criterion ("intense fear, helplessness or horror") had to be met, following the DSM-IV requirements that qualify life events as traumatic (American Psychiatric Association, 2007). The control groups are the same as in the parallel study on patients with PTSD (Schindler et al., 2020). For further participant characteristics, see Table 1 and Supplementary Materials (type of lifetime trauma history).

CBT for MD groups was conducted within the outpatient unit based on established manuals (Hautzinger, 1998, 2008) and supervised by experienced therapists. After 25 sessions, MD+T+ and MD+T- patients were re-invited for clinical and cognitive testing, with MD-T- participants being contacted in a parallel fashion (no difference regarding months between assessments: M = 13.5, SD = 3.86; M = 11.56, SD = 4.03; and M = 14.76, SD = 6.97, respectively; $F(2, 68)$ = 1.78, p = .177, η_p^2 = .05). Among the 63 patients with MD examined at baseline, 6 (9.5%) were only interested in the cross-sectional study, 16 (25.4%) dropped out of CBT, and 41 (65.1%) completed therapy. Between those who dropped out of CBT and those who did not, no differences emerged regarding pre-treatment clinical variables (all ps ≥ .219). All participants had provided written informed consent before study inclusion. The study protocol was approved by the ethics committee of the Technische Universität Dresden (EK 65022010) and conducted in accordance with the Declaration of Helsinki.

Clinical and Psychological Measures

Self-developed questionnaires were applied for socio-demographic (age, sex, education status) and health-related variables (smoking, chronic physical diseases, regular medication intake). Depressive symptoms over the previous two weeks were assessed via the *Beck Depression Inventory-II* (*BDI-II,* Hautzinger et al., 2006). The *PDS* (Ehlers et al., 1996) provided insights into the presence or absence of lifetime trauma history and the severity of symptoms associated with posttraumatic stress according to *DSM-IV* criteria. The *Trauma History Questionnaire* (*THQ,* Maercker, 2002) provided an overview over number and frequency of potentially traumatic events fulfilling the *DSM-IV* A1, but not A2 criterion (Hooper et al., 2011). Furthermore, to obtain information on the severity of childhood maltreatment (irrespective of fulfilling *DSM-IV* A criteria), the *Childhood Trauma Questionnaire* (*CTQ,* Gast et al., 2001) was used. At follow-up, patients with MD additionally received the revised version of the *Questionnaire of Changes in Experience and Behavior* (*Veränderungsfragebogen des Erlebens und Verhaltens VEV-R;* Zielke & Kopf-Mehnert, 2001). This allowed a classification of patient-evaluated therapy effects via 42 items of opposite polarity (e.g., "Compared with the time prior to initiation of

Table 1

Baseline Demographic, Health-Related, and Clinical Characteristics of Patients With Major Depression With (MD+T+) and Without (MD+T−) as well as Controls With (MD−T+) and Without (MD−T−) Lifetime Trauma History

Participants' characteristics	MD+T+ (n = 37)	MD+T− (n = 26)	MD−T+ (n = 39)	MD−T− (n = 45)	Test statistic	p
Demographics						
Age (M, SD)	37.59 (11.91)	39.27 (11.8)	41.46 (12.82)	35.31 (13.8)	$F(3, 143) = 1.71$	.167
Female sex (%)	27 (73)	17 (65.4)	32 (82.1)	38 (84.4)	$X^2_3 = 4.34$	.227
Highest educational status					$X^2_{12} = 22.26$	.035
Academic degree (%)	7 (18.9)	2 (8)[a]	16 (41)	11 (24.4)		
Professional training/college degree (%)	10 (27)	12 (48)[a]	13 (33.3)	15 (33.3)		
A level (%)	13 (35.1)	5 (20)[a]	7 (17.9)	16 (35.6)		
High school diploma/lower (%)	7 (18.9)	6 (24)[a]	3 (7.7)	3 (6.7)		
Smoking (%)	8 (21.6)	11 (42.3)	4 (10.3)	10 (22.2)	$X^2_3 = 9.24$	.026
Physical disease (%)	19 (51.4)	11 (42.3)	17 (43.6)	12 (26.7)	$X^2_3 = 5.58$	.134
Regular medication (%)	21 (56.8)	16 (61.5)	11 (28.2)	8 (17.8)	$X^2_3 = 21.02$	< .001
Psychiatric (%)[b]	16 (43.2)	12 (46.2)	0	0		
Non-psychiatric (%)	10 (27)	8 (30.8)	11 (28.2)	8 (17.8)		
PDS score (M, SD)	12.4 (11.45)	n.a.	4.56 (6.18)	n.a.	$t_{54.72} = 3.68$	.001
THQ number of A1 traumatic events (M, SD)	4.3 (2.5)	2.81 (2.08)	4.21 (3.17)	1.33 (1.21)	$F(3, 143) = 14.87$	< .001[I]
THQ frequency of A1 traumatic events (M, SD)	7.33 (6.71)[c]	6.12 (5.1)	7.26 (8.42)	2.64 (3.79)	$F(3, 143) = 5.25$	.002[II]
CTQ score (M, SD)	39.59 (10.48)	37.45 (10.83)	37.94 (13)	29.02 (4.87)	$F(3, 143) = 9.43$	< .001[III]
BDI-II score (M, SD)	20.78 (9.11)	22.15 (8.75)	5.1 (6.8)	4.52 (0.67)	$F(3, 143) = 66.66$	< .001[IV]

Note. PDS = Posttraumatic Stress Diagnostic Scale; THQ = Trauma History Questionnaire; CTQ = Childhood Trauma Questionnaire; BDI-II = Beck Depression Inventory-II.

[a]refers to n = 25. [b]Included antidepressants (n = 27), anticonvulsives (n = 3), neuroleptics (n = 2), sedatives (n = 2). [c]refers to n = 36.
[I]MD+T+ = MD−T+ > MD−T− in between. [II]MD+T+ = MD−T+ > MD−T− (ps ≤ .006), with MD+T− in between. [III]MD+T+ = MD−T+ = MD−T+ > MD−T− (ps ≤ .005). [IV]MD+T+ = MD+T−>MD−T+ = MD−T− (ps ≤ .001).

therapy, I feel more relaxed/no change/more tense.") into three categories (i.e., symptom improvement, no change, and worsening).

Cognitive Tasks

WM was examined using the *Wechsler Memory Scale* digit span task (Wechsler, 1997). Participants repeated a series of numbers read out loud by the experimenter in a forward (information maintenance) or backward fashion (information manipulation). Interference susceptibility and conflict adaptation were assessed by a number version of the *Simon task* (Fischer et al., 2008). In brief, participants categorized the numbers 1 to 9, except 5, as smaller or larger than five by pressing a left (Alt) or right (Alt Gr) key on a QWERTZ keyboard with their left or right index finger, respectively. Although task-irrelevant, stimulus location automatically facilitates the pressing of the corresponding response button, either in accordance, or in conflict with the required action, resulting in compatible and incompatible trials, respectively. The resulting difference in RTs and PEs comprises the Simon, and the typical reduction of interference susceptibility after conflict trials the conflict adaptation effect (Botvinick et al., 2004; Simon, 1990). Participants completed a 16-trial practice, followed by three 64-trial test blocks, resulting in 192 test trials (for further details, see Schindler et al., 2020; Steudte-Schmiedgen et al., 2014). Indices for interference susceptibility (I − C) and conflict adaptation [(cI − cC) − (iI − iC)] (lowercase letters: compatibility of the previous, uppercase letters: compatibility of the current trial, larger values indicating more pronounced effects) were calculated (van Steenbergen et al., 2010).

Autobiographical memory specificity was assessed via the standardized *Autobiographical Memory Test* (Williams & Broadbent, 1986). Participants were instructed to read words out loud (practice phase: three neutral words, testing phase: five positive and five negative words in a pseudo-randomized order, starting with a positive word and alternating valence) and briefly describe a related specific autobiographical memory. The words were randomly chosen from a word pool from a previous study (Schönfeld & Ehlers, 2006) matched for word frequency, emotionality, imagery, and pleasantness (apart from positive words rated as more pleasant than negative ones; Hager & Hasselhorn, 1994), with different sets used at baseline and follow-up. Answers were tape-recorded, transcribed and coded by trained research assistants. As an outcome variable, the number of specific memories was used, defined as having happened at a particular place and time more than one week ago and having lasted for one day or less. If no answer was provided within 30 seconds, the trial was considered an omission. For assessing inter-rater-reliability, a second, independent rater re-assessed a random sample (10%) of the tape-recorded sequences, resulting in κ = .76 for the baseline and κ = .82 for the follow-up assessment.

Statistical Analyses

Analyses were conducted via SPSS for Windows, version 25 (IBM, Armonk, NY), R (R Core Team, 2017), and STATA 15.1 (StataCorp LLC, 2017). Cross-sectional group comparisons were carried out via univariate analyses of variance (ANOVAs; continuous variables) and X^2 contingency tables (dichotomous variables). For the Simon task, the first trial of each block (1.6%), posterror trials (3%), target repetitions (11.3%), and, for RT analyses, error trials (3%) were excluded. AMT data from one MD+T- and one MD-T- participant were missing.

For longitudinal analyses, as a first step, participants from the MD+T+, MD+T-, and MD-T- groups with available longitudinal data were re-examined regarding baseline demographic and clinical differences. Simon task data from one MD+T+ and three MD-T- participants, and AMT data from one MD+T- and two MD-T- participants were missing. Again, the first trial of each block (1.6%), posterror trials (baseline: 2.8%, follow-up: 2.7%), target repetitions (baseline: 11.5%, follow-up: 10.9%) and, for RT analyses, error trials (baseline: 2.8%, follow-up: 2.7%) were excluded. Repeated-measures ANOVAs with time [2, baseline vs. follow-up] as within-subject and group [3, MD+T+ vs. MD+T- vs. MD-T-] as between-subject factor were applied to assess clinical and cognitive changes over CBT.

Exploratory linear/logistic regression analyses were conducted for examining the predictive value of lifetime trauma history (*PDS*; yes/no) for changes of depressive symptom severity (*BDI-II*) and dropout from care as core outcome measures, respectively. Due to the small sample size for the longitudinal analyses, and the high correlations between depressiveness (*BDI-II*) and the subjectively evaluated therapy effects (*VEV-R*, $r = -.66$, $p < .001$) at the follow-up assessment, we decided to omit the *VEV-R* from the predictive analyses. For the *BDI-II*, a change score was computed by subtracting baseline from follow-up values, and baseline values were included as a covariate to the regression analyses. As a second step, baseline cognitive performance (centered around the mean to avoid multicollinearity issues), and, as a third, the interaction of lifetime trauma history (yes/no) and baseline cognitive performance were added to the model.

Whenever hypothesis testing referred to one major cognitive domain (i.e., EF and learning/memory) and were not exploratory in nature, Holm-Bonferroni correction (Holm, 1979) for family-wise error (FWER) per respective domain was applied. As the assumptions of conventional GLMs (ANOVA, linear regression) are frequently violated in psychological data possibly leading to poor power and inaccurate effect sizes (Field & Wilcox, 2017), we repeated hypothesis testing using robust regressions. These drop GLM assumptions by using a robust sandwich estimation of standard errors, down-weighting observations with large residuals, and omitting outlying residuals (Royall, 1986). Predictive analyses were repeated using mixed-effects regressions with random intercept parameter addressing regression to the mean, which can otherwise yield biased results (Oberg & Mahoney, 2007). However, due to the higher prevalence and familiarity of

conventional GLMs in the field, whenever both analyses yielded the same results, conventional GLMs were reported.

Results

Sample Characteristics, Clinical Symptomatology, and Baseline Cognitive Functioning

The groups were well-matched regarding age, sex, and physical diseases (all $ps \geq .134$, see Table 1). However, group differences emerged for educational status ($X^2_{12} = 22.26$, $p = .035$) and smoking ($X^2_3 = 9.24$, $p = .026$). Furthermore, both clinical groups reported higher medication intake than the non-clinical ones ($X^2_3 = 21.02$, $p < .001$), mainly driven by psychiatric medication. However, including these variables as covariates did not change the cross-sectional results. For depressive symptom severity (*BDI-II*), both MD+T- and MD+T+ individuals reported higher levels than the control groups (all $ps \leq .001$), with post-hoc analyses indicating no difference between them. For number and frequency of *DSM-IV* A1 traumatic events, both MD+T+ and MD-T+ scored higher than MD-T- individuals, with MD+T- individuals in between (*THQ*, all $ps \leq .006$). For the severity of childhood maltreatment, both MD+ groups as well as the MD-T+ participants scored higher than the MD-T- group (*CTQ*, all $ps \leq .005$).

No group differences emerged for forward, backward, and overall digit span (all $ps \geq .283$, see Table 2). For the Simon task, groups differed regarding conflict adaptation of median RTs with a medium effect size, $F(3, 143) = 3.23$, $p = .024$, $\eta^2_p = .063$, 90% CI [0, .12], see Figure 1), with higher levels in MD+T+ compared to MD-T- individuals ($p = .017$) and no other differences (all $ps \geq .43$). Neither for conflict adaptation of mean PEs, nor for interference susceptibility did group differences emerge (all $ps \geq .424$). Regarding OGM, for positive and negative words and the overall score, MD+T- participants scored lower than both MD-T+ and MD-T- ones with, again, medium effect sizes (all $ps \leq .002$), and no other differences (all $ps \geq .118$).

While OGM results remained stable after Holm-Bonferroni correction for FWER, the group difference for conflict adaptation of median RTs lost statistical significance ($p = .168$). Applying robust regressions did not considerably change the results, except for the difference between MD-T+ and MD-T- participants regarding conflict adaptation of median RTs and the interference effect of median RTs emerging as non-significant trends, $\beta = -15.2$, 95% CI [-30.7, 0.2], $p = .053$ and $\beta = -11.5$, 95% CI [-24.2, 1.2], $p = .076$).

Table 2

Baseline Working Memory in the Digit Span Task, Interference and Conflict Adaptation Effects of Means of Median Reaction Times (RTs) and Percentage of Errors (PE) in the Simon Task, and Autobiographical Memory Specificity in the Autobiographical Memory Test (AMT) of Patients With Major Depression With (MD+T+) and Without (MD+T-) as well as Controls With (MD-T+) and Without (MD-T-) Lifetime Trauma History

| Facet of cognitive functioning	MD+T+ (n = 37)	MD+T- (n = 26)	MD-T+ (n = 39)	MD-T- (n = 45)	Test statistic	p	η²p 90% CI [LL, UL]	Adjusted p (Holm-Bonferroni correction)
Digit Span								
Total	17.57 (3.52)	17.77 (3.82)	17.33 (3.74)	18.62 (3.03)	$F_{(3, 143)} = 1.1$	.351	.023 [0, .06]	1
Forward	9.84 (1.94)	9.81 (1.98)	9.72 (1.75)	10.13 (1.7)	$F_{(3, 143)} = 0.41$	.746	.009 [0, .03]	1
Backward	7.73 (2.06)	7.96 (2.36)	7.62 (2.4)	8.49 (2.14)	$F_{(3, 143)} = 1.28$	.283	.026 [0, .07]	1
Simon task: RT								
Interference effect	27.32 (29.05)	24.88 (28.2)	31.87 (28.54)	21.99 (24.81)	$F_{(3, 143)} = 0.94$	.424	.019 [0, .05]	1
Conflict adaptation effect	68.93 (42.42)	51.15 (44.22)	57.74 (39.49)	43.14 (28.83)	$F_{(3, 143)} = 3.23$	.024[I]	.063 [0, .12]	.168
Simon task: PE								
Interference effect	1.83 (3.39)	1.73 (3.57)	2.05 (4.48)	1.32 (3.57)	$F_{(3, 143)} = 0.27$	.846	.006 [0, .02]	1
Conflict adaptation effect	5.23 (5.88)	5 (5.79)	5.26 (7.12)	4.6 (6.42)	$F_{(3, 143)} = 0.1$	.962	.002 [0, 0]	1
AMT number of specific memories								
Total	5.68 (2.72)	4.36 (2.46)[a]	6.69 (2.18)	6.73 (1.86)[b]	$F_{(3, 141)} = 7.16$	< .001[II]	.132 [.05, .21]	< .001
Positive cues	2.92 (1.44)	2.4 (1.35)[a]	3.49 (1.36)	3.57 (1.23)[b]	$F_{(3, 141)} = 5.2$	.002[III]	.1 [.02, .17]	.002
Negative cues	2.76 (1.54)	1.96 (1.34)[a]	3.21 (1.2)	3.16 (1.16)[b]	$F_{(3, 141)} = 5.72$	.001[IV]	.11 [.03, .18]	.002

Note. Data are presented as *M (SD)*. CI = confidence interval; *LL* = lower level; *UL* = upper level; RT = median reaction times; PE = percentages of error.

[a]refers to *n* = 25. [b]refers to *n* = 44.

[I]MD+T- > MD-T- (*p* = .017). [II]MD+T- < MD-T+ = MD+T+ = MD-T- (*ps* ≤ .001). [III]MD+T- < MD-T+ = MD+T+ = MD-T- (*ps* ≤ .011). [IV]MD+T- < MD-T+ = MD+T+ = MD-T- (*ps* ≤ .002).

Figure 1

Mean (± SEM) (A) Conflict Adaptation of Median RTs (Simon Task) and (B) Specificity of Autobiographical Memory (Autobiographical Memory Test) of Patients With Major Depression With (MD+T+) and Without (MD+T-) as well as Controls With (MD-T+) and Without (MD-T-) Lifetime Trauma History at Baseline
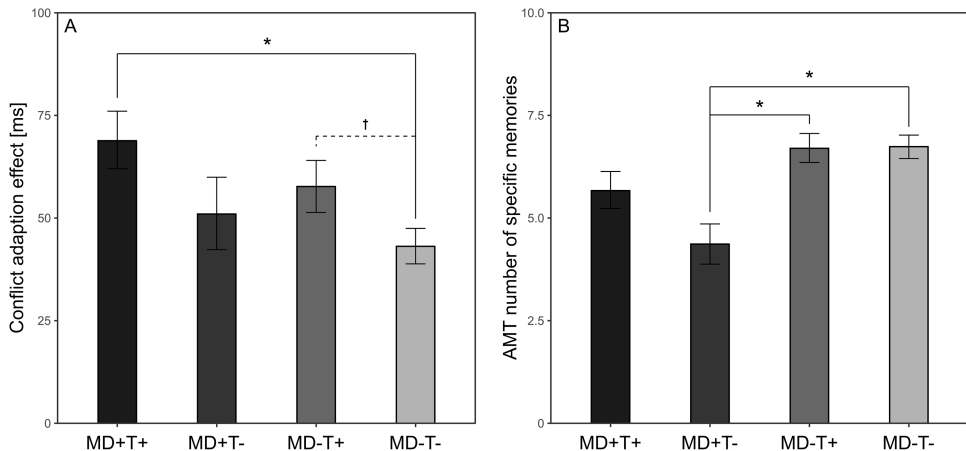
Note. $*p < .05$, $†p < .10$, dotted lines indicate differentiating results between general linear and robust models.

Clinical and Cognitive Treatment Outcome Under Consideration of Lifetime Trauma History

MD+T+ ($n = 21$), MD+T- ($n = 16$), and MD-T- participants ($n = 34$) available for longitudinal analyses did not differ regarding baseline demographic/health-related characteristics (all $ps ≥ .136$, see Supplementary Materials), except for higher medication intake in both MD groups ($X^2_2 = 13.9$, $p = .001$). However, including it as a covariate did not affect the longitudinal results. MD+T+ individuals reported a higher number of *DSM-IV* A1 traumatic events (*THQ*) than MD+T- ones, which, in turn, reported more than MD-T- individuals (all $ps ≤ .036$). With respect to their frequency (*THQ*), as well as for childhood maltreatment severity (*CTQ*), both MD+ groups scored higher than the MD-T- one (all $ps ≤ .035$ and all $ps ≤ .002$, respectively).

Notably, while CBT led to substantial clinical improvements, MD+T+ and MD+T- individuals did not differ regarding depressive symptom changes (*BDI-II*), subjectively evaluated therapy effects (*VEV-R*), and percentage of dropouts (all $ps ≥ .605$, see Table 3). Furthermore, no cognitive improvements over CBT in the clinical groups emerged (all $ps ≥ .272$, see Table 3). However, for digit span, medium-to-large time effects indicated better performance at follow-up over all groups (all $ps ≤ .009$). Robust regressions yielded similar results.

Regression analyses on the predictive value of lifetime trauma history (yes/no) for therapy outcome (*BDI-II* changes of depressive symptom severity and dropout status,

Table 3

Clinical Improvement and Changes in Working Memory in the Digit Span Task, Interference and Conflict Adaptation Effects of Means of Median Reaction Times (RTs) and Percentage of Errors (PE) in the Simon Task, and Autobiographical Memory Specificity in the Autobiographical Memory Test (AMT) Between Baseline and Follow-Up Assessment in Patients With Major Depression With (MD+T+) and Without (MD+T-) as well as Controls Without (MD-T-) Lifetime Trauma History

Participants' characteristics	MD+T+ (n = 21)		MD+T- (n = 16)		MD-T- (n = 34)		Test statistic	p	η²p 90% CI [LL, UL]	Adjusted p (Holm-Bonferroni correction)
	T1	T2	T1	T2	T1	T2				
BDI-II score (M, SD)	21.67 (7.72)	12.14 (9.52)	23.12 (8.83)	11.25 (8.19)	3.97 (4.76)	4.59 (5.3)	$F_{2, 68} = 23.54^{a,b}$	< .001[1]	.409 [.25, .52]	-
BDI-II change score (M, SD)	-9.52 (8.48)		-11.88 (7.69)		0.62 (5.32)					-
VEV-R symptom improvement (%)	16 (76.2)		13 (86.7)		n.a.		$X^2_2 = 1.01$	.605	n.a.	-
Dropouts (%)	9 (29)		7 (26.9)		n.a.		$X^2_2 = 0.31$	1.00	n.a.	-
Digit Span										
Total	18.57 (3.16)	19.81 (4.09)	18.13 (3.57)	18.94 (3.09)	18.59 (3.2)	20.38 (3.54)	$F_{2, 68} = 0.77^a$	.47	.022 [0, .09]	1
Forward	10.1 (1.73)	10.81 (2.21)	9.81 (1.8)	9.94 (1.39)	10 (1.67)	10.94 (1.98)	$F_{2, 68} = 1.17^a$	.318	.033 [0, .11]	1
Backward	8.48 (1.94)	9 (2.15)	8.31 (2.3)	9 (2.03)	8.59 (2.29)	9.44 (2.21)	$F_{2, 68} = 0.18^a$	.833	.005 [0, .04]	1
Simon task: RT										
Interference effect	26.48 (29.66)^b	20.05 (26.93)^b	20.38 (27.09)	15.78 (23.59)	20.71 (25.65)^c	24.48 (24.12)^c	$F_{2, 64} = 1.33^a$	.272	.04 [0, .12]	1
Conflict adaptation effect	66.6 (29.95)^b	60.1 (31.4)^b	54.88 (50.85)	65.84 (35.65)	41.02 (23.4)^c	50.32 (37.03)^c	$F_{2, 64} = 0.96^a$	.387	.029 [0, .1]	1
Simon task: PE										
Interference effect	2.24 (1.84)^b	1.55 (3.27)^b	1 (3.35)	1.45 (2.58)	1.31 (4.05)^c	2.41 (2.76)^c	$F_{2, 64} = 1.23^a$	.299	.037 [0, .12]	1
Conflict adaptation effect	4.76 (5.49)^b	4.22 (6.16)^b	4.4 (5.75)	4.24 (5.12)	5 (6.26)^c	3.32 (6.2)^c	$F_{2, 64} = 0.25^a$	.78	.008 [0, .05]	1
AMT number of specific memories										
Total	6.05 (2.71)	6.67 (2.31)	4.47 (2.13)^d	4.53 (2.23)^d	6.75 (2)^e	6.25 (2.24)^e	$F_{2, 65} = 1.21^a$	.305	.036 [0, .11]	1
Positive cues	3.1 (1.38)	3.29 (1.31)	2.47 (1.3)^d	2.4 (1.3)^d	3.63 (1.29)^e	3.19 (1.42)^e	$F_{2, 65} = 1.2^a$	.308	.036 [0, .11]	1
Negative cues	2.95 (1.47)	3.38 (1.32)	2 (1.13)^d	2.13 (1.13)^d	3.12 (1.24)^e	3.06 (1.16)^e	$F_{2, 65} = 0.64^a$	.529	.019 [0, .08]	1

Note. Data is presented as *M (SD)*. *CI* = confidence interval; *LL* = lower level; *UL* = upper level; BDI-II = Beck Depression Inventory-II; RT = median reaction times; PE = percentages of error.

[a]Group x time interaction. [b]refers to *n* = 20. [c]refers to *n* = 31. [d]refers to *n* = 15. [e]refers to *n* = 32.

[1]MD+T+ = MD+T- > MD-T- (*p* < .001).

respectively) yielded no associations (all *p*s ≥ .391). When, in a second step, adding respective facets of baseline cognitive functioning, more pronounced reductions of depressive symptom severity (*BDI-II*) emerged with smaller forward digit span, *b* = 1.48, 95% CI [0.07; 2.90], *p* = .041), while for all other measures of cognitive functioning, no predictive value emerged (all *p*s ≥ .059). Adding, in a third step, interaction terms of lifetime trauma history (yes/no) and baseline cognitive functioning did not predict CBT outcome regarding *BDI-II* and dropout status (all *p*s ≥ .058). Notably, robust regressions led to similar results.

Discussion

The aim of the study was to assess associations of (i) lifetime trauma history according to the *DSM-IV* (American Psychiatric Association, 2007) and (ii) facets of cognitive functioning (i.e., WM, interference susceptibility, conflict adaptation, and OGM) as well as (iii) their interaction with CBT outcome among patients with MD. At baseline, more pronounced conflict adaptation emerged in individuals with MD and lifetime trauma history in contrast to non-exposed healthy controls, while autobiographical memory was found to be primarily affected in MD without lifetime trauma history compared to both control groups. Notably, individuals with MD with and without lifetime trauma history did not differ regarding treatment outcome, and the cognitive parameters proved stable over CBT. Exploratory analyses suggested no direct or interacting association of lifetime trauma history, and only a tentative one of forward digit span, but no other aspects of cognitive functioning with treatment outcome.

Baseline Cognitive Functioning

On a cross-sectional level, the results support the role of lifetime trauma history for cognitive functioning in MD. While no differences emerged for interference susceptibility and WM, MD+T+ patients showed higher conflict adaptation of median RTs than MD-T- participants, with MD+T- and MD-T+ in between. This corresponds with previous findings from our group of more pronounced conflict adaptation in traumatized individuals with and possibly also without PTSD (Schindler et al., 2020; Steudte-Schmiedgen et al., 2014). However, as there also are suggested associations of conflict adaptation and depressive symptom severity (van Steenbergen et al., 2012), albeit without considering trauma history, further studies are desirable.

Interestingly, autobiographical memory yielded contrasting findings: MD+T- patients showed more pronounced OGM compared to the healthy control groups, corresponding with our previous findings of OGM in PTSD, but not trauma exposure *per se* (Schindler et al., 2020), and suggestions from reviews and meta-analyses (Moore & Zoellner, 2007; Ono et al., 2016; Sumner et al., 2010; Williams et al., 2007). Further, it supports the findings of

Kuyken, Howell, and Dalgleish (2006) of OGM only in MD without (childhood) trauma history, but not those of Aglan et al. (2010) of OGM in MD with history of CSA. In sum, neither the results for conflict adaptation, nor those for OGM speak for a mere additive effect of trauma and MD on cognitive functioning, but rather for complex patterns with different impacts on different processes, and, potentially, different implications for clinical practice.

Clinical and Cognitive Treatment Outcome Under Consideration of Lifetime Trauma History

In contrast to several previous studies particularly on childhood trauma (reviewed in Nemeroff, 2016; Teicher & Samson, 2013), our data suggest CBT to be equally effective in individuals with MD with and without the history of at least one traumatic event according to the *DSM-IV*. Several aspects may contribute to this divergence. Firstly, it is plausible that lifetime trauma, as examined in this study, does exert different effects than childhood trauma. Importantly, in our study, MD+T+ and MD+T- groups reported equal *CTQ* childhood maltreatment severity, and it is conceivable that this may have contributed to lacking group differences with respect to CBT effectiveness. Notably, also with respect to the *THQ*, the MD+T+ and the MD+T- groups did only differ on a descriptive level. However, it is important to consider that this instrument refers to the number and frequency of *potentially* traumatic events, for which the presence of the complete *DSM-IV* criteria are not checked. In order to better understand the role of childhood and adulthood trauma for CBT effectiveness, studies explicitly contrasting individuals with MD (i) without lifetime trauma, (ii) with exclusively childhood, and (iii) with exclusively adulthood trauma as defined by the current diagnostic criteria are necessary. Furthermore, treatment differences might have played a role. Most prominently, the majority of studies reporting similar therapy outcome for MD with and without (particularly childhood) trauma history had applied combined psychotherapy and antidepressant medication (Lewis et al., 2010; Miniati et al., 2010; Nemeroff et al., 2003; but Asarnow et al., 2009), as was the case for approximately half of our sample. Further, we cannot rule out whether, in our study, trauma status had led to slight individual treatment adaptations by the responsible therapists. This might, for instance, have led to combined modifications of trauma-related and -unrelated automatic thought patterns, or the encouraging of restarting activities avoided after the trauma during behavioral interventions within the context of the utilized CBT manuals (Hautzinger, 1998, 2008). Thus, future studies applying more strictly manualized CBT and investigating larger MD groups with and without medication intake are required.

Additionally, the results corroborate previous findings of cognitive alterations in MD being highly stable over CBT (reviewed in Köhler et al., 2015; Moore & Zoellner, 2007; Snyder & Hankin, 2019), and of this to be irrespective of trauma history. While WM improved from baseline to follow-up, this is presumably attributable to practice/habituation

effects, as it also concerned MD-T- individuals. As cognitive impairments are assumed to be associated with worse psychosocial functioning and increased relapse risk in MD (Rock et al., 2014), the continuous finding of this to not be adequately addressed by CBT shows the necessity to strive for "cognitive", next to clinical remission in MD (Bernhardt et al., 2019; Bortolato et al., 2016). For example, this might be achieved by directly targeting cognitive functioning during MD-centered CBT. While research on EF training in MD is still in its infancy (for a meta-analysis, see, e.g., Motter et al., 2016), there are promising results that OGM, as well as MD symptomatology itself may be influenced by interventions directly focusing on autobiographical recall, albeit with long-term stability still questionable (for a meta-analysis, see Barry, Sze, & Raes, 2019).

The exploratory predictive analyses on lifetime trauma history and cognitive functioning for CBT do not provide clear results from which robust next steps could be derived. What can be clearly stated as of now is that there, again, was no evidence for a relevant role of lifetime trauma history. Further, only a singular association with cognitive parameters emerged, suggesting smaller WM to be associated with more pronounced depressiveness-related CBT effects. In sum, this pattern, albeit stemming from a very small sample size, supports the findings of Goodkind et al. (2016) on interference susceptibility, but stands at variance with those of Sumner et al. (2010) suggesting a predictive role of autobiographical memory specificity in MD. Future studies are needed to follow up on autobiographical memory in this context, or investigate whether other cognitive markers might be more suitable to predict clinical outcome after standardized psychotherapeutic/pharmacological treatment (e.g., Groves et al., 2018) with or without taking trauma history into account.

Strengths, Limitations, and Outlook

One central strength of the study is the naturalistic, highly ecologically valid study design. While the inclusion of a waiting control group of MD+T+/MD+T- patients not receiving CBT was impossible for ethical reasons, the fact that a healthy control group was studied longitudinally alongside the MD individuals is a further major strength, as it allowed the separation of CBT-associated and mere practice effects on cognitive functioning. However, limitations resulting from the naturalistic design are the heterogeneous manifestations of psychopathology and medication and the group differences in educational status and smoking. Further limitations include the lack of an objective, observer-rated outcome of depressiveness (e.g., the Hamilton Rating Scale for Depression; Hamilton, 1960), as well as the small sample sizes and the thus reduced statistical power for detecting especially interactive relationships. However, the fact that the vast majority of associations were confirmed in robust analyses corroborates the validity of the findings. Finally, behavioral tasks established in cognitive psychology, such as the ones used in our study, are characterized by task impurity, which describes the impossibility of assessing "pure" cognitive processes without simultaneously eliciting others (Miyake

et al., 2000; e.g., Scott et al., 2015). In order to maximize transparency in data reporting, we chose to report subscale scores of the cognitive tasks for which different properties are discussed (Botvinick et al., 2004; Wechsler, 1997; Williams & Broadbent, 1986). In addition, we acknowledge that for any of the assessed tasks, additional cognitive processes such as processing speed, attention, and motivation – while not directly studied – are inevitably involved.

Conclusions

In conclusion, the study is the first to examine lifetime trauma history, cognitive functioning, and their interaction in the context of CBT in patients with MD. On a cross-sectional level, conflict adaptation and autobiographical memory specificity emerged to be differentially affected in MD with and without lifetime trauma history. Contrary to previous research on childhood trauma, we found no evidence for a differential treatment response in patients with MD with and without lifetime trauma history as defined by the *DSM-IV*. Further, the cognitive parameters were stable over CBT, and only a singular predictive association of forward digit span, but no other facets of baseline cognitive functioning, lifetime trauma history, or their interaction with treatment outcome emerged. These insights into the interaction between lifetime trauma history and cognitive functioning provide unique extensions for research on MD psychopathology and treatment and underline the relevance of "cognitive" remission (Bernhardt et al., 2019; Bortolato et al., 2016). For achieving this aim, further research is required to allow more profound, neuroscience-informed diagnostic processes and personalized, multi-modal treatment approaches depending on patients' individual manifestation of cognitive functioning (De Raedt, 2020).

Supplementary Materials

The Supplementary Materials contain the following items (for access see Index of Supplementary Materials below):

- Supplement 1 (Lifetime trauma history and mental disorder comorbidities)
- Supplement 2 (Baseline demographic, health-related, and clinical characteristics of patients with major depression with (MD+T+) and without (MD+T-) as well as controls without (MD-T-) lifetime trauma history available for longitudinal analyses)

Index of Supplementary Materials

Schindler, L., Stalder, T., Kirschbaum, C., Plessow, F., Schönfeld, S., Hoyer, J., Trautmann, S., Weidner, K., & Steudte-Schmiedgen, S. (2021). *Supplementary materials to "Lifetime trauma history and cognitive functioning in major depression and their role for cognitive-behavioral therapy outcome"* [Additional information]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.5073

PsychOpen GOLD

References

Aglan, A., Williams, J. M. G., Pickles, A., & Hill, J. (2010). Overgeneral autobiographical memory in women: Association with childhood abuse and history of depression in a community sample. *British Journal of Clinical Psychology, 49*(3), 359-372. https://doi.org/10.1348/014466509X467413

Ajilchi, B., Nejati, V., Town, J. M., Wilson, R., & Abbass, A. (2016). Effects of intensive short-term dynamic psychotherapy on depressive symptoms and executive functioning in major depression. *The Journal of Nervous and Mental Disease, 204*(7), 500-505. https://doi.org/10.1097/NMD.0000000000000518

American Psychiatric Association. (2007). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR* (4th ed., text revision, 10th print). American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). American Psychiatric Publishing.

Asarnow, J. R., Emslie, G., Clarke, G., Wagner, K. D., Spirito, A., Vitiello, B., Iyengar, S., Shamseddeen, W., Ritz, L., Birmaher, B., Ryan, N., Kennard, B., Mayes, T., De Bar, L., Mccracken, J., Strober, M., Suddath, R., Leonard, H., Porta, G., Keller, M., & Brent, D. (2009). Treatment of selective serotonin reuptake inhibitor – Resistant depression in adolescents: Predictors and moderators of treatment response. *Journal of the American Academy of Child and Adolescent Psychiatry, 48*(3), 330-339. https://doi.org/10.1097/CHI.0b013e3181977476

Barry, T. J., Sze, W. Y., & Raes, F. (2019). A meta-analysis and systematic review of Memory Specificity Training (MeST) in the treatment of emotional disorders. *Behaviour Research and Therapy, 116*, 36-51. https://doi.org/10.1016/j.brat.2019.02.001

Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., Jüni, P., & Cuijpers, P. (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: A network meta-analysis. *PLoS Medicine, 10*(5), Article e1001454. https://doi.org/10.1371/journal.pmed.1001454

Beblo, T., Baumann, B., Bogerts, B., Wallesch, C.-W., & Herrmann, M. (1999). Neuropsychological correlates of major depression: A short-term follow-up. *Cognitive Neuropsychiatry, 4*(4), 333-341. https://doi.org/10.1080/135468099395864

Bernhardt, M., Klauke, S., & Schröder, A. (2019). Longitudinal course of cognitive function across treatment in patients with MDD: A meta-analysis. *Journal of Affective Disorders, 249*, 52-62. https://doi.org/10.1016/j.jad.2019.02.021

Bortolato, B., Miskowiak, K. W., Köhler, C. A., Maes, M., Fernandes, B. S., Berk, M., & Carvalho, A. F. (2016). Cognitive remission: A novel objective for the treatment of major depression? *BMC Medicine, 14*, Article 9. https://doi.org/10.1186/s12916-016-0560-3

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624-652. https://doi.org/10.1037/0033-295X.108.3.624

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences, 8*(12), 539-546. https://doi.org/10.1016/j.tics.2004.10.003

Cuijpers, P., Karyotaki, E., Weitz, E., Andersson, G., Hollon, S. D., & van Straten, A. (2014). The effects of psychotherapies for major depression in adults on remission, recovery and improvement: A meta-analysis. *Journal of Affective Disorders, 159*, 118-126. https://doi.org/10.1016/j.jad.2014.02.026

Dannehl, K., Rief, W., & Euteneuer, F. (2017). Childhood adversity and cognitive functioning in patients with major depression. *Child Abuse & Neglect, 70*, 247-254. https://doi.org/10.1016/j.chiabu.2017.06.013

De Raedt, R. (2020). Contributions from neuroscience to the practice of Cognitive Behaviour Therapy: Translational psychological science in service of good practice. *Behaviour Research and Therapy, 125*, Article 103545. https://doi.org/10.1016/j.brat.2019.103545

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Ehlers, A., Steil, R., Winter, H., & Foa, E. (1996). *Deutsche Übersetzung der Posttraumatic Stress Diagnostic Scale (PDS)*. Warneford Hospital, Department of Psychiatry.

Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy, 98*, 19-38. https://doi.org/10.1016/j.brat.2017.05.013

Fischer, R., Dreisbach, G., & Goschke, T. (2008). Context-sensitive adjustments of cognitive control: Conflict-adaptation effects are modulated by processing demands of the ongoing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(3), 712-718. https://doi.org/10.1037/0278-7393.34.3.712

Gast, U., Rodewald, F., Benecke, H.-H., & Driessen, M. (2001). *CTQ – Childhood Trauma Questionnaire – deutsche Fassung*. Unpublished Manuscript, Medizinische Hochschule Hannover, Hannover, Germany.

Goodkind, M. S., Gallagher-Thompson, D., Thompson, L. W., Kesler, S. R., Anker, L., Flournoy, J., Berman, M. P., Holland, J. M., & O'Hara, R. M. (2016). The impact of executive function on response to cognitive behavioral therapy in late-life depression. *International Journal of Geriatric Psychiatry, 31*(4), 334-339. https://doi.org/10.1002/gps.4325

Groves, S. J., Douglas, K. M., & Porter, R. J. (2018). A systematic review of cognitive predictors of treatment outcome in major depression. *Frontiers in Psychiatry, 9*, Article 382. https://doi.org/10.3389/fpsyt.2018.00382

Hager, W., & Hasselhorn, M. (1994). *Handbuch deutschsprachiger Wortnormen*. Hogrefe.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*, 56-62. https://doi.org/10.1136/jnnp.23.1.56

Hans, E., & Hiller, W. (2013). Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: A meta-analysis of nonrandomized effectiveness studies. *Journal of Consulting and Clinical Psychology, 81*(1), 75-88. https://doi.org/10.1037/a0031080

Hautzinger, M. (1998). *Depression* (Fortschritte der Psychotherapie, Bd. 4). Hogrefe.

Hautzinger, M. (2008). *Kognitive Verhaltenstherapie bei Depressionen: Behandlungsanleitungen und Materialien* (Materialien für die klinische Praxis, Bd. 6., neu bearb. Aufl. [Nachdr.]). Beltz PVU.

PsychOpen GOLD

Hautzinger, M., Keller, F., & Kühner, C. (2006). *Beck Depressions-Inventar (BDI-II). Revision.* Harcourt Test Services.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70.

Hooper, L. M., Stockton, P., Krupnick, J. L., & Green, B. L. (2011). Development, use, and psychometric properties of the Trauma History Questionnaire. *Journal of Loss and Trauma, 16*(3), 258-283. https://doi.org/10.1080/15325024.2011.572035

Köhler, C. A., Carvalho, A. F., Alves, G. S., McIntyre, R. S., Hyphantis, T. N., & Cammarota, M. (2015). Autobiographical memory disturbances in depression: A novel therapeutic target? *Neural Plasticity, 2015*, Article 759139. https://doi.org/10.1155/2015/759139

Kolassa, I.-T., & Elbert, T. (2007). Structural and functional neuroplasticity in relation to traumatic stress. *Current Directions in Psychological Science, 16*(6), 321-325. https://doi.org/10.1111/j.1467-8721.2007.00529.x

Kuyken, W., Howell, R., & Dalgleish, T. (2006). Overgeneral autobiographical memory in depressed adolescents with, versus without, a reported history of trauma. *Journal of Abnormal Psychology, 115*(3), 387-396. https://doi.org/10.1037/0021-843X.115.3.387

Lahr, D., Beblo, T., & Hartje, W. (2007). Cognitive performance and subjective complaints before and after remission of major depression. *Cognitive Neuropsychiatry, 12*(1), 25-45. https://doi.org/10.1080/13546800600714791

Lewis, C. C., Simons, A. D., Nguyen, L. J., Murakami, J. L., Reid, M. W., Silva, S. G., & March, J. S. (2010). Impact of childhood trauma on treatment outcome in the treatment for adolescents with depression study (TADS). *Journal of the American Academy of Child and Adolescent Psychiatry, 49*(2), 132-140. https://doi.org/10.1016/j.jaac.2009.10.007

Maercker, A. (2002). *THQ – Trauma History Questionnaire – deutsche Fassung.* Unpublished Manuscript, Universität Zürich, Zürich, Switzerland.

McIntyre, R. S., Cha, D. S., Soczynska, J. K., Woldeyohannes, H. O., Gallaugher, L. A., Kudlow, P., Alsuwaidan, M., & Baskaran, A. (2013). Cognitive deficits and functional outcomes in major depressive disorder: Determinants, substrates, and treatment interventions. *Depression and Anxiety, 30*(6), 515-527. https://doi.org/10.1002/da.22063

Miniati, M., Rucci, P., Benvenuti, A., Frank, E., Buttenfield, J., Giorgi, G., & Cassano, G. B. (2010). Clinical characteristics and treatment outcome of depression in patients with and without a history of emotional and physical abuse. *Journal of Psychiatric Research, 44*(5), 302-309. https://doi.org/10.1016/j.jpsychires.2009.09.008

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language, 21*(2), 169-183. https://doi.org/10.1055/s-2000-7563

Moore, S. A., & Zoellner, L. A. (2007). Overgeneral autobiographical memory and traumatic events: An evaluative review. *Psychological Bulletin, 133*(3), 419-437. https://doi.org/10.1037/0033-2909.133.3.419

PsychOpen GOLD

Motter, J. N., Pimontel, M. A., Rindskopf, D., Devanand, D. P., Doraiswamy, P. M., & Sneed, J. R. (2016). Computerized cognitive training and functional recovery in major depressive disorder: A meta-analysis. *Journal of Affective Disorders, 189*, 184-191. https://doi.org/10.1016/j.jad.2015.09.022

Nanni, V., Uher, R., & Danese, A. (2012). Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: A meta-analysis. *The American Journal of Psychiatry, 169*(2), 141-151. https://doi.org/10.1176/appi.ajp.2011.11020335

Nemeroff, C. B. (2016). Paradise lost: The neurobiological and clinical consequences of child abuse and neglect. *Neuron, 89*(5), 892-909. https://doi.org/10.1016/j.neuron.2016.01.019

Nemeroff, C. B., Heim, C. M., Thase, M. E., Klein, D. N., Rush, A. J., Schatzberg, A. F., Ninan, P. T., McCullough, J. P., Jr., Weiss, P. M., Dunner, D. L., Rothbaum, B. O., Kornstein, S., Keitner, G., & Keller, M. B. (2003). Differential responses to psychotherapy versus pharmacotherapy in patients with chronic forms of major depression and childhood trauma. *Proceedings of the National Academy of Sciences of the United States of America, 100*(24), 14293-14296. https://doi.org/10.1073/pnas.2336126100

Oberg, A. L., & Mahoney, D. W. (2007). Linear mixed effects models. *Methods in Molecular Biology, 404*, 213-234. https://doi.org/10.1007/978-1-59745-530-5_11

Ono, M., Devilly, G. J., & Shum, D. H. K. (2016). A meta-analytic review of overgeneral memory: The role of trauma history, mood, and the presence of posttraumatic stress disorder. *Psychological Trauma: Theory, Research, Practice, and Policy, 8*(2), 157-164. https://doi.org/10.1037/tra0000027

Peeters, F., Wessel, I., Merckelbach, H., & Boon-Vermeeren, M. (2002). Autobiographical memory specificity and the course of major depressive disorder. *Comprehensive Psychiatry, 43*(5), 344-350. https://doi.org/10.1053/comp.2002.34635

R [Computer software]. (2017). R Foundation for Statistical Computing. Vienna, Austria.

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: A systematic review and meta-analysis. *Psychological Medicine, 44*(10), 2029-2040. https://doi.org/10.1017/S0033291713002535

Royall, R. M. (1986). Model robust confidence intervals using maximum likelihood estimators. *International Statistical Review / Revue Internationale De Statistique, 54*(2), 221-226. https://doi.org/10.2307/1403146

Saleh, A., Potter, G. G., McQuoid, D. R., Boyd, B., Turner, R., MacFall, J. R., & Taylor, W. D. (2017). Effects of early life stress on depression, cognitive performance and brain morphology. *Psychological Medicine, 47*(1), 171-181. https://doi.org/10.1017/S0033291716002403

Schindler, L., Stalder, T., Kirschbaum, C., Plessow, F., Schönfeld, S., Hoyer, J., Trautmann, S., & Steudte-Schmiedgen, S. (2020). Cognitive functioning in posttraumatic stress disorder before and after cognitive-behavioral therapy. *Journal of Anxiety Disorders, 74*, Article 102265. https://doi.org/10.1016/j.janxdis.2020.102265

PsychOpen GOLD

Schmid, M., & Hammar, A. (2013). A follow-up study of first episode major depressive disorder: Impairment in inhibition and semantic fluency-potential predictors for relapse? *Frontiers in Psychology, 4*, Article 633. https://doi.org/10.3389/fpsyg.2013.00633

Schönfeld, S., & Ehlers, A. (2006). Overgeneral memory extends to pictorial retrieval cues and correlates with cognitive features in posttraumatic stress disorder. *Emotion, 6*(4), 611-621. https://doi.org/10.1037/1528-3542.6.4.611

Scott, J. C., Matt, G. E., Wrocklage, K. M., Crnich, C., Jordan, J., Southwick, S. M., Krystal, J. H., & Schweinsburg, B. C. (2015). A quantitative meta-analysis of neurocognitive functioning in posttraumatic stress disorder. *Psychological Bulletin, 141*(1), 105-140. https://doi.org/10.1037/a0038039

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry, 59*(Suppl 20), 22-33.

Sherin, J. E., & Nemeroff, C. B. (2011). Post-traumatic stress disorder: The neurobiological impact of psychological trauma. *Dialogues in Clinical Neuroscience, 13*(3), 263-278. https://doi.org/10.31887/DCNS.2011.13.2/jsherin

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 31–86). Elsevier.

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychological Bulletin, 139*(1), 81-132. https://doi.org/10.1037/a0028727

Snyder, H. R., & Hankin, B. L. (2019). Cognitive function in mood and anxiety disorders. In R. A. Stern & M. L. Alosco (Eds.), *The Oxford handbook of adult cognitive disorders* (pp. 249–273). Oxford University Press.

Stata Statistical Software: Release 15 [Computer software]. (2017). College Station, TX, USA.

Steudte, S., Kirschbaum, C., Gao, W., Alexander, N., Schonfeld, S., Hoyer, J., & Stalder, T. (2013). Hair cortisol as a biomarker of traumatization in healthy individuals and posttraumatic stress disorder patients. *Biological Psychiatry, 74*(9), 639-646. https://doi.org/10.1016/j.biopsych.2013.03.011

Steudte-Schmiedgen, S., Stalder, T., Kirschbaum, C., Weber, F., Hoyer, J., & Plessow, F. (2014). Trauma exposure is associated with increased context-dependent adjustments of cognitive control in patients with posttraumatic stress disorder and healthy controls. *Cognitive, Affective & Behavioral Neuroscience, 14*(4), 1310-1319. https://doi.org/10.3758/s13415-014-0299-2

Sumner, J. A., Griffith, J. W., & Mineka, S. (2010). Overgeneral autobiographical memory as a predictor of the course of depression: A meta-analysis. *Behaviour Research and Therapy, 48*(7), 614-625. https://doi.org/10.1016/j.brat.2010.03.013

PsychOpen GOLD

Teicher, M. H., & Samson, J. A. (2013). Childhood maltreatment and psychopathology: A case for ecophenotypic variants as clinically and neurobiologically distinct subtypes. *The American Journal of Psychiatry, 170*(10), 1114-1133. https://doi.org/10.1176/appi.ajp.2013.12070957

van Steenbergen, H., Band, G. P. H., & Hommel, B. (2010). In the mood for adaptation: How affect regulates conflict-driven control. *Psychological Science, 21*(11), 1629-1634. https://doi.org/10.1177/0956797610385951

van Steenbergen, H., Booij, L., Band, G. P. H., Hommel, B., & van der Does, A. J. W. (2012). Affective regulation of cognitive-control adjustments in remitted depressive patients after acute tryptophan depletion. *Cognitive, Affective & Behavioral Neuroscience, 12*(2), 280-286. https://doi.org/10.3758/s13415-011-0078-2

Vasterling, J. J., & Arditte Hall, K. A. (2018). Neurocognitive and information processing biases in posttraumatic stress disorder. *Current Psychiatry Reports, 20*(11), Article 99. https://doi.org/10.1007/s11920-018-0964-1

Wechsler, D. (1997). *Wechsler Memory Scale – Third Edition (WMS-III).* The Psychological Corporation.

Williams, J. M. G., Barnhofer, T., Crane, C., Herman, D., Raes, F., Watkins, E., & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin, 133*(1), 122-148. https://doi.org/10.1037/0033-2909.133.1.122

Williams, J. M. G., & Broadbent, K. (1986). Autobiographical memory in suicide attempters. *Journal of Abnormal Psychology, 95*(2), 144-149. https://doi.org/10.1037/0021-843X.95.2.144

Wittchen, H.-U., & Pfister, H. (1997). *DIA-X-Interviews: Manual für Screening-Verfahren und Interview; Interviewheft Längsschnittuntersuchung (DIA-X-Lifetime); Ergänzungsheft (DIA-X-Lifetime); Interviewheft Querschnittuntersuchung (DIA-X-12Monate); Ergänzungsheft (DIA-X-12Monate); PC-programm zur Durchführung des Interviews (Längs- und Querschnittuntersuchung); Auswertungsprogramm.* Swets & Zeitlinger.

Zielke, M., & Kopf-Mehnert, C. (2001). Der VEV-R-2001: Entwicklung und testtheoretische Reanalyse der revidierten Form des Veränderungsfragebogens des Erlebens und Verhaltens (VEV). *Praxis Klinische Verhaltensmedizin und Rehabilitation, 53*, 7-19.

Repetitive Negative Thinking About Suicide: Associations With Lifetime Suicide Attempts

Tobias Teismann[1], Thomas Forkmann[2], Johannes Michalak[3], Julia Brailovskaia[1]

[1] *Mental Health Research and Treatment Center, Ruhr-Universität Bochum, Bochum, Germany.* **[2]** *Department of Clinical Psychology, University of Duisburg-Essen, Essen, Germany.* **[3]** *Department of Clinical Psychology and Psychotherapy, Universität Witten-Herdecke, Witten, Germany.*

Corresponding Author: Tobias Teismann, Mental Health Research and Treatment Center, Ruhr-Universität Bochum, Germany, Massenbergstraße 9-13, 44787 Bochum. Phone: 0049-234-3227787. E-mail: tobias.teismann@rub.de

Supplementary Materials: Materials [see Index of Supplementary Materials]

Abstract

Background: Repetitive negative thinking has been identified as an important predictor of suicide ideation and suicidal behavior. Yet, only few studies have investigated the effect of suicide-specific rumination, i.e., repetitive thinking about death and/or suicide on suicide attempt history. On this background, the present study investigated, whether suicide-specific rumination differentiates between suicide attempters and suicide ideators, is predictive of suicide attempt history and mediates the association between suicide ideation and suicide attempts.

Method: A total of 257 participants with a history of suicide ideation (55.6% female; Age M = 30.56, Age SD = 11.23, range: 18–73 years) completed online measures on suicidality, general and suicide-specific rumination.

Results: Suicide-specific rumination differentiated suicide attempters from suicide ideators, predicted suicide attempt status (above age, gender, suicide ideation, general rumination) and fully mediated the association between suicide ideation and lifetime suicide attempts.

Conclusion: Overall, though limited by the use of a non-clinical sample and a cross-sectional study design, the present results suggest that suicide-specific rumination might be a factor of central relevance in understanding transitions to suicidal behavior.

Keywords

repetitive negative thinking, rumination, suicide ideation, suicide attempts

Highlights

- Suicide-specific rumination was investigated in participants with a history of suicide ideation.
- Suicide-specific rumination differentiated suicide attempters from suicide ideators.
- Suicide-specific rumination predicted suicide attempt status.
- Suicide-specific rumination mediated the association between suicide ideation and lifetime suicide attempts.
- Suicide-specific rumination might be a factor of central relevance in understanding transitions to suicidal behavior.

Repetitive negative thinking (RNT) is defined as a style of thinking about one's problems or negative experiences with three key characteristics: the thinking is repetitive, it is at least partly intrusive, and it is difficult to disengage from. Two additional features of RNT are that individuals perceive it as unproductive and it captures mental capacity (Ehring et al., 2011). The two most intensively studied types of RNT are worry and depressive rumination. RNT – in the form of rumination and worry – has been identified as a critical factor in the development and maintenance of psychiatric symptoms and disorders (Ehring & Watkins, 2008; Teismann & Ehring, 2019; Watkins, 2008). In prospective studies, rumination was found to predict the future onset of a major depressive episode (Nolen-Hoeksema, 2000; Nolen-Hoeksema et al., 2007; Robinson & Alloy, 2003; Wilkinson et al., 2013) and to mediate the effect of various risk factors on the onset of depression (Spasojević & Alloy, 2001). Additional studies have shown that rumination prospectively predicts the onset of post-traumatic stress disorder (Moulds et al., 2020; Szabo et al., 2017) and is linked to the maintenance of social anxiety disorder (Penney & Abbott, 2014), insomnia (Takano et al., 2014) and eating disorder psychopathology (Smith, Mason, & Lavender, 2018). Moreover, a close association between RNT, suicide ideation and suicide attempts has been shown in cross-sectional and longitudinal studies (Rogers & Joiner, 2017) – even when different types of RNT as well as different methodologies, samples (clinical and non-clinical) and measures of suicidality were used (Kerkhof & van Spijker, 2011; Law & Tucker, 2018). As such, rumination significantly predicted suicide ideation in prospective studies using student and community samples (Miranda & Nolen-Hoeksema, 2007; Smith, Alloy, & Abramson, 2006). Furthermore, rumination was found to be more common in suicide attempters than in non-attempters (e.g., Horwitz et al., 2019). Galynker (2017) understands intensive, persistent and uncontrollable brooding (*ruminative flooding*) as a core feature of an acute suicidal state, the so-called suicide crisis syndrome. Taken together, there is strong empirical evidence for the importance of RNT with respect to understanding suicide ideation and behavior.

In the vast majority of these studies, the relationship between general RNT and suicidal ideation and suicide attempts was investigated. However, Rogers and Joiner (2018a, 2018b) have recently started to study the effect of suicide-specific rumination,

that is, RNT about death and/or suicide. They found that suicide-specific rumination is associated with lifetime suicide attempts over and above a large array of known risk factors, including suicide ideation, general rumination, depression and anxiety (Rogers & Joiner, 2018a). Furthermore, they could show that the association between suicide-specific rumination and lifetime suicide attempts is mediated by an acute suicidal state, called acute suicidal affective disturbance (ASAD; Rogers & Joiner, 2018b). In both of these studies, suicide-specific rumination was assessed using either a 5-item (Rogers & Joiner, 2018b) or an 8-item (Rogers & Joiner, 2018a) version of the Suicide Rumination Scale (SRS). This scale assesses the tendency to ruminate or fixate on one's suicidal thoughts, intention and plans. However, it cannot be excluded that some items of the SRS may confound general preparation behavior ("When I have thoughts of suicide, I think about how I want to kill myself"; "… I wonder what the fastest and easiest way to die is") or so called flash forwards ("When I have thoughts of suicide, I imagine the process of how I want to kill myself"), with generic features of RNT ("When I have thoughts of suicide, I have trouble getting the suicidal thoughts out of my mind"). It is therefore unclear whether the significant association between suicide-specific rumination – as assessed with the SRS – and lifetime suicide attempts are in fact due to RNT or rather a consequence of increased preparation and planning behavior.

On this background, the current study aims at investigating the association between suicide-specific rumination and suicidal behavior with a suicide-specific version of the Perseverative Thinking Questionnaire (PTQ; Ehring et al., 2011), a self-report measure designed to assess core characteristics of RNT (repetitiveness, intrusiveness, difficulties with disengagement, perceived unproductiveness). The study had three aims: 1. To investigate whether suicide-specific rumination – as assessed with an unconfounded measure – differentiates between lifetime suicide attempters and non-attempters; 2. To investigate, whether suicide specific rumination is associated with lifetime suicide attempts – above and beyond age, gender, current suicide ideation and general rumination; 3. To investigate whether suicide-specific rumination mediates the association between current suicide ideation and lifetime suicide attempts. Since most suicide ideators do not show suicidal behavior, the necessity to understand what differentiates attempters from ideators has recently been highlighted (May & Klonsky, 2016).

Method and Materials

Participants and Procedure

Between March and May 2019, $N = 300$ (58% female; $M_{age} = 32.25$, $SD_{age} = 13.68$, range: 18−77 years) and again between February and June 2020, $N = 276$ (67% female; $M_{age} = 32.08$, $SD_{age} = 10.73$, range: 18−64 years) participants took part in a single assessment using an online survey. The assessments took part within the context of two other

studies (Teismann & Brailovskaia, 2020; Teismann et al., 2020), that were advertised as investigating the association between well-being and psychological strain. It was assured that no participant took part in both of these studies. Of the two samples, *n* = 257 (55.6% female; M_{age} = 30.56, SD_{age} = 11.23, range: 18–73 years) reported lifetime suicide ideation and were included in the present study. One-hundred and twenty-nine participants (50.2%) reported some suicide ideation in the last four weeks (SSEV- score ≥1); fifty-two participants (20.2%) indicated that they had attempted suicide at least once in their lifetime (range: 1–6). All participants – except for one Asian participant – were Caucasian.

Participants were recruited through postings at local university as well as social media postings on Facebook and Twitter. Data was collected through an anonymous online survey using the SoSci-server (https://www.soscisurvey.de/). Participation in the study was not compensated; yet, participating students were eligible to receive course credits. In order to take part in the study, participants had to be at least 18 years old and to give their consent to participation at the beginning of the study. Prior to assessments, all participants were informed about the purpose of the study, the voluntary nature of their participation, data storage and security. The study was approved by the responsible Ethics Committee.

Measures

Suicide Ideation and Behavior Scale (SSEV)

The SSEV (Teismann et al., 2021) assesses with six items the frequency of suicide ideation in the past four weeks (e.g., "During the past four weeks, … I thought it would be better if I wasn't alive, … I've been thinking about killing myself, … I have seriously considered killing myself"). All items are answered on a 6-point Likert scale ranging from "1=*never*" to "5=*many times every day*", with higher scores indicating greater severity of suicidal ideation. Occurrence ("In the course of my life I have tried to kill myself - and I really wanted to die") and number of lifetime suicide attempts ("How many times have you tried to kill yourself?") are assessed with two further SSEV-items. The scale has been shown to have a good internal consistency (Cronbach's α ≥ .92; Teismann et al., 2021). Accordingly, internal consistency was good in the current sample, (α = .84).

Perseverative Thinking Questionnaire (PTQ)

The PTQ (Ehring et al., 2011) is a 15-item self-report measure designed to assess process characteristics of perseverative thinking ("The same thoughts keep going through my mind again and again"; "I keep asking myself questions without finding an answer"; "Thoughts intrude into my mind"; "My thoughts take up all my attention"). All items are to be answered on a 5-point scale ranging from 0 ("*never*") to 4 ("*almost always*"). The scale has been shown to have good internal consistencies (Cronbach`s α ≥ .93; Ehring

et al., 2011). Accordingly, internal consistencies were excellent in the current sample, $\alpha = .95$.

Perseverative Thinking about Suicide Questionnaire (PTSQ)

The PTSQ (Teismann, 2018) is modeled after the PTQ and assesses with nine items suicide specific rumination ("I can´t stop dwelling about suicide"; "I am thinking about suicide the whole time"; "Thoughts about suicide intrude into my mind"; "My thoughts about suicide repeat themselves"). In the adaption process the word "thoughts" from the original PTQ was replaced by the term "suicidal thoughts" in the PTSQ: For example the PTQ-item "The same thoughts keep going through my mind again and again" became the PTSQ-item "The same thoughts about suicide keep going through my mind again and again". Items from the PTQ that were not adjustable in the described manner (i.e., "I think about many problems without solving any of them") were not included in the PTSQ. The adaptation was conducted by the first author and consented with all co-authors. All items are to be answered on a 5-point scale ranging from 0 (*"never"*) to 4 (*"almost always"*). Participants are only asked to answer all these items, if they affirm a first screening item ("In my lifetime I have thought about suicide"). The scale has been shown to have high internal consistency (Cronbach`s $\alpha = .94$; Höller et al., in preparation). Accordingly, an exploratory factor analysis (EFA) using principal component analysis (PCA; rotation method: varimax) revealed a unidimensional factor structure within the present sample as well as excellent internal consistency, $\alpha = .95$.

Statistical Analyses

Statistical analyses were conducted with SPSS 26 and the Process macro version 3.5 (www.processmacro.org/index.html; Rockwood & Hayes, 2020). Descriptive statistics and zero-order bivariate correlations between the investigated variables were calculated. Differences between groups (lifetime suicide ideators: $n = 205$ vs. lifetime attempters: $n = 52$) were analyzed using one-way ANOVAs. Considering the different sizes of both groups, Hedges'g was included as effect size (see Hedges, 1981). Notably, the current data fit the assumptions for the calculation of multivariate analyses (no significant outliners > 3 and < -3, number of significant outliners > 2 and < -2 below 5%; no violation of multicollinearity assumption as all values of tolerance > 0.25, and all variance inflation factor values < 5; interaction between the independent variables and their logarithmic transformations is not significant) (see Field, 2013; Tabachnick & Fidell, 2014; Urban & Mayerl, 2006). Next, a three-step multiple logistic regression analysis was calculated to examine the relative contribution of current suicide ideation (SSEV), general rumination (PTQ) and suicide-specific rumination (PTSQ) to the prediction of lifetime suicide attempt status (coded: 0 = no attempts, 1 = attempts). The variable age was significantly correlated with current suicide ideation ($r = -.163$, $p < .01$), general rumination ($r = -.189$, $p < .05$), and suicide-specific rumination ($r = -.161$, $p < .05$). The variable gender (coded: 0

= woman, 1 = man) was negatively correlated with general rumination (r = -.147, p < .05), and lifetime suicide attempt status (r = -.129, p < .05). Considering the relationships of age and gender with the potential predictors and the outcome of the regression model, both were included as control variables. Thus, age and gender were included in Step 1 of the regression model, current suicide ideation and general rumination were included in Step 2, and suicide-specific rumination was included in Step 3. Finally, a mediation analysis was conducted that included current suicide ideation (predictor), suicide-specific rumination (mediator), and number of lifetime suicide attempts (outcome). The basic association between current suicide ideation and lifetime suicide attempts was denoted by c (the total effect). The path of current suicide ideation to suicide specific rumination was denoted by a, and the path of suicide specific rumination to lifetime suicide attempts was denoted by b. The combined effect of path a and path b presented the indirect effect. The direct effect of current suicide ideation on lifetime suicide attempts after inclusion of suicide specific rumination in the model was denoted by c'. The mediation effect was assessed by the bootstrapping procedure (10.000 samples) that provides percentile bootstrap confidence intervals (95% CI).

Results

Descriptive Statistics, Correlations and Group Differences

Descriptive statistics for each measure and correlations are presented in Table 1. Correlation analyses indicated that all study variables correlated significantly with each other (see Table 1). The correlations ranged between r = .354 and r = .806 (all: p < .01), indicating medium to large effects (see Cohen, 1988).

Table 1

Means, Standard Deviations and Correlations of Study Variables

Measure	*M (SD)*	*Min–Max*	Skewness	Kurtosis	2	3	4
1. SSEV	8.00 (3.35)	6–23	2.130	4.449	.354**	.806**	.406**
2. SSEV-SA	0.32 (0.84)	0–6	3.932	18.382	–	.463**	.265**
3. PTSQ	15.07 (7.32)	9–44	1.509	1.813		–	.490**
4. PTQ	47.04 (13.07)	16–75	-.043	-.448			–

Note. N = 257; *M* = Mean; *SD* = Standard Deviation; *Min* = Minimum; *Max* = Maximum; SSEV = Suicide Ideation and Behavior Scale; SSEV-SA = Suicide Ideation and Behavior Scale – lifetime number of suicide attempts; PTSQ = Perseverative Thinking about Suicide Questionnaire; PTQ = Perseverative Thinking Questionnaire. SSEV-SA was dichotomized (0 = no attempts, 1 = attempts) for the correlation analyses.
**p < .01.

Lifetime suicide ideators (assessed with the PTSQ-screening item) and lifetime suicide attempters differed significantly in PTSQ-scores (suicide ideators: n = 205; M = 13.50, SD = 5.88, range: 9–36; suicide attempters: n = 52; M = 21.29, SD = 9.02, range: 9–44), $F(1,255)$ = 57.52, p < .001, effect size: Hedges'g = 1.17 (large effect). Furthermore, lifetime suicide ideators (M = 45.36, SD = 12.61, range: 16–75) and lifetime suicide attempters (M = 53.67, SD = 12.89, range: 23–75) differed significantly in PTQ-scores, $F(1,255)$ = 17.89, p < .001, effect size: Hedges'g = 0.65 (medium effect); with suicide attempters reporting more RNT than suicide ideators.

Prediction of Lifetime Suicide Attempts

Associations between study variables and lifetime suicide attempts are shown in Table 2. In the multiple logistic regression model, current suicide ideation (*OR*: 1.19; small effect, see Chen, Cohen, & Chen, 2010) and general rumination (*OR*: 1.03; small effect, see Chen et al., 2010) served as a significant predictor of lifetime suicide attempts in Step 2. However, in Step 3, only the new included variable suicide-specific rumination emerged as a significant predictor of lifetime suicide attempts (*OR*: 1.14; small effect, see Chen et al., 2010).

Table 2

Results From a Three-Step Multiple Logistic Regression Analysis Predicting Lifetime Suicide Attempts (Dichotomized: 0 = no attempts, 1 = attempts)

| Step	OR (95% CI)	p
Step 1		
Age	0.98 [0.95-1.01]	.163
Gender	0.47 [0.25-0.90]	.023
Step 2		
Age	0.99 [0.96-1.03]	.720
Gender	0.56 [0.28-1.14]	.110
SSEV	1.19 [1.09-1.31]	< .001
PTSQ	1.03 [1.00-1.06]	.047
Step 3		
Age	1.00 [0.96-1.03]	.796
Gender	0.50 [0.24-1.04]	.065
SSEV	0.97 [0.85-1.12]	.719
PTQ	1.01 [0.98-1.04]	.461
PTSQ	1.14 [1.06-1.23]	< .001

Note. N = 257; SSEV-SI = Suicide Ideation and Behavior Scale; PTQ = Perseverative Thinking Questionnaire; PTSQ = Perseverative Thinking about Suicide Question-naire; *OR* = odds ratio from logistic regression; *CI* = confidence interval.

Mediation Analysis

Figure 1 shows results of the bootstrapped mediation analysis. The basic relationship between current suicide ideation (predictor) and lifetime suicide attempts (outcome) was significant (total effect, *c: p* < .001). The association between current suicide ideation and suicide-specific rumination (mediator) (*a: p* < .001), as well as the link between suicide-specific rumination and lifetime suicide attempts (*b: p* < .001) were also significant. In contrast, the relationship between current suicide ideation and lifetime suicide attempts was no longer significant after the inclusion of suicide-specific rumination in the model (direct effect, *c': p* = .445). The indirect effect (*ab*) was significant, *b* = .10, *SE* = .03, 95% CI [.04, .17]. Thus, suicide-specific rumination significantly mediated the relationship between current suicide ideation and lifetime suicide attempts.

Figure 1

Mediation Model With Suicide Ideation (Predictor), Suicide-Specific Rumination (Mediator), and Lifetime Suicide Attempts (Outcome).
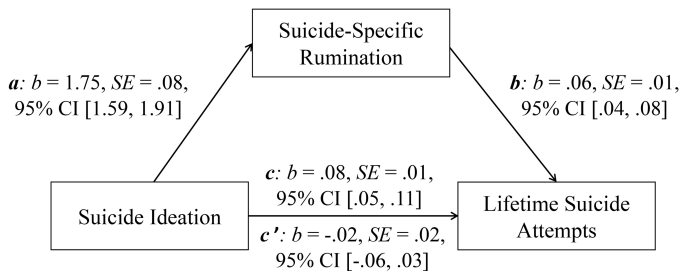
Note. c = total effect; *c'* = direct effect; *b* = standardized regression coefficient; *SE* = standard error; CI = confidence interval.

Discussion

The present study investigated the association between RNT – that is, suicide-specific rumination and general rumination – and (lifetime) suicide attempts. The main findings were as follows: (1.) General rumination and suicide-specific rumination differentiated between lifetime suicide attempters and suicide ideators; (2.) Suicide-specific rumination was predictive of lifetime suicide attempt status – controlling for age, gender, current suicide ideation and general rumination; (3.) The association between current suicide ideation and lifetime suicide attempts was fully mediated by suicide-specific rumination.

These results complement previous research showing an association between general rumination and suicide ideation/behavior (Rogers & Joiner, 2017) as well as between suicide-specific rumination and lifetime suicide attempts (Rogers & Joiner, 2018a, 2018b). In accordance with findings by Rogers and Joiner (2018a) it was shown that suicide-specific

PsychOpen GOLD

rumination outperformed other suicide risk factors – including current suicide ideation – in the prediction of lifetime suicide attempt status. Of note, findings could be replicated with a new – potentially unconfounded – measure of suicide-specific rumination. Though further study results have to be awaited, these findings suggest a rather robust effect of suicide-specific rumination. Accordingly, it seems as if RNT about suicide may be more pernicious in increasing the risk for suicidal behavior than ruminative thoughts about one's distress more generally.

Nonetheless, both general rumination and suicide-specific rumination differentiated between (lifetime) suicide attempters and (lifetime) suicide ideators (cf., Horwitz et al., 2019). Klonsky and May (2015) recently emphasized that it is crucial to understand factors that differentiate those who consider suicide from those who make suicide attempts. Yet, in a comprehensive meta-analysis May and Klonsky (2016) found only few studies that directly compared suicide ideators and suicide attempters and only few variables that differentiated the two groups. Though the importance of single factors in differentiating suicide attempters and suicide ideators has recently been disputed (Huang et al., 2020), these findings point to the potential potency of (suicide-specific) RNT in understanding transitions to suicidal behavior.

A further analysis showed that the association between current suicide ideation and (lifetime) suicide attempts is completely mediated by suicide-specific rumination, that is, the risk of suicidal behaviour only increases when suicide is considered in a repetitive way. Within the metacognitive theory of emotional disorders, Wells and Matthews (2015) state that a psychological disorder results from an unhelpful thinking style called the Cognitive Attentional Syndrome (CAS). The CAS incorporates worry/rumination, threat monitoring and unhelpful thought control strategies. According to the theory, not single thoughts, assumptions or beliefs create emotional turmoil, but the way a person deals with these thoughts: only if respective thoughts activate the CAS, emotional and behavioral problems will follow. On this background one may assume that thoughts of suicide per se do not pose a great risk for suicidal behaviour (cf., McHugh et al., 2019), unless individuals engage in such thoughts in a repetitive manner. In future studies, the association between suicide-specific rumination and other variables of the metacognitive model should be investigated more closely.

The results of the current study should be interpreted with consideration of the following limitations. First, the PTSQ was developed for the current study and has only recently been subjected to stringent psychometric evaluation (Höller et al., in preparation). However, no direct comparison between the PTSQ and the Suicide Rumination Scale (SRS; Rogers & Joiner, 2018a) was made. Therefore, no conclusions with respect to the relationship between the two measures can be drawn, or determined whether one of the two measures is more valid in assessing suicide-specific rumination. Second, Rogers and Joiner (2018a) included a large number of control variables (e.g., depression, anxiety, insomnia, agitation, emotion regulation, general RNT) in their study on suicide specific

RNT, whereas in the present study only age, gender, general RNT and current suicide ideation were included as control variables. Future studies should therefore strive to investigate, whether suicide specific RNT – as assessed with the PTSQ – also outperforms such a great number of suicide risk factors in predicting the presence of a lifetime suicide attempt. Third, general RNT is understood as a trait (Watkins & Nolen-Hoeksema, 2014) and both the PTSQ and the SRS capture suicide specific RNT in a trait-like manner. Nevertheless, it is unclear whether suicide specific RNT is indeed stable over time and across suicidal crises and/or whether it is (only) associated with more intense suicidal crises (cf., Galynker, 2017). Prospective studies with repeated measurements are needed. Fourth, all of the constructs included in this study were measured exclusively via self- report assessments. Although it may be difficult to gather information regarding the frequency of particular thought patterns, participants may be prone to inaccuracy and uncertainty when responding to self-report items. Finally, the use of a cross-sectional research design and a sample comprised of predominantly Caucasians, limits the generalizability of the results and the discussion of temporal/causal relationships between study variables. This limitation is of specific importance considering the interpretation of the results of the mediation analysis: As all data were collected at a single measurement time-point and the outcome measure (i.e., lifetime suicide attempts) is retrospective, it might be more appropriate to frame the findings as indirect effects rather than as mediation effects. A replication of this study in treatment-seeking samples with prospective research designs would help to indicate whether the study results remain consistent in more at-risk populations. Still, it is important to emphasize that all participants within the current study reported lifetime suicidal ideation, and in this sense are a group of clinical interest.

Not least therefore, the current study does exhibit potential clinical implications: First of all, it might be important to account for the presence of suicide-specific rumination in addition to other risk factors, when assessing individuals for suicide risk. Furthermore, suicide-specific rumination may be a potential target in treatment to reduce one's suicidality. As such, (general) rumination has been shown to be malleable through treatments such as cognitive behavioral therapy (Teismann & Ehring, 2019) or mindfulness-based cognitive therapy (Gu et al., 2015). Therefore, it should be tested, whether suicide-specific rumination might be modifiable by similar interventions and techniques than general rumination. On the background of findings regarding the relevance of (suicide-specific) RNT in understanding suicidal behavior, respective studies seem highly warranted. Should the current findings be confirmed in further studies, it also seems reasonable to integrate suicide-specific rumination as a relevant factor with respect to the transition from suicide ideation to suicidal behavior within the current models of suicide ideation/behavior (cf., O'Connor & Kirtley, 2018).

Competing Interests: The dataset reported here is not part of any published or currently in-press works. The authors have no conflict of interests to declare.

Supplementary Materials

Perseverative Thinking about Suicide Questionnaire (PTSQ). The PTSQ is modeled after the Perseverative Thinking Questionnaire (Ehring et al., 2011) and assesses with nine items suicide specific rumination (for access see Index of Supplementary Materials below).

Index of Supplementary Materials

Teismann, T., Forkmann, T., Michalak, J., & Brailovskaia, J. (2021). *Supplementary Materials to "Repetitive negative thinking about suicide: Associations with lifetime suicide attempts"* [Questionnaire]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.5036

References

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics – Simulation and Computation, 39*(4), 860-864. https://doi.org/10.1080/03610911003650383

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ, USA: Lawrence Erlbaum.

Ehring, T., & Watkins, E. R. (2008). Repetitive negative thinking as a transdiagnostic process. *International Journal of Cognitive Therapy, 1*, 192-205. https://doi.org/10.1521/ijct.2008.1.3.192

Ehring, T., Zetsche, U., Weidacker, K., Wahl, K., Schönfeld, S., & Ehlers, A. (2011). The Perseverative Thinking Questionnaire (PTQ): Validation of a contend-independent measure of repetitive negative thinking. *Journal of Behavior Therapy and Experimental Psychiatry, 42*, 225-232. https://doi.org/10.1016/j.jbtep.2010.12.003

Field, A. (2013). *Discovering Statistic Using SPSS* (4th ed.). London, United Kingdom: SAGE.

Galynker, I. (2017). *The suicidal crisis*. New York, NY, USA: Oxford University Press.

Gu, J., Strauss, C., Bond, R., & Cavanagh, K. (2015). How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of meditation studies. *Clinical Psychology Review, 37*, 1-12. https://doi.org/10.1016/j.cpr.2015.01.006

PsychOpen GOLD

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics, 6*, 107-128. https://doi.org/10.3102/10769986006002107

Höller, I., Teismann, T., & Forkmann, T. (in preparation). Perseverative Thinking about Suicide Scale (PTSQ): Validation of a new measure to assess suicide specific rumination.

Horwitz, A. G., Czyz, E. K., Berona, J., & King, C. A. (2019). Rumination, brooding and reflection: Prospective associations with suicide ideation and suicide attempts. *Suicide and Life-Threatening Behavior, 49*, 1085-1093. https://doi.org/10.1111/sltb.12507

Huang, X., Ribeiro, J. D., & Franklin, J. C. (2020). The differences between suicide ideators and suicide attempters: Simple, complicated, or complex? *Journal of Consulting and Clinical Psychology, 88*, 554-569. https://doi.org/10.1037/ccp0000498

Kerkhof, A., & van Spijker, B. (2011). Worrying and rumination as proximal risk factors for suicidal behavior. In R. O'Connor, S. Platt, & J. Gordon (Eds.), *International handbook of suicide prevention: Research policy and practice* (pp. 199-210). Oxford, United Kingdom: John Wiley & Sons.

Klonsky, E. D., & May, A. M. (2015). The Three-Step Theory (3ST): A new theory of suicide rooted in the "Ideation-to-Action" framework. *International Journal of Cognitive Therapy, 8*, 114-129. https://doi.org/10.1521/ijct.2015.8.2.114

Law, K. C., & Tucker, R. P. (2018). Repetitive negative thinking and suicide: A burgeoning literature with need for further exploration. *Current Opinion in Psychology, 22*, 68-72. https://doi.org/10.1016/j.copsyc.2017.08.027

May, A. M., & Klonsky, E. D. (2016). What distinguishes suicide attempter from suicide ideators? A meta-analysis of potential factors. *Clinical Psychology: Science and Practice, 23*, 5-20. https://doi.org/10.1111/cpsp.12136

McHugh, C. M., Corderoy, A., Ryan, C. J., Hickie, I. B., & Large, M. M. (2019). Association between suicidal ideation and suicide: Meta-analyses of odds ratios, sensitivity, specificity and positive predictive value. *BJPsych Open, 5*(2), Article e18. https://doi.org/10.1192/bjo.2018.88

Miranda, R., & Nolen-Hoeksema, S. (2007). Brooding and reflection: Rumination predicts suicidal ideation at 1-year follow-up in a community sample. *Behaviour Research and Therapy, 45*, 3088-3095. https://doi.org/10.1016/j.brat.2007.07.015

Moulds, M. L., Bisby, M. A., Wild, J., & Bryant, R. A. (2020). Rumination in posttraumatic stress disorder: A systematic review. *Clinical Psychology Review, 82*, Article 101910. https://doi.org/10.1016/j.cpr.2020.101910

Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology, 109*, 504-511. https://doi.org/10.1037/0021-843X.109.3.504

Nolen-Hoeksema, S., Stice, E., Wade, E., & Bohon, C. (2007). Reciprocal relations between rumination and bulimic, substance abuse, and depressive symptoms in female adolescents. *Journal of Abnormal Psychology, 116*, 198-207. https://doi.org/10.1037/0021-843X.116.1.198

O'Connor, R. C., & Kirtley, O. J. (2018). The Integrated Motivational-Volitional Model of Suicidal Behaviour. *Philosophical Transactions of the Royal Society of London: Series B. Biological Sciences, 373*, Article 20170268. https://doi.org/10.1098/rstb.2017.0268

Penney, E. S., & Abbott, M. J. (2014). Anticipatory and post-event rumination in social anxiety disorder: A review of the theoretical and empirical literature. *Behaviour Change, 31*, 79-101. https://doi.org/10.1017/bec.2014.3

Robinson, M. S., & Alloy, L. B. (2003). Negative cognitive styles and stress-reactive rumination interact to predict depression: A prospective study. *Cognitive Therapy and Research, 27*, 275-291. https://doi.org/10.1023/A:1023914416469

Rockwood, N. J., & Hayes, A. F. (2020). Mediation, moderation, and conditional process analysis: Regression-based approaches for clinical research. In A. G. C. Wright & M. N. Hallquist (Eds.), *Handbook of research methods in clinical psychology* (pp. 396-414). Cambridge, United Kingdom: Cambridge University Press.

Rogers, M. L., & Joiner, T. E. (2017). Rumination, suicidal ideation and suicide attempts: A meta-analytic review. *Review of General Psychology, 21*, 132-142. https://doi.org/10.1037/gpr0000101

Rogers, M. L., & Joiner, T. E. (2018a). Suicide-specific rumination relates to lifetime suicide attempts above and beyond a variety of other suicide risk factors. *Journal of Psychiatric Research, 98*, 78-86. https://doi.org/10.1016/j.jpsychires.2017.12.017

Rogers, M. L., & Joiner, T. E. (2018b). Lifetime acute suicidal affective disturbance symptoms account for the link between suicide-specific rumination and lifetime past suicide attempts. *Journal of Affective Disorders, 235*, 428-433. https://doi.org/10.1016/j.jad.2018.04.023

Smith, J. M., Alloy, L. B., & Abramson, L. Y. (2006). Cognitive vulnerability to depression, rumination, hopelessness, and suicidal ideation: Multiple pathways to self-injurious thinking. *Suicide & Life-Threatening Behavior, 36*, 443-454. https://doi.org/10.1521/suli.2006.36.4.443

Smith, K. E., Mason, T. B., & Lavender, J. M. (2018). Ruination and eating disorder psychopathology: A meta-analysis. *Clinical Psychology Review, 61*, 9-23. https://doi.org/10.1016/j.cpr.2018.03.004

Spasojević, J., & Alloy, L. B. (2001). Rumination as a common mechanism relating depressive risk factors to depression. *Emotion, 1*, 25-37. https://doi.org/10.1037/1528-3542.1.1.25

Szabo, Y. Z., Warnecke, A. J., Newton, T. L., & Valentine, J. C. (2017). Rumination and posttraumatic stress symptoms in trauma-exposed adults: A systematic review and meta-analysis. *Anxiety, Stress, and Coping, 30*, 396-414. https://doi.org/10.1080/10615806.2017.1313835

Tabachnick, B., & Fidell, L. (2014). *Using multivariate statistics* (5th ed.). Boston, MA, USA: Pearson.

Takano, K., Sakamoto, S., & Tanno, Y. (2014). Repetitive thought impairs sleep quality: An experience sampling study. *Behavior Therapy, 45*, 67-82. https://doi.org/10.1016/j.beth.2013.09.004

Teismann, T. (2018). *Perseverative Thinking about Suicide Questionnaire.* Unpublished manuscript.

Teismann, T., & Brailovskaia, J. (2020). Entrapment, positive psychological functioning and suicide ideation: A moderation analysis. *Clinical Psychology & Psychotherapy, 27*, 34-41. https://doi.org/10.1002/cpp.2403

PsychOpen GOLD

Teismann, T., Brailovskaia, J., Schaumburg, S., & Wannemüller, A. (2020). High Place Phenomenon: Prevalence and clinical correlates in two German samples. *BMC Psychiatry, 20*, Article 478. https://doi.org/10.1186/s12888-020-02875-8

Teismann, T., & Ehring, T. (2019). *Pathologisches Grübeln*. Göttingen, Germany: Hogrefe.

Teismann, T., Forkmann, T., Glaesmer, H., Juckel, G., & Cwik, J. C. (2021). Skala Suizidales Erleben und Verhalten (SSEV): Factor structure and psychometric properties. *Diagnostica, 67*, 115-125. https://doi.org/10.1026/0012-1924/a000269

Urban, D., & Mayerl, J. (2006). *Regressionsanalyse: Theorie, Technik und Anwendung* (2. Aufl.). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin, 134*, 163-206. https://doi.org/10.1037/0033-2909.134.2.163

Watkins, E. R., & Nolen-Hoeksema, S. (2014). A habit-goal framework of depressive rumination. *Journal of Abnormal Psychology, 123*, 24-34. https://doi.org/10.1037/a0035540

Wells, A., & Matthews, G. (2015). *Attention and emotion: A clinical perspective (Classic Edition)*. Hove, United Kingdom: Psychology Press.

Wilkinson, P. O., Croudace, T. J., & Goodyer, I. M. (2013). Rumination, anxiety, depressive symptoms and subsequent depression in adolescents at risk for psychopathology: A longitudinal cohort study. *BMC Psychiatry, 13*, Article 250. https://doi.org/10.1186/1471-244X-13-250

PsychOpen GOLD

Research Articles

Burnout Subtypes: Psychological Characteristics, Standardized Diagnoses and Symptoms Course to Identify Aftercare Needs

Gianandrea Pallich [1,2] ID , Martin grosse Holtforth [3,4], Barbara Hochstrasser [1]

[1] *Center for Psychiatry and Psychotherapy, Private Hospital Meiringen, Meiringen, Switzerland.* **[2]** *Department of Clinical Psychology and Psychotherapy, University of Zurich, Zurich, Switzerland.* **[3]** *Department of Clinical Psychology & Psychotherapy, University of Bern, Bern, Switzerland.* **[4]** *Psychosomatic Medicine, Department of Neurology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland.*

Corresponding Author: Gianandrea Pallich, University of Zurich, Department of Psychology, Binzmühlestrasse 14, Box 1, 8050 Zurich, Switzerland. E-mail: g.pallich@psychologie.uzh.ch

Supplementary Materials: Materials [see Index of Supplementary Materials]

Abstract

Background: To better understand individual differences between burnout inpatients and improve individually tailored treatments in a psychiatric hospital, cluster analysis based on a number of self-report measures was used to investigate psychosocial characteristics of 96 participants.

Method: Group membership was analyzed regarding associations with standardized measures of psychiatric and personality disorders. Moreover, symptom levels of burnout, depression, and general mental health were used to characterize the groups and to observe differential trajectories at admission, discharge, and follow-up.

Results: As in previous research, we identified four subtypes that differed in comorbidity, psychological characteristics and treatment outcome. This calls for tailored interventions for the more vulnerable patients.

Conclusion: The replicated and enriched characterization of burnout inpatients can help to optimally meet the differential needs of burnout patients.

Keywords

depression, burnout, aftercare needs, diagnoses, symptoms, cluster analysis, subtypes

> **Highlights**
> - Four burnout subtypes were found: Functional, Dysfunctional, Straightforward Pragmatist, and Unhappy Altruist.
> - Psychosocial characteristics and symptom levels at admission, discharge, and follow-up were described to better characterize the subtypes.
> - The replicated and enriched characterization of burnout inpatients improves individually tailored treatments.

The Importance of Burnout

The term burnout was introduced to the scientific discussion of psychological ailments in the 1970s by Freudenberger as a label of a negative affective state after having been exposed to continued work-related stress experiences (Freudenberger, 1974). Later, Maslach and colleagues (Maslach & Jackson, 1981) embossed the concept of burnout, recognizing emotional exhaustion, depersonalization, and a reduced sense of personal accomplishment with a sense of a diminished level of performance to be the key dimensions of this phenomenon. Criticisms of this definition notwithstanding, the related questionnaire, the Maslach Burnout Inventory (MBI), has become the gold standard in research and literature (Burisch, 2014). Since then, the phenomenon of burnout has been described in more than 60 different professions and professional subgroups (Kaschka, Korczak, & Broich, 2011), showing a prevalence of burnout varying between 3.5% and 50% (Nil et al., 2010). Not surprisingly, the conception and improvement of the clinical treatment of burnout inpatients have also become an important research focus (Hochstrasser, Von Bardeleben, Ruckstuhl, & Soyka, 2008).

Long-Term Effects of an Inpatient Treatment Program for Burnout

Due to the heterogeneity and multifactorial etiology of burnout, a multimodal and individual treatment has been shown to be warranted (Hochstrasser et al., 2008; Schwarzkopf, Conrad, Straus, Porschke, & Von, 2016). Yet, the majority of studies on burnout interventions have not been performed with clinical samples, but in groups of volunteers who exhibited a level of burnout allowing them to maintain active engagement at work (Ahola et al., 2017; Awa et al., 2010; van der Klink et al., 2001). Patients with burnout who need inpatient care are those who are more afflicted, i.e., those who suffer from clinical burnout. Despite the importance of an adequate and effective inpatient treatment for burnout, to date, only few studies have examined the short- or long-term effects of inpatient treatment programs for burnout (Elkuch et al., 2010; Perski et al., 2017; Schwarzkopf et al., 2016). A previous study (Elkuch et al., 2010) examining a multimodal inpatient treatment at a private psychiatric hospital has found evidence of positive effects. Treatment included cognitive-behavioral individual and group psychotherapy, various relaxation techniques, body therapy, physical exercise, and psycho-

pharmacological treatment. However, one limitation of the previous study was that assessments were performed only at admission and at follow-up, but not at discharge. Moreover, it has to be considered that the potential long-term effects of the inpatient treatment program and its sustainability may develop in the period between discharge and follow-up. Thus, assessing patients at admission, discharge, and at follow-up allows the examination of the short-term effects and the unfolding process of long-term effects more accurately. The expected results promise to yield valuable information serving the ongoing optimization of the future inpatient treatment of burnout.

The Importance of Characterizing Patients Discharged From Inpatient Treatment for Burnout

Identifying burnout patients' subtypes is crucial to tailoring treatment to patient characteristics and thereby improving burnout treatment. At an empirical level, some studies have identified subjects with burnout symptoms as one of several types of respondents in the workforce. Schaarschmidt and Fischer (2001) used self-report data on personal experiences with work-related stress and typical coping behaviors using the AVEM questionnaire (Work-related Behavior and Experience Patterns; German: Arbeitsbezogenes Verhaltens- und Erlebensmuster; Schaarschmidt & Fischer, 1996) to empirically categorize subjects in the workforce. The AVEM assesses stress experiences and coping behaviors in three domains and 11 subscales of six items each: *work commitment*, *resistance to stress/emotions*, and *subjective well-being* (Schaarschmidt & Fischer, 2001). The domains and subscales were identified by factor analyses of responses of 1598 subjects of diverse professions, and the AVEM has been subsequently used in various studies (Schulz et al., 2011; Voltmer et al., 2007, 2010, 2011). In the original study, Schaarschmidt and Fischer (2001) empirically identified four types of subjects based on scores in the 11 subscales: Healthy (Pattern G), Unambitious (Pattern S), Overexertion (Risk pattern A), and Burnout (Risk pattern B). In a recent study based on a sample of 1766 health care employees, Leiter and Maslach (2016) proposed five empirical profiles emerging from latent profile analyses of their dimensions of burnout (i.e., emotional exhaustion, depersonalization, and a reduced sense of personal accomplishment): Burnout (high on all three dimensions), Engagement (low on all three), Overextended (high on exhaustion only), Disengaged (high on cynicism only), and Ineffective (high on inefficacy only).

At a theoretical level, Montero-Marin and colleagues (Montero-Marín et al., 2009) proposed a three-partite classification of burnout patients based on a general proposal by Farber (1991): frenetic (involved and ambitious subjects who sacrifice their health and personal lives for their jobs); under-challenged (indifferent and bored workers who fail to find personal development in their job); and worn-out (subjects who feel they have little control over results and that their efforts go unacknowledged). Haberthür and colleagues (Haberthür et al., 2009) empirically classified burnout inpatients using self-report data on various interpersonal and intrapersonal aspects of functioning, such

PsychOpen GOLD

as social support, interpersonal problems, coping styles, emotion regulation, and motivational incongruence. The authors identified four groups by cluster analyses: Functional, Dysfunctional, Straightforward Pragmatist, and Unhappy Altruist.

For the current study, data were collected in the same private hospital and the same treatment unit as in the Haberthür et al. study. To our knowledge, the results of Haberthür et al.'s study (Haberthür et al., 2009) have not been replicated yet. The study did not assess standardized clinical diagnoses of psychiatric disorders and personality disorders, or comorbid somatic diagnoses, nor did it assess outcome at discharge.

The present study attempts to overcome these limitations and to replicate the former empirical classification of burnout inpatients to allow practitioners to tailor individual treatments to improve treatment outcomes. The self-reported person characteristics examined in Haberthür et al.'s study were motivational incongruence (motive satisfaction), interpersonal problems, social support, regulation of emotions, and coping styles. In the current study, the self-report measures used for clustering were the same as those used by Haberthür and colleagues, with the addition of as a self-report screening tool for personality dysfunction. Refining the clinical assessment methodology, structured interviews for psychiatric diagnosis and personality disorder were conducted. At admission, discharge, and follow-up, we assessed levels of depression, general symptoms and burnout.

Aims

The aims of this study are: 1. To constructively replicate and improve a previously empirically derived description and categorization of burnout inpatients in an analogous treatment setting according to psychosocial parameters; 2. To characterize the patients and patient groups according to psychiatric diagnostic criteria; 3. To observe how group membership corresponds to different levels of psychological symptoms (depression and burnout) and general mental health at admission, discharge, and follow-up.

Material and Method

Sample, Treatment, and Recruitment

The present study was approved by the ethics committee of the canton Bern (Switzerland) and was conducted in the Private Hospital Meiringen. The sample comprised 96 inpatients of a specialized burnout ward. The therapeutic program includes individual psychotherapy, group therapy, relaxation techniques, body therapy, massages, sports activities and fitness instructions, psychopharmacotherapy, and selected interventions from complementary medicine (e.g., traditional Chinese medicine). A detailed description of the treatment program can be found in Hochstrasser et al. (2008).

The specialized burnout ward admits only patients being referred by a physician, having a burnout syndrome that arose primarily in the context of the work environment, and with a diagnostically confirmed burnout syndrome at admission evaluated in a clinical interview before admission. In this context, it is important to note that in the ICD-10, burnout is not considered to qualify as an independent psychiatric disorder but is listed as a syndrome being associated with difficulties pertaining to life circumstances (i.e., ICD-10, Z.73.0). An association of burnout with mental disorders, especially depression, has often been described, such that a recent overview on the overlap between depression and burnout postulated that clinical burnout corresponds to an atypical depression (Bianchi et al., 2015). Consequently, various comorbid primary psychiatric diagnoses according to ICD-10, Chapter F, were given on the basis of a clinical interview and in accordance with the patients' symptomatic presentation at admission. To be included, patients had to be at least 18 years old. Patients were excluded if they exhibited current alcohol or drug addictions (if not stopped at admission), inability to participate in the treatment (e.g., due to psychological disorders or dementia), insufficient knowledge of the German language, or acute suicidality or psychotic symptoms. Between February 2017 and December 2017, a total of 173 inpatients were asked to participate in the study, a total of 113 inpatients gave their consent, and, due to missing data in cluster-relevant questionnaires, a total of 96 individuals, $n = 96$, $f = 33$ (34.4%), $m = 63$ (65.6%), were included in the analyses.

Instruments

During the first week after admission, participants completed paper-pencil versions of different questionnaires and participated in two clinical interviews (Mini-DIPS and SCID-II) (Fydrich et al., 1997; Margraf, 2013) administered by the study psychologist. The discharge assessment was done in the last week of their stay, and the follow-up assessment was administered three months after discharge via paper-pencil questionnaires sent by mail with a pre-paid return envelope.

As in Haberthür et al. (2009), psychological characteristics were measured using the following self-report instruments: First, a short version of the Incongruence Questionnaire (German: Inkongruenzfragebogen, K-INK; grosse Holtforth & Grawe, 2003) was used to assess the degree of insufficient motivational satisfaction (approach incongruence and avoidance incongruence). The German 32-item short version of the Inventory for Interpersonal Problems was used to assess problematic interpersonal behaviors (IIP-SC; Soldz, Budman, Demby, & Merry, 1995; German: grosse Holtforth, 2005). The 32 items are an equivalent subset of the German IIP (IIP-D; Horowitz, Strauss, & Kordy, 2000). To measure the subjective appraisal of received or anticipated social support from persons in the social environment, the German short version of the Questionnaire of Social Support was used (German: Fragebogen zur sozialen Unterstützung, F-SOZU-K-22; Fydrich, Sommer, & Brähler, 2007). To evaluate different ways of coping with stressful

situations (task-oriented, emotion-oriented, avoidance-oriented), the German version of the Coping Inventory for Stressful Situations (CISS; Kälin, 1995) was used. The Questionnaire for the Self-Evaluation of Emotional Competency (German: Fragebogen zur Selbsteinschätzung emotionaler Kompetenzen, SEK; Berking & Znoj, 2008) was used to measure deficits and resources in emotion regulation with the following scales: attention, awareness of bodily sensations, clarity, understanding, regulation, acceptance, resilience, self-support, and goal-oriented readiness to confront. The scale scores can be summarized by a total score. As mentioned before, previous studies did not assess personality and personality dysfunctions. To fill this gap, we added the Inventory of Personality Organization (IPO-16; Zimmermann et al., 2013) for clustering purposes. The 16-item short version of the Inventory of Personality Organization (IPO-16) is a self-report measure assessing the severity of personality dysfunction.

The level of symptoms and problems were assessed using the Beck Depression Inventory (BDI; Hautzinger et al., 1995), a brief version of the Symptom Checklist SCL-90 (SCL-9; Klaghofer & Brähler, 2001) and the Maslach Burnout Inventory – Human Services Survey (MBI-HSS; Maslach et al., 1997). The BDI is a self-report instrument assessing the degree of depressive symptomatology. The brief version of the Symptom Checklist assesses the general level of symptoms in one scale (Hautzinger et al., 1995). The MBI-HSS is considered the gold standard for burnout assessment and measures burnout in three dimensions (emotional exhaustion, depersonalization, and a sense of reduced personal effectiveness) (Maslach et al., 1997).

Data Analytical Approach / Statistical Analysis

All statistical analyses were performed using the SPSS program (version 23.0) and Jamovi (version 0.8.6.0) (an interface program based on R). In a first step, a hierarchical cluster analysis (Ward's Method) was performed to determine the appropriate number of clusters. Euclidean distance, which does not weigh outliers as strongly as the quadrated Euclidean distance, was used. According to these criteria, a cluster solution of four groups was considered optimal.

The following questionnaires and scales were used for clustering: the Incongruence Questionnaire (K-INK; Approach and Avoidance incongruence), the Inventory for Interpersonal Problems (IIP-SC/IIP-D; Dominance and Affiliation Dimensions), the Questionnaire of Social Support (F-SOZU-K-22; general score), the Self-Report Measure for the Assessment of Emotion Regulation Skills (SEK; general score); the Coping Inventory for Stressful Situations (CISS; task-oriented coping, emotion-oriented coping, avoidance-oriented coping) and the Inventory of Personality Organization (IPO-16; general score).

The number of clusters, i.e., four, corresponds to the number of clusters proposed by Elkuch et al. (2010). On the basis of the solutions suggested by the hierarchical cluster analysis, we further calculated confirmatory *k*-means cluster analyses for four cluster solutions. To exclude bias resulting from differing scaling of the various variables, all

cluster analyses were performed using *z*-standardized values. Consequently, the values of the resulting groups were *z*-standardized and are presented as norm-related *z*-standardized values (see Figure 1). Descriptive statistics were used to describe the frequencies of different clinical diagnoses and personality disorder diagnoses (see Table 1).

Figure 1

z-Standardized Levels of Psychological Characteristics Used for the Formation of the Four Groups and Duration of Hospital Stay (No Grouping Characteristic)
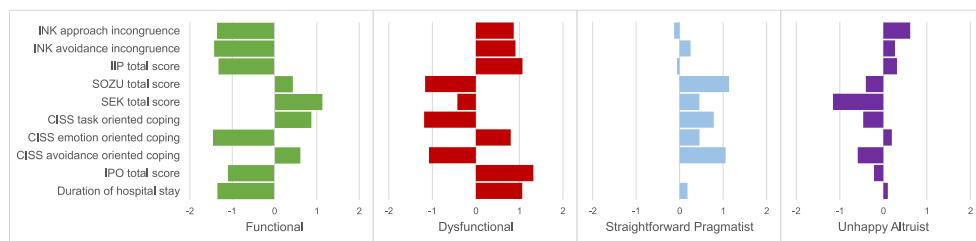
BDI, SCL-9, and MBI-HSS were used to further characterize the groups (but not as factors to identify the groups) and to observe longitudinal development of symptoms. To evaluate differences in symptom levels among the resulting groups, we calculated a repeated measure analysis of variance (ANOVA) (see Figure 2).

Results

Sample Description

A total of 96 patients were included in the analyses. The mean age at admission was 48.02 years (*SD* = 8.78; 27.44 − 62.79 years). 33 (34.4%) of the participants were female, 63 (65.6%) were male. 50 married, 14 divorced, 25 singles, 3 separated, 1 widowed, and 1 unknown. The mean duration of the hospital stay was 57.31 days (*SD* = 16.04; 9 − 94 days). All the participants received medication during clinical stay. The duration between the time of discharge from the hospital and the follow-up assessment was 3 months. At follow-up, 14.6% participants were unemployed, 10.4% were fully employed, 33.3% were working part-time, 2.1% were working in their own household, 1.0% was in training for a different job, 3.1% were in a rehabilitation program, 6.3% were receiving a pension (i.e. an amount of money paid regularly by a government or company to somebody who has retired from work) or a disability pension (i.e. a form of pension given to those people who are permanently or temporarily unable to work due to a disability), and the employment status of 29.1% was unknown. In comparison to before the impatient stay, 6.3% were unemployed, 49.0% were fully employed, 20.8% were working part-time,

Table 1

Main Psychiatric Diagnoses, Presence of an Additional Psychiatric Diagnosis, Presence of an Additional Somatic Diagnosis, and Personality Disorders for the Four Groups at Admission

	Total sample	Functionals	Dysfunctionals	Straightforward Pragmatists	Unhappy Altruists
Main Psychiatric Diagnoses (Mini-DIPS)					
F31.x (Bipolar disorder)	5.21%	5.88%	15.79%	0.00%	2.70%
	($n = 5$)	($n = 1$)	($n = 3$)	($n = 0$)	($n = 1$)
F32.x (Major depressive disorder, single episode)	34.37%	17.65%	15.79%	43.48%	45.95%
	($n = 33$)	($n = 3$)	($n = 3$)	($n = 10$)	($n = 17$)
F33.x (Major depressive disorder, recurrent)	46.87%	47.05%	57.90%	39.13%	45.95%
	($n = 45$)	($n = 8$)	($n = 11$)	($n = 9$)	($n = 17$)
F43.x (Reaction to severe stress, and adjustment disorders)	6.25%	11.76%	5.26%	13.04%	0.00%
	($n = 6$)	($n = 2$)	($n = 1$)	($n = 3$)	($n = 0$)
Missing	7.29%	17.65%	5.26%	4.35%	5.40%
	($n = 7$)	($n = 3$)	($n = 1$)	($n = 1$)	($n = 2$)
Presence of Comorbid Psychiatric Diagnoses (Mini-DIPS)	33.33%	23.52%	42.08%	24.21%	37.80%
	($n = 32$)	($n = 4$)	($n = 8$)	($n = 6$)	($n = 14$)
Presence of Comorbid Somatic Diagnoses	33.33%	35.28%	31.56%	21.75%	40.50%
	($n = 32$)	($n = 6$)	($n = 6$)	($n = 5$)	($n = 15$)
Avoidant personality Disorder (PD) & Obsessive-Compulsive PD (possibly comorbid with additional PD)	6.25%	0.00%	21.05%	4.35%	1.70%
	($n = 6$)	($n = 0$)	($n = 4$)	($n = 1$)	($n = 1$)
Avoidant PD (possibly comorbid with additional PD)	4.17%	5.88%	5.26%	0.00%	5.41%
	($n = 4$)	($n = 1$)	($n = 1$)	($n = 0$)	($n = 2$)
Obsessive-compulsive PD (possibly comorbid with an additional PD)	23.96%	5.88%	21.05%	34.78%	27%
	($n = 23$)	($n = 1$)	($n = 4$)	($n = 8$)	($n = 10$)
Other PD	5.21%	5.88%	10.52%	4.35%	1.70%
	($n = 5$)	($n = 1$)	($n = 2$)	($n = 1$)	($n = 1$)
No PD	60.42	82.35%	42.11%	56.52%	62.16%
	($n = 58$)	($n = 14$)	($n = 8$)	($n = 13$)	($n = 23$)

PsychOpen GOLD

Figure 2

Repeated ANOVAs for the Four Groups (Functionals, Dysfunctionals, Straightforward Pragmatists and Unhappy Altruists) at Intake, Discharge and Follow-Up for BDI and SCL and at Intake and Follow-Up for the Three Dimensions of MBI (Emotional Exhaustion, Depersonalization and Sense of Reduced Personal Effectiveness)
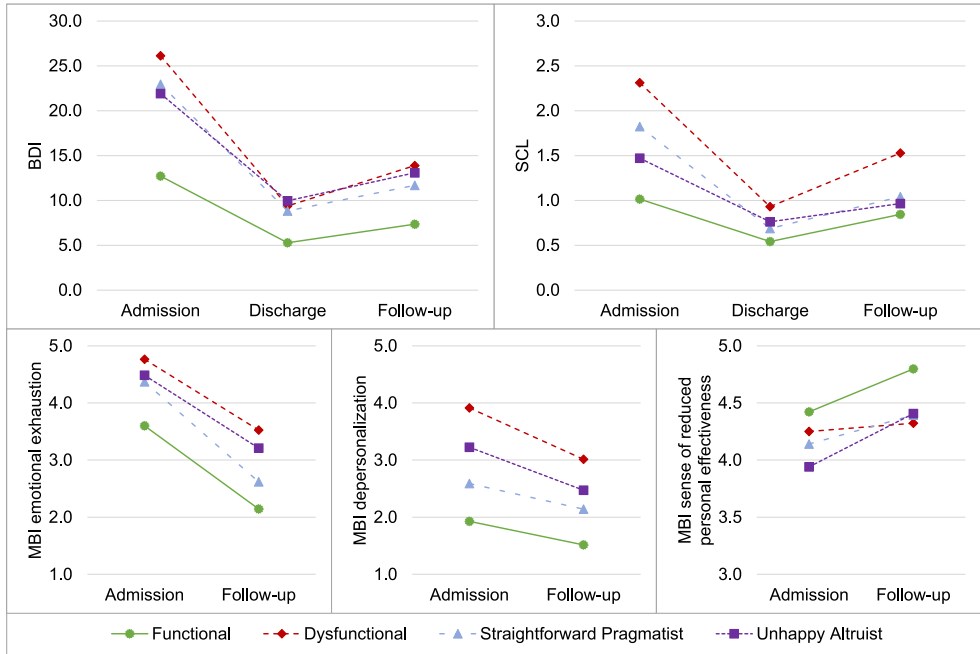
1.0% was working in their own household, none was in training, in a rehabilitation program, or receiving a pension / disability pension, and the employment status of 22.9% participants was unknown.

Psychosocial Characteristics

Generally, the group labels are intended to be maximally comprehensive summaries of the respective characteristics. With the current sample and measures, we found that the obtained clusters corresponded closely to the previous grouping by Haberthür et al. (2009), so that we decided to keep the previous labels: (a) Functional, (b) Dysfunctional, (c) Straightforward Pragmatist, and (d) Unhappy Altruist.

Functionals

Participants categorized in this group, $n = 17$, $f = 3$ (17.6%), $m = 14$ (82.4%), experienced little avoidance incongruence ($z = -1.43$) and approach incongruence ($z = -1.36$). The

PsychOpen GOLD

patients mentioned few interpersonal problems (z = -1.32) and having good social support (z = 0.43). In addition, they reported good emotional competences (z = 1.13). The Functionals group used many task-oriented (z = 0.87) but just a few emotion-oriented (z = -1.45) coping strategies. They reported having little personality dysfunctions (z = -1.10). In general, they had a shorter stay in the hospital (z = -1.35).

Most of the participant in the group of Functionals had an F33.x (i.e., major depressive disorder, recurrent) diagnosis (n = 8, 47.05%), three (17.65%) had an F32.x (i.e., major depressive disorder, single episode) diagnoses, two (11.76%) an F43.x (i.e., reaction to severe stress, and adjustment disorders diagnoses) and one (5.88%) an F31.x (i.e., bipolar disorder) diagnoses. A total of four (23.52%) had a secondary psychiatric diagnosis (e.g., F10.1, F40.2, F41.0, F42.2). Six participants (35.29%) of this group additionally had one or more somatic diagnoses (e.g., E78.0, G44.0, G44.2, H95.1, I10.90, I10.91, R05, R73.1, Z61, Z62, Z73). For three participants it was not possible to use the Mini-DIPS for assessing standardized diagnoses.

Most of the participants categorized in the group of the Functionals showed no personality disorder (82.35%, n = 14).

Dysfunctionals

Compared to the other three groups, participants categorized in the group of Dysfunctionals, n = 19, f = 7 (36.8%), m = 12 (63.2%), showed the highest average approach incongruence (z = 0.87) as well as avoidance incongruence (z = 0.91). They showed strong interpersonal problems (z = 1.07). Additionally, they reported least social support (z = -1.17), generally insufficient emotional competence (z = -0.43) and mainly emotional coping (z = 0.80) and little task-oriented (z = -1.19) and avoidance-oriented (z = -1.08) coping strategies. In addition, they reported many personality dysfunctions (z = 1.32). In general, they had a longer stay in the hospital (z = 1.06).

Most of the participants in this group had an F33.x diagnosis (n = 11, 57.90%), three (15.79%) had an F32.x diagnosis, three (15.79%) an F31.x diagnosis and one (5.26%) an F43.x diagnosis. A total of eight (42.12%) had a secondary psychiatric diagnosis (e.g., F13.2, F40.2, F41.0, F41.1, F42.1, F43.1, F50.5). Six participants (31.59%) of this group additionally had one or more somatic diagnoses (e.g., A49.8, E03.9, E14.91, E78.5, G40.9, G43.9, G47.0, G47.39, I10.9, J45.0, M54.4, N48.0). One participant was not diagnosed systematically with Mini-DIPS.

The Dysfunctionals showed the highest association with a combination of avoidant and obsessive-compulsive personality disorders (21.05%, n = 4), and a high percentage had an obsessive-compulsive personality disorder (21.05%, n = 4) or other personality disorders (10.52%, n = 2).

Straightforward Pragmatists

On average, this group, n = 23, f = 13 (56.5%), m = 10 (43.5%), showed more avoidance incongruence (z = 0.25) than approach incongruence (z = -0.13). Generally, they reported a low level of interpersonal problems (z = -0.06). This group reported having good social support (z = 1.14). They generally had good emotional competences (z = 0.45). They reported using emotional coping (z = 0.46) and task-oriented coping (z = 0.78) at similar levels. This group showed average personality dysfunction (z = 0.00). The hospital stay was a little higher than average (z = 0.18).

Most of the participants in this group had an F32.x diagnosis (n = 10, 43.48%), nine (39.13%) had an F33.x diagnosis, three (13.04%) an F31.x diagnosis and none (0.00%) an F41.x diagnosis. A total of six (26.09%) had a secondary psychiatric diagnosis (e.g., F40.0, F40.2, F41.0, F44.4, F50.3). Five participants (21.74%) of this group additionally had one or more somatic diagnoses (e.g., E66.99, G35.9, G43.9, G47.31, H93.1, I10.90, I49.8). For one participant it was not possible to use the Mini-DIPS for assessing standardized diagnoses.

The group of the Straightforward Pragmatists showed the highest prevalence of obsessive-compulsive personality disorder compared to the other groups.

Unhappy Altruists

The members of this group, n = 37, f = 10 (27.0%), m = 27 (73.0%), showed higher average approach incongruence (z = 0.62) than avoidance incongruence (z = 0.27). Overall, they tended to show above-average scores in interpersonal problems (z = 0.31). Additionally, they reported bad social support (z = -0.40). Furthermore, this group showed an emotional competence below the average (z = -1.16). The members of this group primarily used emotion-oriented coping strategies (z = 0.19) and few task-oriented coping strategies (z = -0.46). This group showed little personality dysfunction (z = - 0.22). The hospital stay was a little longer than average (z = 0.10).

Seventeen participants (45.95%) of this group had an F32.x diagnosis. Seventeen participants 45.95%) had an F33.x diagnosis, and one participant (2.70%) had an F31.x diagnosis. No participants were diagnosed with F43.x in this group. A total of fourteen patients in this group (37.83%) had a secondary psychiatric diagnosis (e.g., F10.1, F13.2, F40.1, F40.2, F41.0, F44.2, F61.0). Fifteen participants (40.54%) had one or more somatic diagnoses in addition (e.g., D17.3, E11.90, E78.0, G25.0, G25.81, G43.0, G43.9, G47.1, G47.31, H93.1, H93.3, I10.90, K91.1, M17.9, M19.91, M53.0). For two participants it was not possible to use the Mini-DIPS for assessing standardized diagnoses.

The group of the Unhappy Altruists, similar to the group of the Straightforward Pragmatists, showed a higher percentage of obsessive-compulsive personality disorder compared to the other groups.

Symptom Course in Groups of Burnout Patients

Data were analyzed using repeated measures ANOVAs for depression, general symptoms, and burnout with a within-subjects factor (admission, discharge, follow-up) and a between-subject factor of subtypes (Functional, Dysfunctional, Straightforward Pragmatist, and Unhappy Altruist). For missing values, list-wise deletion of cases was applied. For the three MBI dimensions, there were just two measured time points (admission and follow-up). Only for the repeated measures ANOVA for depression, the Mauchly's test indicated that the assumption of sphericity had been violated, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.876$).

For the interested reader we report mean, median, standard deviation and range for BDI, SCL and MBI for the four groups (i.e. Functional, Dysfunctional, Straightforward Pragmatist and Unhappy Altruist) in the Supplementary Material

Depression (Beck Depression Inventory, BDI)

A repeated measures ANOVA (see Figure 2) with a Greenhouse-Geisser correction showed that mean depression scores differed significantly between time points, $F(1.75, 127.86) = 117.55$, $p < .001$, $\eta_p^2 = .617$, and between groups, $F(3, 73) = 4.46$, $p < .01$, $\eta_p^2 = .158$. The interaction between time and groups was also significant, $F(5.25, 127.86) = 2.51$, $p < .05$, $\eta_p^2 = .093$. Post hoc tests revealed that the depression scores for the Functionals group was the lowest and differed highly significantly from those of all other groups regarding ($p < .001$). For the remaining groups, depressive symptoms were higher at admission, all at similar and non-significantly different levels. At discharge, levels of depressive symptoms did not differ significantly between the four groups. However, at follow-up, the average depression levels of the Functional and Dysfunctional groups differed significantly ($p < .05$), and also a significant difference between Functionals and Unhappy Altruists ($p < .05$) was found. All patient groups showed a significant decrease of depressive symptoms from admission to discharge ($p < .05$). Whereas Functionals, Straightforward Pragmatists, and Unhappy Altruists reported no significant increase of depressive symptoms between discharge and follow-up, the Dysfunctionals showed a significant increase of depressive symptoms ($p < .05$).

General Symptoms (Brief Symptom Checklist, SCL)

A repeated measures ANOVA (see Figure 2) showed that the mean general symptoms scores differed significantly between time points, $F(2, 154) = 53.23$, $p < .001$, $\eta_p^2 = .409$, and between groups, $F(3, 77) = 7.24$, $p < .001$, $\eta_p^2 = .220$. The interaction between time and groups was also significant, $F(6, 154) = 2.60$, $p < .05$, $\eta_p^2 = .092$. The Functionals had the lowest symptom level, and the Dysfunctionals the highest at admission, discharge, and follow-up. Straightforward Pragmatists and Unhappy Altruists showed similar levels of general symptoms, with levels being lower than those of Dysfunctionals but higher compared to the Functional group at all three measurement points. Post hoc tests

revealed that at admission, the Functionals differed highly significantly regarding the general symptoms from the Dysfunctional and Straightforward Pragmatists ($p < .01$) and significantly from the Unhappy Altruists ($p < .05$). The group of Dysfunctionals differed highly significantly from the group of Unhappy Altruists ($p < .01$) and significantly from the group of Straightforward Pragmatists ($p < .05$). Regarding general symptoms at discharge, no significant differences could be found between the four groups. At follow-up Dysfunctionals differed from all other groups ($p < .05$). All groups showed a significant decrease of symptoms between admission and discharge ($p < .05$). Whereas Dysfunctionals and Straightforward Pragmatists showed significant increases of symptoms between discharge and follow-up ($p < .05$), this was not the case for the Functionals and the Unhappy Altruists.

Burnout (Maslach Burnout Inventory, MBI)

A repeated measures ANOVA for the dimension of emotional exhaustion showed a significant difference between admission and follow-up, $F(1, 71) = 56.87$, $p < .001$, $\eta_p^2 = .445$, and between groups, $F(3, 71) = 3.10$, $p < .05$, $\eta_p^2 = .116$. The interaction between time and groups was not significant. Post hoc tests revealed that regarding emotional exhaustion, all groups showed a significant ($p < .01$) decrease between admission and follow-up. Functionals and Dysfunctionals showed a significantly different level of emotional exhaustion at admission ($p < .01$).

A second repeated measures ANOVA for MBI for the dimension of depersonalization showed a significant difference between admission and follow-up, $F(1, 71) = 11.83$, $p < .001$, $\eta_p^2 = .143$, and between groups, $F(3, 71) = 4.54$, $p < .01$, $\eta_p^2 = .161$. The interaction between time and groups was not significant. Post hoc tests showed that Functionals differed from Dysfunctionals and Unhappy Altruists significantly concerning depersonalization at admission ($p < .05$). Additionally, Dysfunctionals differend from Straightforward Pragmatists ($p < .05$). Dysfunctionals' and Unhappy Altruists' level of depersonalization decreased significantly ($p < .05$) between admission and follow-up.

A last repeated measures ANOVA for the third MBI dimension, the sense of reduced personal effectiveness, showed a significant difference between admission and follow-up, $F(1, 71) = 4.61$, $p < .05$, $\eta_p^2 = .061$, and no significant difference between groups. The interaction between time and groups was not significant. In a post hoc analysis, Unhappy Altruists reported, as the only group, a significant ($p < .05$) improvement in the sense of reduced personal effectiveness.

Discussion

In the present study, we set out to reproduce the results and improve the previous descriptions of burnout inpatients of Haberthür and colleagues (2009). First, a cluster analysis was used to group burnout patients. Second, we characterized the burnout

patients according to psychosocial parameters. Additionally, the groups were described regarding their residual symptoms at admission, discharge, and follow-up. Finally, we described the psychiatric, somatic, and personality disorder diagnoses of the sample. Four groups were identified based on clustering (i.e., Functional, Dysfunctional, Straightforward Pragmatist, and Unhappy Altruist).

The Functional group was characterized by low levels of motivational incongruence, interpersonal problems, emotion-oriented coping, and personality dysfunction and showed good social support, emotional regulation, and mainly task-oriented and avoidance-oriented coping. The members of the Dysfunctional group had an almost reversed profile showing high levels for incongruence, interpersonal problems, emotion-oriented coping, and personality dysfunction in addition to low social support, emotional regulation, task- and avoidance-oriented coping. The other two groups (Straightforward Pragmatists and Unhappy Altruists) did not show characteristically extreme values in the above-mentioned variables. Straightforward Pragmatists reported good social support, emotional competences, and using all three coping strategies. Unhappy Altruists reported levels of incongruence and interpersonal problems a little above average, low social support, and emotional competences as well as stronger use of emotion-oriented than task-oriented or avoidance-oriented coping strategies.

All psychosocial characteristics of the Functional group could be reproduced without exception as described by Haberthür and colleagues (2009). The group of the Dysfunctionals had similar psychosocial parameters as found in the previous study, with the exception of task-oriented coping that was found to be low instead of average. For the other groups (i.e., Straightforward Pragmatists and Unhappy Altruists), we found similar psychosocial parameters as in the previous study. Only the emotional competence of the Straightforward Pragmatists was found to be high and not average, and reported levels of emotional competence and social support of the Unhappy Altruists were found to be low instead of average.

Above and beyond replicating the description by psychosocial parameters, also psychiatric, somatic, and personality disorder diagnoses were assessed for the four subtypes of burnout patients. The standardized assessment of psychiatric diagnoses showed most of the participants of the groups of Functionals and Dysfunctionals having a recurrent major depressive disorder. Furthermore, most of the participants in the group of the Straightforward Pragmatists had a single major depressive disorder. Finally, the participants of the group of the Unhappy Altruists had the same frequency of recurrent major depressive disorder and single major depressive disorder. Interestingly, all groups showed some comorbidity of somatic diagnoses. Around one third of the Functionals and Dysfunctionals had, in addition to psychiatric diagnoses, also a somatic diagnosis, whereas around one quarter of the group of the Straightforward Pragmatists had a somatic diagnosis. Finally, 40.50% of the Unhappy Altruists had a diagnosis of a somatic disorder.

In previous studies, a high level of overlap between burnout and depression symptomatology was found in all groups of burnout patients, to the point that it has been suggested that clinical burnout may rather be a form of depression (Bianchi et al., 2015). The DGPPN (Berger et al., 2012) proposed to consider depression as a common consequence of prolonged burnout. Yet, the temporal relationship between burnout and depression remains unclear (Ahola & Hakanen, 2007). In order to require inpatient care, burnout patients are likely to be more strongly affected and more impaired regarding daily functioning, which might bias the sample we examined towards those with a depressive disorder or other mental disorders.

Functionals showed almost no personality disorders, Dysfunctionals had a higher prevalence of personality disorders (especially Avoidant PD in combination with Obsessive-compulsive PD and Obsessive-compulsive PD), and both Straightforward Pragmatists and Unhappy Altruists showed a high prevalence of Obsessive-compulsive PD. The results of the IPO-16, assessing the severity of personality dysfunction, seem to confirm these findings, indicating similar personality dysfunctions as found through SCID-II for the four groups (very low for Functionals, very high for Dysfunctionals, average for Straightforward Pragmatists, and low for Unhappy Altruists). These findings could be relevant for planning tailored treatments for burnout patients considering that treatment of personality disorder is a major goal of psychotherapy interventions for all groups except the Functionals.

Symptom level at admission, discharge, and follow-up was assessed for the four groups using the BDI, the SCL, and the MBI. Generally, all four groups improved significantly between admission and discharge regarding depressive symptoms and overall symptom level. Dysfunctionals showed an increase of depressive symptoms between discharge and follow-up. Dysfunctionals and Straightforward Pragmatists showed a significant increase of general symptoms between discharge and follow-up. This worsening should be considered for discharge planning for the group of Straightforward Pragmatists and even more so for the Dysfunctionals. Particularly for Dysfunctionals, increased attention to discharge planning and more intensive support after leaving the clinic seems indicated.

Practical Implications

This study suggests that it is of great importance to attend to the relevant psychological characteristics of burnout patients, and that applying our categorization early in the process could improve the success of treatment and discharge planning. This may be done by clinical judgment or, if available, also by using structured assessment tools. Depending on burnout group membership, the needs of inpatients are likely to be different.

As indicated by group label, Functionals generally show more benign characteristics and are more likely to improve during the inpatient treatment, an effect that appears sustained during follow-up. Under the perspective of optimal resource allocation, frequent

monitoring of patients' mental health might suffice to meet their needs for care after discharge.

In contrast, our data suggest that the group of Dysfunctionals may need the most intensive inpatient treatment and well-organized psychosocial aftercare. Our data suggest that increasing the level of a patient's motivational satisfaction needs to be an important treatment goal, and the identification of the individual sources of motivational incongruence will help to select targeted interventions. Also, due to strong interpersonal problems, lower levels of perceived social support, and insufficient emotional competence, assertiveness training (Rakos, 1991), activation of the patient's social network (Perry & Pescosolido, 2015; Pescosolido & Levy, 2002; Smith & Christakis, 2008), as well as training of emotional skills (Berking, 2015; Cherniss, 2000; Pallich et al., 2020) might be suitable interventions. The enhancement of a task-oriented coping style and related skills may be an important target for longer-term treatment of this group.

Whereas Straightforward Pragmatists showed rather unproblematic profiles at admission with regard to psychological characteristics, they showed significant increases of symptoms between discharge and follow-up. For this reason, assessment during inpatient treatment should exceed patient self-reports to not miss relevant stressors that patients might not be able or willing to report. In addition, the formulation of crisis-response plans may be indicated, as well as close symptom monitoring after discharge.

Finally, also Unhappy Altruists, who show interpersonal problems above the average, bad social support, and emotional competence below the average, could profit from assertiveness training (Rakos, 1991) and a training of emotional competence (Berking, 2015; Cherniss, 2000; Pallich et al., 2020). This group, who showed higher levels of approach incongruence and avoidance incongruence, indicating dissatisfaction of motives, should be analyzed more deeply during treatment. Targeted interventions may be selected after recognizing individual sources of motivational incongruence to increase the level of patients` motivational satisfaction. Finally, especially for the group of Unhappy Altruists, who show the highest rate of comorbid somatic diagnoses, it is suggested to consider specific interventions and treatment for somatic problems.

Additionally, an assessment of personality disorders (Fydrich et al., 1997) seems indicated particularly for Straightforward Pragmatists and Unhappy Altruists, who generally show a high percentage of comorbidity with personality disorder diagnoses. During the inpatient stay and after discharge, a long-lasting psychotherapeutic treatment with a focus on personality disorder should be implemented (Sachse, 2013). Especially strategies for avoidant and obsessive-compulsive personality disorders could play an important role (Sachse, 2013).

The present findings should be considered in light of some methodological limitations. First, the sample was relatively small and recruited in only one clinic, and sample sizes of the groups resulting from cluster analysis differed considerably ($n = 17$ to $n = 37$). Second, because of the heuristic nature of this study, we omitted corrections of

the significance levels for multiple testing (e.g., Bonferroni). Additionally, the intervals between assessments were different across the sample as the length of inpatient stay varied. This may have affected symptom scores at discharge and follow-up. Furthermore, follow-up data were collected only three months after discharge and the patient sample consists of patients of an inpatient ward of one psychiatric hospital in one country and is therefore not representative for the population of burnout patients.

Future research should replicate the classification of former patients in different, larger, and more diverse samples. An additional follow-up later in time could be more informative regarding relapse and promote the development of tailored interventions for the different groups. Following up inpatients months or even years after discharge could provide further information about potential difficulties patients may encounter in the long run regarding the course of symptoms and especially relapse risk for different groups. Additionally, different and tailored intervention programs for acute treatment and maintenance care for the different burnout types should be developed and tested. Those programs should focus on the individual needs and the tailored therapeutic interventions of the different groups, as mentioned above. We assume that observing such a differentiated treatment approach will increase the probability of an effective and long-lasting successful treatment outcome. Further long-term data collection will allow evaluating the effects of more tailored programs on the basis of assessed burnout subtypes in service of further optimizing the acute treatment and aftercare of burnout patients. The development of tailored treatment programs for the different subtypes of burnout patients and their long-term evaluation will be an important next step to optimize the acute treatment and aftercare of burnout patients.

Conclusion

To the best of our knowledge, this is the first study to constructively replicate and improve the attempt to categorize burnout inpatients of Haberthür and colleagues (2009). Overall, we were able to replicate and improve the characterization of the four different groups: Functional, Dysfunctional, Straightforward Pragmatist, and Unhappy Altruist. Additionally, we described psychiatric, somatic, and personality disorder diagnoses. We further showed the symptoms course in the four groups of burnout patients. These findings support the proposition that burnout is a heterogeneous phenomenon. For clinicians it is necessary to consider these different characteristics of burnout inpatients in order to assure an individually tailored treatment program and corresponding discharge and aftercare planning. Future research should focus on tailored treatment programs depending on different subtypes of burnout patients.

PsychOpen GOLD

Supplementary Materials

In the Supplementary Materials we report mean, median, standard deviation and range for BDI, SCL and MBI for the four groups (i.e. Functional, Dysfunctional, Straightforward Pragmatist and Unhappy Altruist) (for access see Index of Supplementary Materials below).

Index of Supplementary Materials

Pallich, G., grosse Holtforth, M., & Hochstrasser, B. (2021). *Supplementary materials to "Burnout subtypes: Psychological characteristics, standardized diagnoses and symptoms course to identify aftercare needs"* [Additional results]. PsychOpen GOLD.
https://doi.org/10.23668/psycharchives.5093

References

Ahola, K., & Hakanen, J. (2007). Job strain, burnout, and depressive symptoms: A prospective study among dentists. *Journal of Affective Disorders, 104*(1-3), 103-110.
https://doi.org/10.1016/j.jad.2007.03.004

Ahola, K., Toppinen-Tanner, S., & Seppänen, J. (2017). Interventions to alleviate burnout symptoms and to support return to work among employees with burnout: Systematic review and meta-analysis. *Burnout Research, 4*, 1-11. https://doi.org/10.1016/j.burn.2017.02.001

Awa, W. L., Plaumann, M., & Walter, U. (2010). Burnout prevention: A review of intervention programs. *Patient Education and Counseling, 78*(2), 184-190.
https://doi.org/10.1016/j.pec.2009.04.008

Berger, M., Linden, M., Schramm, E., Hillert, A., Voderholzer, U., & Maier, W. (2012). *Positionspapier der deutschen Gesellschaft für Psychiatrie, Psychotherapie Und Nervenheilkunde (DGPPN) Zum Thema Burnout.* Berlin, Germany: DGPPN.

Berking, M. (2015). *Training emotionaler Kompetenzen* (3rd ed.). Springer.
https://doi.org/10.1007/978-3-642-54017-2

PsychOpen GOLD

Berking, M., & Znoj, H. (2008). Entwicklung und Validierung eines Fragebogens zur standardisierten Selbsteinschätzung emotionaler Kompetenzen (SEK-27). *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie, 56*(2), 141-153. https://doi.org/10.1024/1661-4747.56.2.141

Bianchi, R., Schonfeld, I. S., & Laurent, E. (2015). Burnout–depression overlap: A review. *Clinical Psychology Review, 36*, 28-41. https://doi.org/10.1016/j.cpr.2015.01.004

Burisch, M. (2014). *Das Burnout-Syndrom* (5th ed.). Springer. https://doi.org/10.1007/978-3-642-36255-2

Cherniss, C. (2000). Social and emotional competence in the workplace. In R. Bar-On & J. D. A. Parker (Eds.), *The handbook of emotional intelligence: Theory, development, assessment, and application at home, school, and in the workplace* (pp. 433–458). Jossey-Bass.

Elkuch, F. M., Haberthür, A. K., Hochstrasser, B., grosse Holtforth, M., & Soyka, M. (2010). Langzeiteffekte einer stationären Burnouttherapie – eine Nachbefragung. *Verhaltenstherapie & Verhaltensmedizin, 31*, 4-18.

Farber, B. A. (1991). *Crisis in education: Stress and burnout in the American teacher.* Jossey-Bass.

Freudenberger, H. J. (1974). Staff burn-out. *The Journal of Social Issues, 30*(1), 159-165. https://doi.org/10.1111/j.1540-4560.1974.tb00706.x

Fydrich, T., Renneberg, B., Schmitz, B., & Wittchen, H.-U. (1997). *SKID II. Strukturiertes Klinisches Interview für DSM-IV, Achse II: Persönlichkeitsstörungen. Interviewheft. Eine deutschsprachige, erw. Bearb. d. amerikanischen Originalversion d. SKID-II von: M. B. First, R. L. Spitzer, M. Gibbon, J. B. W. Williams, L. Benjamin, (Version 3/96).* Hogrefe.

Fydrich, T., Sommer, G., & Brähler, E. (2007). *Fragebogen zur Sozialen Unterstützung (F-SOZU)* [Questionnaire of social support]. Hogrefe.

grosse Holtforth, M. (2005). *Deutsche 32-Item Kurzform des Inventars zur Erfassung Interpersonaler Probleme (IIP-SC-D) (unveröffentlicher Fragebogen)* [German 32-item short version of the inventory for the evaluation of interpersonal problems (IIP-SC) (unpublished questionnaire)]. Universität Bern, Institut für Psychologie.

grosse Holtforth, M., & Grawe, K. (2003). Der Inkongruenzfragebogen (INK). *Zeitschrift für Klinische Psychologie und Psychotherapie, 32*(4), 315-323. https://doi.org/10.1026/0084-5345.32.4.315

Haberthür, A. K., Elkuch, F. M., Holtforth, M. G., Hochstrasser, B., & Soyka, M. (2009). Characterization of patients discharged from inpatient treatment for burnout: Use of psychological characteristics to identify aftercare needs. *Journal of Clinical Psychology, 65*(10), 1039-1055. https://doi.org/10.1002/jclp.20606

Hautzinger, M., Bailer, M., Worall, H., & Keller, F. (1995). *BDI Beck-Depressions-Inventar.* Huber.

Hochstrasser, B., Von Bardeleben, U., Ruckstuhl, L., & Soyka, M. (2008). Therapie des Burnout – Theoretischer Hintergrund, Klinik und Darstellung eines stationären multimodalen Behandlungskonzeptes. *Nervenheilkunde: Zeitschrift für Interdisziplinäre Fortbildung, 27*(1), 11-24. https://doi.org/10.1055/s-0038-1627102

Horowitz, L. M., Strauss, B., & Kordy, H. (2000). *Inventar zur Erfassung Interpersonaler Probleme. Deutsche Version* [Inventory for the Evaluation of Interpersonal Problems. German version, 2nd, revised ed. and with new norms] (2., überarbeitete und neunormierte Aufl.). Beltz Test GmBH.

Kälin, W. (1995). *Deutsche 24-Item Kurzform des 'Coping Inventory for Stressful Situations' (CISS) von Endler, N. S., & Parker, J. D. A. Basierend auf der Übersetzung von Semmer, N., Tschan, F., & Schade, V. (unveröffentlichter Fragebogen)* [German 24-item short form of the Coping Inventory for Stressful Situations (CISS) by Endler, N. S., & Parker, J. D. A. Based on the translation by Semmer, N., Tschan, F., & Schade, V. (unpublished questionnaire)]. Universität Bern, Institut für Psychologie.

Kaschka, W. P., Korczak, D., & Broich, K. (2011). Modediagnose Burn-out. *Deutsches Ärzteblatt, 108*(46), 781-787.

Klaghofer, R., & Brähler, E. (2001). Konstruktion und Teststatistische Prüfung einer Kurzform der SCL-90–R. [Construction and test statistical evaluation of a short version of the SCL-90–R.]. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie, 49*(2), 115-124.

Leiter, M. P., & Maslach, C. (2016). Latent burnout profiles: A new approach to understanding the burnout experience. *Burnout Research, 3*(4), 89-100. https://doi.org/10.1016/j.burn.2016.09.001

Margraf, J. (2013). *Mini-DIPS: Diagnostisches Kurz-Interview bei psychischen Störungen*. Springer.

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Organizational Behavior, 2*(2), 99-113. https://doi.org/10.1002/job.4030020205

Maslach, C., Jackson, S., & Leiter, M. (1997). The Maslach Burnout Inventory Manual. In *Evaluating stress: A book of resources* (Vol. 3, pp. 191–218). Scarecrow Press.

Montero-Marín, J., García-Campayo, J., Mera, D. M., & del Hoyo, Y. L. (2009). A new definition of burnout syndrome based on Farber's proposal. *Journal of Occupational Medicine and Toxicology, 4*, Article 31. https://doi.org/10.1186/1745-6673-4-31

Nil, R., Jacobshagen, N., Schächinger, H., Baumann, P., Höcke, P., Hättenschwiler, J., Ramseier, F., Seifritz, E., & Holsboer-Trachsler, E. (2010). Burnout—Eine Standortbestimmung. *Schweizer Archiv für Neurologie und Psychiatrie, 161*(2), 72-77. https://doi.org/10.4414/sanp.2010.02140

Pallich, G., Blättler, L., Gomez Penedo, J. M., grosse Holtforth, M., & Hochstrasser, B. (2020). Emotional competence predicts outcome of an inpatient treatment program for burnout. *Journal of Affective Disorders, 274*, 949-954. https://doi.org/10.1016/j.jad.2020.05.139

Perry, B. L., & Pescosolido, B. A. (2015). Social network activation: The role of health discussion partners in recovery from mental illness. *Social Science & Medicine, 125*, 116-128. https://doi.org/10.1016/j.socscimed.2013.12.033

Perski, O., Grossi, G., Perski, A., & Niemi, M. (2017). A systematic review and meta-analysis of tertiary interventions in clinical burnout. *Scandinavian Journal of Psychology, 58*(6), 551-561. https://doi.org/10.1111/sjop.12398

Pescosolido, B. A., & Levy, J. A. (2002). The role of social networks in health, illness, disease and healing: The accepting present, the forgotten past, and the dangerous potential for a complacent future. In *Social networks and health* (Vol. 8, pp. 3-25). https://doi.org/10.1016/S1057-6290(02)80019-5

PsychOpen GOLD

Rakos, R. F. (1991). *Assertive behavior: Theory, research, and training*. Taylor & Francis/Routledge.

Sachse, R. (2013). *Persönlichkeitsstörungen: Leitfaden für die Psychologische Psychotherapie* (2., überarbeitete und erweiterte Auflage 2013). Hogrefe.

Schaarschmidt, U., & Fischer, A. (1996). *AVEM: arbeitsbezogene Verhaltens-und Erlebnismuster*. Swets Test Services.

Schaarschmidt, U., & Fischer, A. (2001). *Coping with professional demands: A new diagnostic approach*. na.

Schulz, M., Damkröger, A., Voltmer, E., Löwe, B., Driessen, M., Ward, M., & Wingenfeld, K. (2011). Work-related behaviour and experience pattern in nurses: Impact on physical and mental health. *Journal of Psychiatric and Mental Health Nursing, 18*(5), 411-417. https://doi.org/10.1111/j.1365-2850.2011.01691.x

Schwarzkopf, K., Conrad, N., Straus, D., Porschke, H., & Von, R. K. (2016). Einmal Burnout ist nicht immer Burnout: Eine stationäre multimodale Psychotherapie ist eine effektive Burnout-Behandlung [Effectiveness of an inpatient multimodal psychiatric-psychotherapeutic program for the treatment of job burnout]. *Praxis, 105*(6), 315-321. https://doi.org/10.1024/1661-8157/a002301

Smith, K. P., & Christakis, N. A. (2008). Social networks and health. *Annual Review of Sociology, 34*(1), 405-429. https://doi.org/10.1146/annurev.soc.34.040507.134601

Soldz, S., Budman, S., Demby, A., & Merry, J. (1995). A Short Form of the Inventory of Interpersonal Problems Circumples Scales. *Assessment, 2*(1), 53-63. https://doi.org/10.1177/1073191195002001006

van der Klink, J. J., Blonk, R. W. B., Schene, A. H., & van Dijk, F. J. H. (2001). The benefits of interventions for work-related stress. *American Journal of Public Health, 91*(2), 270-276. https://doi.org/10.2105/AJPH.91.2.270

Voltmer, E., Kieschke, U., & Spahn, C. (2007). Work-related behaviour and experience patterns of physicians compared to other professions. *Swiss Medical Weekly, 137*(31–32), 448-453. https://doi.org/10.4414/smw.2007.11834

Voltmer, E., Spahn, C., Schaarschmidt, U., & Kieschke, U. (2011). Work-related behavior and experience patterns of entrepreneurs compared to teachers and physicians. *International Archives of Occupational and Environmental Health, 84*(5), 479-490. https://doi.org/10.1007/s00420-011-0632-9

Voltmer, E., Schwappach, D. L. B., Frank, E., Wirsching, M., & Spahn, C. (2010). Work-related behavior and experience patterns and predictors of mental health in German physicians in medical practice. *Family Medicine, 42*(6), 433-439.

Zimmermann, J., Benecke, C., Hörz, S., Rentrop, M., Peham, D., Bock, A., Wallner, T., Schauenburg, H., Frommer, J., & Huber, D. (2013). Validierung einer deutschsprachigen 16-item-version des inventars der persönlichkeitsorganisation (IPO-16). *Diagnostica, 59*(1), 3-16. https://doi.org/10.1026/0012-1924/a000076

PsychOpen GOLD

EACLIPT

Clinical Psychology in Europe (CPE) is the official journal of the European Association of Clinical Psychology and Psychological Treatment (EACLIPT).

leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.

PsychOpen GOLD

Imagery Rescripting Versus Cognitive Restructuring for Social Anxiety: Treatment Effects and Working Mechanisms

Miriam Strohm[1], Marena Siegesleitner[1], Anna E. Kunze[1] , Thomas Ehring[1] ,

Charlotte E. Wittekind[1]

[1] *Department of Psychology, LMU Munich, Munich, Germany.*

Corresponding Author: Charlotte E. Wittekind, Department of Psychology, LMU Munich, Leopoldstraße 13, 80802 Munich, Germany. Tel.: +49 (0)89 2180 5196. E-mail: charlotte.wittekind@psy.lmu.de

Supplementary Materials: Materials [see Index of Supplementary Materials]

Abstract

Background: Negative mental images in social anxiety are often linked to memories of distressing social experiences. Imagery Rescripting (ImRs) has been found to be a promising intervention to target aversive memories, but mechanisms underlying ImRs are largely unknown. The present study aimed (a) to investigate the effects of ImRs compared to cognitive restructuring (CR) on social anxiety symptoms and (b) to extend previous research by examining whether ImRs works by fostering reappraisal of negative emotional self-beliefs.

Method: Highly socially anxious individuals (N = 77) were randomly allocated to ImRs, CR, or no intervention control (NIC). A speech task was performed at baseline and at 1-week follow-up.

Results: Only CR significantly reduced social anxiety symptoms from baseline to follow-up. Decreases in negative appraisals and emotional distress in response to the speech task did not differ between conditions. Regarding working mechanisms, ImRs led to stronger increases in positive emotions than CR and NIC. Both CR and ImRs yielded short-term reductions in emotionally anchored idiosyncratic self-beliefs, but CR was superior to ImRs at follow-up.

Conclusions: The present study provides evidence for the efficacy of a single-session of CR for social anxiety symptoms. As one specific version of ImRs was applied, it is conceivable that other or optimized versions of ImRs might be more effective.

Keywords

> **Highlights**
> - CR was more effective than ImRs and no intervention to reduce social anxiety symptoms.
> - CR more effectively reduced dysfunctional beliefs.
> - ImRs led to strongest increase of positive emotions.

Cognitive models of social anxiety disorder (SAD) suggest that negative mental images of the self are a key maintaining factor of the disorder (Clark & Wells, 1995; Hofmann, 2007; Rapee & Heimberg, 1997). Image content is often linked to former aversive social experiences (Hackmann et al., 2000). Therefore, specifically targeting these aversive memories during treatment might improve therapeutic outcomes (Norton & Abbott, 2017; Wild & Clark, 2011).

Imagery Rescripting (ImRs) is an imagery-based intervention for aversive memories that has increasingly been incorporated in cognitive behavioral therapy (CBT) for SAD (e.g., McEvoy et al., 2020; McEvoy & Saulsman, 2014; Wild & Clark, 2011). During ImRs, patients are instructed to visualize an aversive memory and to change it in imagination according to their emotional needs. ImRs aims to update the meaning of memories thereby reducing associated negative (self-)images, beliefs, and emotions (Arntz, 2012). ImRs may be an efficacious treatment for different disorders including SAD (Morina et al., 2017). Several studies have found that one session of ImRs significantly improved social anxiety symptoms (Lee & Kwon, 2013; Wild et al., 2007, 2008), also when delivered as a stand-alone intervention and without prior cognitive restructuring (CR; Nilsson et al., 2012; Norton & Abbott, 2016; Reimer & Moscovitch, 2015). While ImRs yields promising treatment results, a better understanding of its underlying working mechanisms is needed to eventually optimize treatment efficacy.

It has been proposed that ImRs might work by changing the idiosyncratic meaning of aversive experiences (Arntz, 2012) and, more specifically, by leading to *emotionally anchored reappraisal* of core beliefs (Nilsson et al., 2012; Norton & Abbott, 2016; Wild et al., 2008). During ImRs, positive meanings are offered in the form of images. Based on evidence that mental imagery elicits stronger emotions than verbal thinking (Holmes & Mathews, 2010), it is conceivable that generating images with alternative meanings during ImRs is associated with stronger emotional activation than questioning maladaptive beliefs verbally (Holmes et al., 2009). Consequently, alternative meanings offered in the form of images might be more emotionally anchored, more believable, and more likely to lead to changes in behavior than meanings exclusively generated as verbal representations (Holmes & Mathews, 2010). This assumption is in line with the idea that

one can distinguish between different levels of meaning representations (e.g., Barnard & Teasdale, 1991; but see Power & Dalgleish, 1999). According to the Model of Interacting Cognitive Subsystems (ICS; Barnard & Teasdale, 1991), intellectual beliefs (propositional level) can be distinguished from emotional beliefs (implicational level). Intellectual beliefs are described as knowing something "with the head", whereas emotional beliefs correspond to an implicit sense of knowing "with the heart" or "having a gut feeling" (Barnard & Teasdale, 1991). Cognitive treatments can be expected to change beliefs primarily on a propositional level. ImRs as an experientially oriented intervention invokes different sensory modalities thereby addressing the implicational meaning level, which is suggested to be necessary to then change emotional beliefs (see Arntz, 2012; Wild et al., 2008). Although *emotionally anchored reappraisal* (i.e., changing emotional beliefs) has often been discussed as a mechanism underlying ImRs, empirical evidence is largely missing. One study with a sample of Bulimia Nervosa patients has investigated effects of ImRs on emotional vs. intellectual beliefs (Cooper et al., 2007). ImRs was found to be more effective than a control intervention in reducing emotional self-beliefs. A recent study investigated the effects of ImRs (vs. imaginal exposure [IE] and supportive counselling [SC]) on memory processes in patients with social anxiety disorder (Romano et al., 2020). There were no differences between conditions regarding memory appraisal, but a higher proportion of patients receiving ImRs updated their negative core belief compared to SC (no differences emerged compared to IE). Given the limited number of studies on working mechanisms of ImRs, the aim of the present study was to investigate whether ImRs works by reducing maladaptive emotional beliefs.

The present study aimed to (1) investigate the effects of stand-alone ImRs and CR on social anxiety symptoms, and (2) extend previous research by exploring mechanisms underlying ImRs. Our procedure was based on the study by Norton and Abbott (2016). Highly socially anxious individuals were randomly allocated to either one session of ImRs, one session of CR, or a no-intervention control condition (NIC). Outcomes were assessed at baseline and at 1-week follow-up. A speech task was included to examine intervention effects to a social stressor. In line with previous findings, we hypothesized that ImRs and CR would yield greater decreases in social anxiety symptoms than NIC. We expected ImRs and CR to reduce negative appraisals and emotional responses (subjective arousal and distress) to the speech task more strongly than NIC. Regarding mechanisms, we hypothesized that ImRs would lead to stronger emotional activation than CR. While we expected both ImRs and CR to decrease the maladaptive intellectual self-beliefs, we assumed that ImRs would yield stronger reductions of maladaptive emotional self-beliefs. We additionally explored the relationship between the hypothesized mechanisms and symptomatic change.

Method

Participants

Highly socially anxious individuals were recruited via advertisements on university campus and social media. To be included, participants had to score ≥ 30 (clinical cut-off) on the German version of the Social Interaction Anxiety Scale (SIAS; Stangier et al., 1999). Results of a sample-size calculation (two-tailed, α = .05, power = .80, run with G*Power 3.1; Faul et al., 2007) with medium to large effect sizes (*d* = .70; Morina et al., 2017) showed that a sample size of 76 was required to detect significant differences between active treatments (ImRs + CR) versus NIC.

During the first session, eligible participants were administered the Mini International Neuropsychiatric Interview (M.I.N.I. 5.0.0; Sheehan et al., 1998; German version: Ackenheil et al., 1999) to screen for exclusion criteria: (1) current diagnosis of Major Depressive Disorder, (2) current and/or lifetime diagnosis of Posttraumatic Stress Disorder/Psychotic Disorder/Bipolar Disorder, (3) Substance Dependence during the past 12 months, (4) acute suicidal tendencies. Further exclusion criteria were: (5) age < 18 or > 35 years, (6) current psychological treatment, (7) pregnancy, (8) severe physical illness. The restricted age range was applied to obtain a more homogenous sample regarding age. Participants had to meet the following inclusion criteria: (1) negative mental self-image(s) in feared social situations, (2) aversive social experience related to the image, and (3) maladaptive self-belief (see Section "Imagery Interview").

A total of 96 participants attended Session 1 of whom 16 had to be excluded (*n* = 10 current/lifetime diagnosis of mental disorders specified above; *n* = 4 no negative mental self-image; *n* = 2 no maladaptive self-belief). Three participants did not attend the follow-up session, leaving a final sample of 77 participants (81% female; age: *M* = 22.46, *SD* = 3.88). All participants gave written informed consent and were reimbursed by receiving partial course credit or 20€. The study was approved by the Research Ethics Committee of the Faculty of Psychology and Educational Sciences at LMU Munich.

Clinical Interviews

The M.I.N.I. (Sheehan et al., 1998; German version: Ackenheil et al., 1999) was administered to assess current diagnoses according to DSM-IV (American Psychiatric Association [APA], 2000). Additionally, the SAD module of the Structured Clinical Interview for DSM-IV (SCID-I; First et al., 2002; German version: Wittchen et al., 1997) was administered.

Imagery Interview

The Imagery Interview was based on the Waterloo Images and Memories Interview (WIMI; Moscovitch et al., 2011) and on the interview used by Norton and Abbott (2016).

The semi-structured interview assessed negative self-imagery, aversive memories, and maladaptive self-beliefs. Participants were asked to define their most anxiety-provoking social situation and to imagine themselves being in such a situation. They were instructed to become aware of whether there was a mental image that comes to their mind in this kind of situation and to describe the mental image in detail. Participants were then asked when they first felt the way they did in the image and to visualize and describe the respective event. This was used to determine whether there was an early aversive memory related to the mental image. In order to specify the idiosyncratic self-belief derived from the negative mental image and the aversive memory, participants were asked: "What do the image and the memory tell about you as a person?". Participants were instructed to summarize the meaning in form of a short statement.

Speech Task

In order to measure reactions to a social stressor, participants were asked to give a 3 min video-recorded impromptu speech (Norton & Abbott, 2016) on a given political topic in both sessions (the order of two topics was counterbalanced).

Symptom Measures

The 20-item SIAS (Mattick & Clarke, 1998; German version: Stangier et al., 1999) was used to assess social interaction anxiety during the past seven days on a 5-point scale (0 = *not at all* to 4 = *extremely*). The 12-item Brief Fear of Negative Evaluation Scale-Revised (BFNE-R; Carleton et al., 2006; German version: Reichenberger et al., 2016) was administered to measure fear of negative evaluation by others on a 5-point scale (1 = *not at all characteristic of me* to 5 = *extremely characteristic of me*). In order to test for baseline group differences in depressive symptoms, the Patient Health Questionnaire-9 Item (PHQ-9; Krönke et al., 2001; German version: Löwe et al., 2002) was administered.

Speech Task Measures

In order to verify the relevance of the speech task as a stressor we asked participants to indicate how anxious they had felt or would have felt when giving a speech/presentation during the last week (0 = *not at all anxious* to 3 = *extremely anxious*). The Probability and Consequences Questionnaire (PCQ; Rapee & Abbott, 2007) asks participants to rate their appraisal of the likelihood (7 items) and cost (7 items) of negative evaluation of their speech on a 5-point scale (0 = *not at all likely/bad* to 4 = *extremely likely/bad*). Subjectively experienced levels of distress were assessed using Subjective Units of Distress (SUD, 0 = *not at all distressed* to 100 = *extremely distressed*). Self-assessment manikins (SAM; Bradley & Lang, 1994) were used to assess self-reported physiological arousal (1 = *very calm* to 9 = *very aroused*).

Measures of Underlying Mechanisms

Emotional Activation

The Positive and Negative Affect Schedule-Extended (PANAS-X; Watson & Clark, 1994; German version: Grühn et al., 2010) was administered to assess changes in positive and negative emotions from pre- to post-intervention. Participants were instructed to indicate how they felt at this very moment. We included the general dimensions "positive affect" (PA) and "negative affect" (NA) as well as the subscales "fear", "hostility", "guilt", "sadness", "joviality", "self-assurance", and "attentiveness". Scales range from 1 (*very slightly or not at all*) to 5 (*extremely*).

Intellectual and Emotional Beliefs

The maladaptive self-belief was identified during the Imagery Interview. Participants were asked to rate intellectually and emotionally how much they felt that this belief was true (see Cooper et al., 2007). For the intellectual rating, participants were asked to indicate how much they would rationally agree to their belief (0 = *I do not agree at all* to 100 = *I completely agree).* For the emotional rating, participants were asked how much they *felt* the belief was true, regardless of what they were thinking rationally (0 = *feels not true at all* to 100 = *feels completely true).*

Interventions

Imagery Rescripting

The ImRs procedure was based on protocols by Arntz and Weertman (1999) and Wild and Clark (2011). Stage 1 of ImRs started with participants closing their eyes and vividly imagining the aversive memory from the perspective of their younger-self. Participants were instructed to describe the situation in the first person, present tense, and to include all sensory modalities. Stage 2 of ImRs was initiated by instructing participants to imagine the scene from the perspective of their current adult-self who is witnessing the events as a bystander. Participants were asked to describe what they see is happening to their younger-self and were then encouraged to intervene in any way they wished. When the adult-self felt fully satisfied, Stage 3 was initiated by asking participants to relive the memory again from the perspective of their younger-self, experiencing the interventions of their adult-self. Additionally, the younger-self was encouraged to express further unmet needs. The ImRs procedure was concluded by asking participants to dwell on the final positive image. As we wanted to elucidate the underlying mechanisms of ImRs (vs. CR) on symptom change, we used "pure" interventions and tested ImRs in isolation. Consequently, ImRs was not preceded by cognitive restructuring and we did not explicitly refer to the maladaptive self-belief during ImRs. The mean duration of ImRs was 22.35 min (*SD* = 6.20).

Cognitive Restructuring

The CR procedure was based on the protocol by Wild and Clark (2011). Participants were first asked to outline evidence for their maladaptive self-belief and were then encouraged to challenge the self-belief by collecting evidence against it. To support this process we asked participants to consider alternative explanations for their experiences (including the early aversive memory), and to think of experiences contradicting the self-belief. All evidence for and against the negative self-belief was written down on a worksheet. Finally, participants were instructed to rephrase the original self-belief into a more helpful statement. The mean duration of CR was 23.74 min (SD = 4.40).

No-Intervention Control Condition

Participants in NIC were provided neutral magazines and were instructed to wait for 30 min in the laboratory. They were asked not to use any electronic device.

Procedure

The study comprised two sessions, which were one week apart. Two experimenters carried out different parts of the procedure so that the speech task and intervention were not administered by the same experimenter. During Session 1, Experimenter 1 administered the clinical interviews and baseline measurements (t0: sociodemographic data, SIAS, BFNE-R, public speaking anxiety, SUIS, ERQ), followed by pre-speech measures (SUD, SAM, PCQ) and the speech task. Experimenter 2 then conducted the Imagery Interview and administered pre-treatment questionnaires (t1: intellectual and emotional belief, PANAS-X). Then, participants were randomly allocated to ImRs (n = 25), CR (n = 27), or NIC (n = 25). The allocation sequence was computer-generated and Experimenter 2 was blinded until the beginning of the interventions, Experimenter 1 was blinded during the entire study. Immediately after the interventions or the waiting period, participants completed post-treatment measures (t2: intellectual and emotional belief, PANAS-X). During Session 2, which took place one week later, Experimenter 1 administered the follow-up questionnaire (t3: SIAS, BFNE-R, intellectual and emotional belief) and the second speech task, again including speech task measures administered prior to the speech task (SUD, SAM, PCQ). Finally, participants were fully debriefed.

Statistical Analyses

A series of 2(Time) x 3(Condition) repeated measures ANOVAs were carried out for social anxiety symptoms (t0; t3), for speech task measures (pre-speech1; pre-speech2), and for positive and negative emotions (t1; t2). To follow up significant interactions, planned contrasts on change scores were conducted (ImRs+CR vs. NIC; ImRs vs. CR). Effects on intellectual and emotional self-beliefs were tested with 3(Time) x 3(Condition) repeated measures ANOVAs. Significant interactions were followed up using planned contrasts

(ImRs+CR vs. NIC; ImRs vs. CR). For ImRs, Pearson correlations were computed between mechanisms and symptomatic change. A significance level of α = .05 (two-tailed) was used for all analyses. Partial eta squared (η_p^2) or Cohen's d were used as effect sizes.

Results

Participant Characteristics and Baseline Comparisons

No significant baseline differences between conditions emerged (see Table 1). Mean age at time of the aversive event was 12.86 years (SD = 4.55; range 3-27), with significant differences between groups[1] (ImRs: M = 13.88, SD = 4.90; CR: M = 13.76, SD = 4.60; NIC: M = 10.88, SD = 3.55), $F(2, 74)$ = 3.78, p = .027.

Table 1

Demographic Variables and Pre-Treatment Characteristics

Demographics and pre-treatment characteristics	Overall sample (N = 77)	ImRs (n = 25)	CR (n = 27)	NIC (n = 25)	Statistics
Demographics					
Gender (female/male), n	62/15	21/4	20/7	21/4	$\chi^2(2)$ = 1.10, p = .577
Age in years, M (SD)	22.36 (3.88)	22.64 (3.82)	22.59 (3.92)	21.84 (4.01)	$F(2,74)$ = 0.33, p = .718
Social anxiety symptoms, M (SD)					
SIAS	40.29 (12.55)	40.84 (13.21)	37.93 (12.06)	42.28 (12.49)	$F(2,74)$ = 0.81, p = .447
BFNE-R	40.48 (10.39)	40.20 (11.00)	39.44 (10.36)	41.88 (10.07)	$F(2,74)$ = 0.36, p = .696
SAD Criteria met, n (%)	21 (27)	8 (32)	8 (30)	5 (20)	$\chi^2(2)$ = 1.02, p = .599
Comorbidity (yes/no), n	7/70	3/22	3/24	1/24	
Generalized Anxiety Disorder, n	2	0	1	1	
Dysthymia	3	1	2	0	
Anorexia Nervosa	1	1	0	0	
Bulimia Nervosa	1	1	0	0	
Public Speaking Anxiety, M (SD)	1.94 (0.85)	1.92 (0.95)	1.93 (0.96)	1.96 (0.61)	$F(2,74)$ = 0.02, p = .984

Note. ImRs = Imagery Rescripting; CR = Cognitive Restructuring; NIC = No-Intervention Control; SIAS = Social Interaction Anxiety Scale; BFNE-R = Brief Fear of Negative Evaluation Scale-Revised; SAD = Social Anxiety Disorder.

---

1) We tested whether age of the aversive memory (i.e., time that had passed since the event) had an influence on our main symptomatic outcomes. However, results remained unchanged when including age of the memory as a covariate. Note that age of the aversive memory was not significantly different in the two active treatment conditions (ImRs and CR).

Social Anxiety Symptoms

Social Interaction Anxiety

For SIAS scores (see Figure 1), there was no main effect of Condition, $F(2, 74) = 1.97$, $p = .147$, $\eta_p^2 = .05$, but a significant effect of Time, $F(1, 74) = 17.94$, $p < .001$, $\eta_p^2 = .20$, and a significant interaction, $F(2, 74) = 3.22$, $p = .046$, $\eta_p^2 = .08$. Planned contrasts revealed no difference between the active treatment groups compared to NIC in reducing social interaction anxiety, $t(74) = 1.05$, $p = .298$, $d = 0.26$. However, CR led to stronger decreases than ImRs, $t(74) = 2.29$, $p = .025$, $d = 0.64$.

Figure 1

Effects of ImRs vs. CR vs. NIC on (a) Social Interaction Anxiety (SIAS), and (b) Fear of Negative Evaluation (BFNE-R)
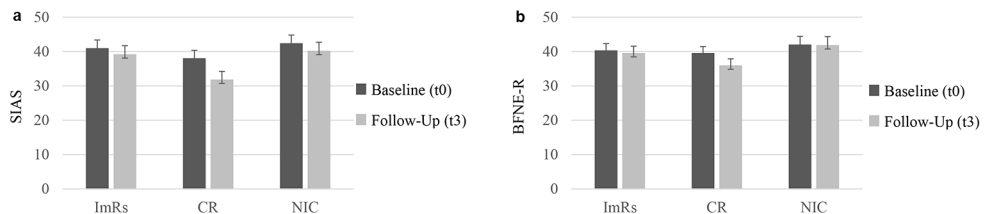
Note. Error Bars Represent SEM.

Fear of Negative Evaluation

Results for BFNE-R revealed a significant main effect of Time, $F(1, 74) = 5.70$, $p = .020$, $\eta_p^2 = .07$, but neither a significant effect of Condition, $F(2, 74) = 1.09$, $p = .342$, $\eta_p^2 = .03$, nor a significant interaction, $F(2, 74) = 2.90$, $p = .061$, $\eta_p^2 = .07$., see Figure 1.

Speech Task Measures

For both subscales of the PCQ[2], there were significant main effects of Time, Fs$(1, 71) > 9.74$, ps $< .003$, η_p^2s $\geq .12$, but no significant interactions, Fs$(2, 71) < 2.28$, ps $> .110$, η_p^2s $\leq .06$. The main effect of Condition was significant for probability, $F(2, 71) = 3.13$, $p = .050$, $\eta_p^2 = .08$, but not for cost of negative evaluation, $F(2, 71) = 1.13$, $p = .330$, $\eta_p^2 = .03$. ImRs and CR did not yield significantly greater reductions of appraisals of negative evaluation than NIC (see Table 2).

For distress (SUD), a significant effect of Time emerged, $F(1, 70) = 17.41$, $p < .001$, $\eta_p^2 = .20$, but neither the main effect of Condition nor the interaction were significant, Fs$(2, 70) < 2.12$, ps $> .128$, η_p^2s $< .06$ (see Table 2).

---

2) In some participants, speech-related questionnaires were erroneously not administered (PCQ: $n = 3$; SUD: $n = 4$; SAM: $n = 2$) and these participants were excluded from the respective analyses.

Results for arousal (SAM) revealed a significant effect of Time, $F(1, 72) = 11.35$, $p = .001$, $\eta_p^2 = .14$, but neither a significant main effect of Condition nor a significant interaction, $Fs(2, 72) < 1.05$, $p > .354$, $\eta_p^2 < .03$ (see Table 2).

Table 2

Means and Standard Deviations for Speech Task Measures Before (Speech 1) and After (Speech 2) Intervention: Means (SD)

	Speech 1	Speech 2
Group	*M (SD)*	*M (SD)*
Negative Evaluation: Probability[a]		
ImRs	15.46 (4.25)	14.79 (4.66)
CR	13.04 (5.62)	10.27 (5.45)
NIC	14.25 (5.93)	13.33 (5.93)
Negative Evaluation: Cost[a]		
ImRs	13.50 (5.87)	12.25 (6.10)
CR	12.46 (6.71)	9.27 (4.64)
NIC	13.79 (5.98)	12.29 (6.52)
Distress (SUD)[b]		
ImRs	66.50 (29.77)	57.42 (27.33)
CR	75.12 (22.01)	55.35 (28.42)
NIC	72.17 (23.10)	65.65 (24.33)
Arousal (SAM)[c]		
ImRs	6.67 (1.61)	5.79 (1.35)
CR	6.62 (1.50)	5.65 (1.67)
NIC	6.28 (1.67)	6.00 (1.61)

Note. ImRs = Imagery Rescripting; CR = Cognitive Restructuring; NIC = No-Intervention Control; SUD = Subjective Units of Distress; SAM = Self-Assessment Manikins.
[a]$n = 74$. [b]$n = 73$. [c]$n = 75$.

Mechanisms

Activation of Positive and Negative Emotions

For PANAS-PA and NA (see Table 3) there were significant effects of Time, $Fs(1, 74) > 35.10$, $ps < .001$, $\eta_p^2 s \geq .32$, but no significant effects of Condition, $Fs(2, 74) < 2.17$, $ps > .121$, $\eta_p^2 s \leq .06$. No significant interaction was found for PANAS-NA, $F(2, 74) = 0.57$, $p = .570$, $\eta_p^2 = .02$. A significant interaction emerged for PANAS-PA, $F(2, 74) = 9.29$, $p < .001$, $\eta_p^2 = .20$. Planned contrasts revealed that active treatments increased positive emotions more strongly than NIC ($M_{diff} = -0.52$, $SD = 4.48$), $t(60.89) = 3.97$, $p < .001$, $d = 0.97$, with ImRs ($M_{diff} = -7.36$, $SD = 6.81$) leading to stronger increases than CR ($M_{diff} = $

-3.52, SD = 5.35), $t(45.54)$ = 2.25, p = .029, d = 0.62. Results for the remaining subscales of PANAS-X are provided in the Supplementary Materials (Table S1).

Table 3

Symptom Measures and Mechanism Variables Before the Interventions (t0/t1), After the interventions (t2) and at Follow-up (t3): Means (SD)

	t0/t1	t2	t3
Group	*M (SD)*	*M (SD)*	*M (SD)*
PANAS-PA			
ImRs	23.12 (5.20)	30.48 (8.21)	
CR	22.89 (6.79)	26.41 (7.12)	
NIC	22.92 (6.13)	23.44 (6.89)	
PANAS-NA			
ImRs	19.04 (6.77)	13.92 (3.64)	
CR	18.19 (6.29)	14.41 (5.37)	
NIC	18.60 (5.58)	13.48 (3.12)	
Intellectual belief			
ImRs	51.60 (27.53)	39.40 (26.91)	48.80 (26.55)
CR	64.74 (29.83)	40.37 (25.79)	42.52 (29.49)
NIC	57.8 (33.32)	55.48 (32.32)	57.24 (29.63)
Emotional belief			
ImRs	90.40 (10.88)	62.52 (19.71)	73.80 (18.10)
CR	84.07 (16.82)	56.11 (29.00)	52.78 (27.92)
NIC	83.08 (20.92)	81.36 (21.90)	79.60 (20.74)

Note. ImRs = Imagery Rescripting; CR = Cognitive Restructuring; NIC = No-Intervention Control; SIAS = Social Interaction Anxiety Scale; BFNE-R = Brief Fear of Negative Evaluation Scale-Revised; PANAS = Positive and Negative Affect Schedule; PA = positive affect; NA = negative affect.

Intellectual and Emotional Beliefs

To check whether participants were able to distinguish between the intellectual and the emotional belief, a correlation between the two measures was computed. The moderate correlation of r_s = .387, p = .001, suggests that the two measures have some overlap but are not identical. For intellectual beliefs, there was no significant effect of Condition, $F(2, 74)$ = 1.00, p = .373, η_p^2 = .03, but a significant effect of Time and a significant interaction, $Fs(1.81, 134.19 / 3.63, 134.19)$ > 6.12, ps < .001, $\eta_p^2 s$ ≥ .14 (see Table 3). Planned contrasts revealed that active treatments led to stronger reductions in intellectual beliefs from pre- to post-intervention than NIC, $t(55.43)$ = 4.58, p < .001, d = 1.12, and from pre to follow-up, $t(74)$ = 2.13, p = .036, d = 0.52. CR led to stronger reductions than ImRs from pre- to post-intervention, $t(35.93)$ = 2.03, p = .050, d = 0.49, and from pre to follow-up, $t(74)$ = 3.04, p = .003, d = 0.84.

For emotional beliefs, there were significant effects of Time and Condition, Fs(2, 148/2, 74) > 5.37, ps ≤ .006, η_p^2s ≥ .13, and a significant interaction, F(4, 148) = 13.94, p < .001, η_p^2 = .27. Planned contrasts revealed that the active treatments reduced emotional beliefs more strongly than NIC from pre- to post-intervention, t(60.66) = 8.51, p < .001, d = 2.07, and from pre to follow-up, t(69.14) = 5.62, p < .001, d = 1.37. CR and ImRs decreased emotional beliefs from pre- to post-intervention equally effective, t(49.78) = -0.16, p = .878, d = 0.04, but CR led to stronger reductions than ImRs from pre to follow-up, t(48.13) = 2.67, p = .010, d = 0.74.

Correlations Between Mechanisms and Symptomatic Change

Within the ImRs group, symptomatic change was not significantly correlated with changes in emotions (PA x SIAS: r = -.08; PA x BFNE-R: r = .26; NA x SIAS: r = -.35; NA x BFNE-R: r = .11; ps ≥ .085) nor with pre-post changes in emotional beliefs and symptomatic change (SIAS: r = -.39; BFNE-R: r = -.15; all ps ≥ .055). The same non-significant pattern emerged in the CR group (PA x SIAS: r = -.25; PA x BFNE: r = .07; NA x SIAS: r = -.13; NA x BFNE: r = -.12; rational belief x SIAS: r = .14; rational belief x BFNE: r = .09, ps ≥ .217).

Discussion

The present study examined the effects of single-session ImRs vs. CR for socially anxious individuals compared to NIC.

Effects on Social Anxiety Symptoms

Contrary to hypothesis, we found that one session of cognitive restructuring (CR) is more effective than one session of imagery rescripting (ImRs) and no intervention control (NIC) in reducing social interaction anxiety. No significant differences between groups emerged for fear of negative evaluation. When confronted with the speech task, participants in all conditions demonstrated equal reductions in distress, arousal, and negative appraisals suggesting that if CR and ImRs are administered as very brief interventions no beneficial effects emerge over and above mere exposure to the speech. The speech task represents a strength of the study, but our findings suggest that the speech task may be susceptible to exposure effects, thereby reducing its ability to capture between-group differences in anxiety across time. Taken together, we could not replicate previous findings regarding the effects of the interventions on responses to a social stressor (Norton & Abbott, 2016). Our findings support previous evidence that one session of CR exerts positive effects on social anxiety symptoms (e.g., Norton & Abbott, 2016; Shikatani et al., 2014). Contrary to expectations, we were not able to replicate earlier findings on the benefits of stand-alone ImRs (Nilsson et al., 2012; Norton & Abbott, 2016;

Reimer & Moscovitch, 2015) on social anxiety symptoms. This result is surprising given the similarities between studies (i.e., one session of ImRs, no cognitive preparation); however, a sub-clinical sample was included in our study whereas participants were diagnosed with SAD in previous research (Nilsson et al., 2012; Norton & Abbott, 2016; Reimer & Moscovitch, 2015). Although the severity of self-reported interaction anxiety in our study was comparable to previous studies (ø40 [this study]; ø37 [Nilsson et al, 2012]; ø44 [Norton & Abbott, 2016]), the low rate of diagnoses in the present sample could indicate that the impairment caused by the social anxiety symptoms was not sufficient to fulfill diagnostic criteria and that participants are able to cope with their negative mental images.

As our ImRs procedure closely followed the procedure of Norton and Abbott (2016), it seems rather unlikely that procedural differences explain the inconsistent findings. Alternatively, ImRs as used in this study might need to be optimized. First, ImRs might not have been optimally delivered (e.g., insufficient reactivation of emotions or the hot-spot; short duration of ImRs [ø 22min in the present study]). Second, we do not know to what extent participants were able to put themselves in their younger self´s perspective. Third, in order to ensure internal validity we used a highly standardized ImRs protocol whereas other studies administered ImRs in a more individualized way and with a more active therapist/ experimenter (e.g., Norton & Abbott, 2016). Fourth, participants were instructed to introduce changes themselves in the present study. Finally, as dysfunctional self-beliefs were not explicitly addressed during ImRs it cannot be ruled out that the rescripting did not show a good enough match with the dysfunctional self-beliefs in the sense of providing corrective information and experiences to modify this belief. This may provide another explanation why ImRs was not associated with long-term effects in our study. Therefore, as the ImRs protocol used in the present study represents only one specific implementation of ImRs, it is conceivable that other versions of ImRs might have yielded more stable effects. For example, in accordance with the protocol by Wild and Clark (2011), a combination of ImRs with CR (Lee & Kwon, 2013; Wild et al., 2008) might yield more stable treatment effects. Different ImRs techniques have been applied in both research and clinical practice; however, it remains an open question how ImRs is best realized (e.g., with or without cognitive preparation; active vs. passive role of patient/therapist), therefore, future research is clearly needed to identify the most effective implementation of ImRs.

Mechanisms Underlying Imagery Rescripting

In line with our hypothesis and with previous evidence (Holmes & Mathews, 2010), a single session of ImRs led to stronger increases of positive emotions than CR and NIC. In contrast, negative emotions significantly decreased across time with no differences between conditions. ImRs and CR more strongly reduced maladaptive intellectual and emotional beliefs from pre- to post-intervention compared to NIC, but only for CR

reductions remained stable across time. In ImRs, neither changes in positive emotions nor in emotional beliefs correlated with symptomatic outcomes.

Although our results indicate that brief ImRs led to beneficial (short-term) effects, it remains to be tested whether the aforementioned mechanisms play a role in producing symptomatic change, as ImRs did not yield improvements on symptom measures in the present study. Moreover, our results challenge the notion that emotionally anchored reappraisal is a mechanism specific to ImRs. In fact, brief CR seems to be more effective in targeting maladaptive emotional beliefs in the longer-term, counter to the theoretical idea that cognitive treatment strategies primarily change intellectual meaning levels (i.e., propositional level). However, after a single session of CR mean levels of emotional beliefs were still high at follow-up and more systematic research is needed to test whether emotional beliefs can be further reduced with multiple treatment sessions.

Limitations

ImRs and CR were delivered as very brief interventions within a non-therapeutic setting. Thus, the interventions deviate from treatment as used in clinical practice limiting its generalizability. However, laboratory-based studies in healthy or subclinical samples are a valuable means to investigate mechanisms involved in psychological treatments under highly controlled and standardized conditions (e.g., Van Den Hout et al., 2017). Although we inquired about the meaning of the mental image, we did not assess how distressing and how relevant the image was regarding participants´ social anxiety symptoms. The distress/impairment caused by the image should be inquired in future studies as it is conceivable that only the modification of distressing images might be associated with long-term effects on social anxiety symptoms. Moreover, it remains unclear whether participants adhered to the ImRs instructions and how distressed they were during ImRs as distress during ImRs was not assessed. Therefore, we cannot verify the correct implementation of ImRs and that emotional activation was sufficient. Emotional beliefs were rated on a one-item VAS, which might reduce reliability.

Conclusion

The present study compared the effects of ImRs vs. CR as stand-alone single-session interventions in socially anxious individuals and aimed to examine mechanisms underly-ing symptomatic change. Results indicate that a single session of CR effectively reduces social anxiety symptoms. The present study raises the question how ImRs for socially anxious individuals should optimally be implemented in order to yield symptomatic change. We propose that more individualized ImRs protocols, higher treatment intensity, cognitive preparation, and/or directly targeting dysfunctional self-beliefs might be neces-sary to yield therapeutic effects.

PsychOpen GOLD

Supplementary Materials

The supplementary material contains a table containing the means and standard deviations of the positive and negative emotions as well as the results of the statistical analyses (for access see Index of Supplementary Materials below).

Index of Supplementary Materials

Strohm, M., Siegesleitner, M., Kunze, A. E., Ehring, T., & Wittekind, C. E. (2021). *Supplementary materials to "Imagery rescripting versus cognitive restructuring for social anxiety: Treatment effects and working mechanisms"* [Additional results]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.5098

References

Ackenheil, M., Stotz-Ingenlath, G., Dietz-Bauer, R., & Vossen, A. (1999). *Mini International Neuropsychiatric Interview – German version 5.0.0.* München, Germany: Psychiatrische Universitätsklinik München.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC, USA: Author.

Arntz, A. (2012). Imagery rescripting as a therapeutic technique: Review of clinical trials, basic studies, and research agenda. *Journal of Experimental Psychopathology, 3*, 189-208. https://doi.org/10.5127/jep.024211

Arntz, A., & Weertman, A. (1999). Treatment of childhood memories: Theory and practice. *Behaviour Research and Therapy, 37*, 715-740. https://doi.org/10.1016/S0005-7967(98)00173-9

Barnard, P. J., & Teasdale, J. D. (1991). Interacting cognitive subsystems: A systematic approach to cognitive-affective interaction and change. *Cognition and Emotion, 5*, 1-39. https://doi.org/10.1080/02699939108411021

PsychOpen GOLD

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*, 49-59. https://doi.org/10.1016/0005-7916(94)90063-9

Carleton, R. N., McCreary, D. R., Norton, P. J., & Asmundson, G. J. (2006). Brief Fear of Negative Evaluation Scale – Revised. *Depression and Anxiety, 23*, 297-303. https://doi.org/10.1002/da.20142

Clark, D., & Wells, A. (1995). A cognitive model of social phobia. In R. Heimberg, M. Liebowitz, D. Hope, & R. Schneier (Eds.), *Social phobia: Diagnosis, assessment and treatment* (pp. 69-93). New York, NY, USA: Guilford Press.

Cooper, M. J., Todd, G., & Turner, H. (2007). The effects of using imagery to modify core emotional beliefs in bulimia nervosa: An experimental pilot study. *Journal of Cognitive Psychotherapy, 21*, 117-122. https://doi.org/10.1891/088983907780851577

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. https://doi.org/10.3758/BF03193146

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders.* Research Version, Patient Edition (SCID-I/P). New York, NY, USA: Biometrics Research, New York State Psychiatric Institute.

Grühn, D., Kotter-Grühn, D., & Röcke, C. (2010). Discrete affects across the adult lifespan: Evidence for multidimensionality and multidirectionality of affective experiences in young, middle-aged and older adults. *Journal of Research in Personality, 44*, 492-500. https://doi.org/10.1016/j.jrp.2010.06.003

Hackmann, A., Clark, D. M., & McManus, F. (2000). Recurrent images and early memories in social phobia. *Behaviour Research and Therapy, 38*, 601-610. https://doi.org/10.1016/S0005-7967(99)00161-8

Hofmann, S. G. (2007). Cognitive factors that maintain social anxiety disorder: A comprehensive model and its treatment implications. *Cognitive Behaviour Therapy, 36*, 193-209. https://doi.org/10.1080/16506070701421313

Holmes, E. A., Lang, T. L., & Shah, D. M. (2009). Developing interpretation bias modification as a "cognitive vaccine" for depressed mood: Imagining positive events makes you feel better than thinking about them verbally. *Journal of Abnormal Psychology, 118*, 76-88. https://doi.org/10.1037/a0012590

Holmes, E. A., & Mathews, A. (2010). Mental imagery in emotion and emotional disorders. *Clinical Psychology Review, 30*, 349-362. https://doi.org/10.1016/j.cpr.2010.01.001

Krönke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*, 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lee, S. W., & Kwon, J.-H. (2013). The efficacy of Imagery Rescripting (IR) for social phobia: A randomized controlled trial. *Journal of Behavior Therapy and Experimental Psychiatry, 44*, 351-360. https://doi.org/10.1016/j.jbtep.2013.03.001

PsychOpen GOLD

Löwe, B., Spitzer, R. L., Zipfel, S., & Herzog, W. (2002). *Gesundheitsfragebogen für Patienten (PHQ-D)*. Komplettversion und Kurzform (2. Auflage). Karlsruhe, Germany: Pfizer.

Mattick, R., & Clarke, J. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy, 36*, 455-470. https://doi.org/10.1016/S0005-7967(97)10031-6

McEvoy, P. M., Hyett, M. P., Bank, S. R., Erceg-Hurn, D. M., Johnson, A. R., Kyron, M. J., . . . Rapee, R. M. (2020). Imagery-enhanced v. verbally-based group cognitive behavior therapy for social anxiety disorder: A randomized clinical trial. *Psychological Medicine, 11.* Advance online publication. https://doi.org/10.1017/S0033291720003001

McEvoy, P. M., & Saulsman, L. M. (2014). Imagery-enhanced cognitive behavioral group therapy for social anxiety disorder: A pilot study. *Behaviour Research and Therapy, 55*, 1-6. https://doi.org/10.1016/j.brat.2014.01.006

Morina, N., Lancee, J., & Arntz, A. (2017). Imagery rescripting as a clinical intervention for aversive memories: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 55*, 6-15. https://doi.org/10.1016/j.jbtep.2016.11.003

Moscovitch, D. A., Gavric, D. L., Merrifield, C., Bielak, T., & Moscovitch, M. (2011). Retrieval properties of negative vs. positive mental images and autobiographical memories in social anxiety: Outcomes with a new measure. *Behaviour Research and Therapy, 49*, 505-517. https://doi.org/10.1016/j.brat.2011.05.009

Nilsson, J.-E., Lundh, L.-G., & Viborg, G. (2012). Imagery rescripting of early memories in social anxiety disorder: An experimental study. *Behaviour Research and Therapy, 50*, 387-392. https://doi.org/10.1016/j.brat.2012.03.004

Norton, A. R., & Abbott, M. J. (2016). The efficacy of imagery rescripting compared to cognitive restructuring for social anxiety disorder. *Journal of Anxiety Disorders, 40*, 18-28. https://doi.org/10.1016/j.janxdis.2016.03.009

Norton, A. R., & Abbott, M. J. (2017). Bridging the gap between aetiological and maintaining factors in social anxiety disorder: The impact of socially traumatic experiences on beliefs, imagery and symptomatology. *Clinical Psychology & Psychotherapy, 24*, 747-765. https://doi.org/10.1002/cpp.2044

Power, M. J., & Dalgleish, T. (1999). Two routes to emotion: Some implications of multi-level theories of emotion for therapeutic practice. *Behavioural and Cognitive Psychotherapy, 27*, 129-141. https://doi.org/10.1017/S1352465899272049

Rapee, R. M., & Abbott, M. J. (2007). Modelling relationships between cognitive variables during and following public speaking in participants with social phobia. *Behaviour Research and Therapy, 45*, 2977-2989. https://doi.org/10.1016/j.brat.2007.08.008

Rapee, R. M., & Heimberg, R. G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy, 35*, 741-756. https://doi.org/10.1016/S0005-7967(97)00022-3

Reichenberger, J., Schwarz, M., König, D., Wilhelm, F. H., Voderholzer, U., Hillert, A., & Blechert, J. (2016). Angst vor negativer sozialer Bewertung: Übersetzung und Validierung der Furcht vor

negativer Evaluation – Kurzskala. *Diagnostica, 62*, 169-181.
https://doi.org/10.1026/0012-1924/a000148

Reimer, S., & Moscovitch, D. (2015). The impact of imagery rescripting on memory appraisals and core beliefs in social anxiety disorder. *Behaviour Research and Therapy, 75*, 48-59.
https://doi.org/10.1016/j.brat.2015.10.007

Romano, M., Moscovitch, D. A., Huppert, J. D., Reimer, S. G., & Moscovitch, M. (2020). The effects of imagery rescripting on memory outcomes in social anxiety disorder. *Journal of Anxiety Disorders, 69*, Article 102169. https://doi.org/10.1016/j.janxdis.2019.102169

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry, 59*(Suppl 20), 22-33.

Shikatani, B., Antony, M. M., Kuo, J. R., & Cassin, S. E. (2014). The impact of cognitive restructuring and mindfulness strategies on postevent processing and affect in social anxiety disorder. *Journal of Anxiety Disorders, 28*, 570-579. https://doi.org/10.1016/j.janxdis.2014.05.012

Stangier, U., Heidenreich, T., Berardi, A., Golbs, U., & Hoyer, J. (1999). Die Erfassung sozialer Phobie durch Social Interaction Anxiety Scale (SIAS) und die Social Phobia Scale (SPS). *Zeitschrift für Klinische Psychologie, 28*, 28-36. https://doi.org/10.1026//0084-5345.28.1.28

Strohm, M. F. (2019). *Imagery rescripting of aversive autobiographical memories: Effects and working mechanisms* [Doctoral dissertation, LMU Munich].
https://edoc.ub.uni-muenchen.de/25436/1/Strohm_Miriam.pdf

Van Den Hout, M. A., Engelhard, I. M., & McNally, R. J. (2017). Thoughts on experimental psychopathology. *Psychopathology Review, a4*, 141-154. https://doi.org/10.5127/pr.045115

Watson, D., & Clark, L. A. (1994). *Manual for the positive and negative affect schedule – Expanded form.* Ames, IA, USA: University of Iowa.

Wild, J., & Clark, D. M. (2011). Imagery rescripting of early traumatic memories in social phobia. *Cognitive and Behavioral Practice, 18*, 433-443. https://doi.org/10.1016/j.cbpra.2011.03.002

Wild, J., Hackmann, A., & Clark, D. M. (2007). When the present visits the past: Updating traumatic memories in social phobia. *Journal of Behavior Therapy and Experimental Psychiatry, 38*, 386-401. https://doi.org/10.1016/j.jbtep.2007.07.003

Wild, J., Hackmann, A., & Clark, D. M. (2008). Rescripting early memories linked to negative images in social phobia: A pilot study. *Behavior Therapy, 39*, 47-56.
https://doi.org/10.1016/j.beth.2007.04.003

Wittchen, H.-U., Zaudig, M., & Fydrich, T. (1997). *SKID – Strukturiertes Klinisches Interview für DSM-IV: Achse I und II.* Göttingen, Germany: Hogrefe.

PsychOpen GOLD

EACLIPT

Clinical Psychology in Europe (CPE) is the official journal of the European Association of Clinical Psychology and Psychological Treatment (EACLIPT).

ψ leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.

PsychOpen GOLD

Scientific Update and Overview

Check for updates

Body Exposure, its Forms of Delivery and Potentially Associated Working Mechanisms: How to Move the Field Forward

Andrea S. Hartmann [1] iD , Eva Naumann [2], Silja Vocks [1] iD , Jennifer Svaldi [2],

Jessica Werthmann [3] iD

[1] *Institute of Psychology, Osnabrück University, Osnabrück, Germany.* **[2]** *Department of Psychology, Eberhard-Karls University Tübingen, Tübingen, Germany.* **[3]** *Institute of Psychology, Albert-Ludwig University Freiburg, Freiburg, Germany.*

Corresponding Author: Andrea S. Hartmann, Osnabrück University, Institute of Psychology, Knollstr. 15, 49069 Osnabrück, Germany. Telephone: +495419696349, Fax: +49541969402. E-mail: andrea.hartmann@uos.de

Abstract

Background: Body image disturbance (BID) is a hallmark feature of eating disorders (EDs) and has proven to be involved in their etiology and maintenance. Therefore, the targeting of BID in treatment is crucial, and has been incorporated in various treatment manuals. One of the most common techniques in the treatment of BID is body exposure (BE), the confrontation with one's own body. BE has been found to be effective in individuals with EDs or high body dissatisfaction. However, BE is applied in a multitude of ways, most of which are based on one or a combination of the hypothesized underlying working mechanisms, with no differential effectiveness known so far.
Method: The aim of this paper is to selectively review the main hypothesized working mechanisms of BE and their translation into therapeutic approaches.
Results and Conclusion: Specifically, we underline that studies are needed to pinpoint the proposed mechanisms and to develop an empirically informed theoretical model of BE. We provide a framework for future studies in order to identify working mechanisms and increase effectiveness of BE.

Keywords

body exposure, eating disorders, body image disturbance, working mechanisms, intervention

> **Highlights**
> - Body exposure (BE) is an effective intervention for body image disturbance in eating disorders.
> - Ways of delivery vary depending on assumed underlying working mechanisms.
> - Impact of attention focus, verbalization, therapist presence, and dosage should be investigated.
> - Research on working mechanisms will improve BE and maximize results for specific patients.

Body image disturbance (BID) is a distinct risk factor for the development and maintenance of eating disorders (EDs), and potentially contributes to relapse after treatment (e.g., Glashouwer et al., 2019). Furthermore, targeting body dissatisfaction is associated with better overall treatment outcome (Wilson et al., 2002). Thus, the improvement of body image should be a key element of ED treatment, e.g. in the form of body exposure (BE), alongside the normalization of nutrition and eating behaviors. This paper aims to selectively review the theoretical rationales underlying potential working mechanisms of BE, the empirical evidence for these rationales, and the corresponding therapeutic application of BE. Another aim is to review future research ideas on mechanisms, BE delivery, and moderators of BE effects in order to foster clinicians' use of BE as an effective intervention strategy.

Efficacy of Body Exposure

A meta-analytical review indicated that BE is effective as stand-alone intervention for BID (Alleva et al., 2015). The analysis included 62 original studies on the effectiveness of stand-alone interventions to improve body image that had a control group, random allocation to conditions, and at least one pre- and posttest measure. Two interventions that can be broadly viewed as BE - namely exposure exercises and guided imagery exercises - showed significant intervention effects on body image. The meta-analysis further demonstrated that effects were stronger when targeting individuals with body concerns as compared to unselected groups (Alleva et al., 2015). In an extension of this finding, a more recent review (Griffen et al., 2018) focused on summarizing the effects of BE in distinct groups of individuals with various ED diagnoses separately and mixed, as well as individuals with obesity, body dysmorphic disorder, and non-clinical individuals. Their search yielded a total of 15 studies evaluating BE. For all participant groups, at least preliminary effectiveness of BE was shown. However, due to a scarcity of studies no differential effectiveness of various forms of BE could be determined (Griffen et al., 2018).

Notably, some individuals do not benefit from BE, as evidenced by findings that on certain measures, between-group effects are significant while group by time interaction

effects are not (e.g., Delinsky & Wilson, 2006). Research and reports on symptom deterioration or treatment dropouts are rare. In a randomized controlled trial by Hildebrandt and colleagues (2012), self-injurious behaviors and subsequent study dropout occurred in the BE condition but not the control condition. In a study by Delinsky and Wilson (2006), the only dropouts occurred in the BE condition (without significant attrition differences between conditions), and the participants who dropped out also had higher depression scores at the outset. Accordingly, while BE might deteriorate symptoms in emotionally unstable patients, frequency of symptom deterioration or treatment discontinuation cannot be extrapolated from current data.

In sum, BE seems to be effective for the majority of patients. A common characteristic of BE procedures is a systematic examination of one's own body by the patient – in a mirror or through recorded videos – over a varying number of sessions. However, different BE versions exist in which the specific BE approach varies in several aspects, and the (clinical) decision for the specific BE approach often relies on the hypothesized underlying working mechanism.

Hypothesized Working Mechanisms: Theoretical Ideas on How BE Reduces Body Image Disturbance

The theoretical accounts of BE show distinct differences, resulting in a variety of specific intervention approaches. Here, we will briefly review four theoretical ideas that have previously been proposed. Moreover, where available, we present empirical evidence and the respective treatment implications. Of note, the field is only just beginning to develop a comprehensive understanding of how exposure might work, and an integrated model of these rationales is lacking. Thus, while in the following the theoretical ideas are discussed as discrete working mechanisms, it might very well be that they all work alongside each other or interact (Lass-Hennemann et al., 2018). Furthermore, there may also be a general working mechanism, e.g., the generally structured preoccupation with one's body without avoidance or safety behaviors.

First, a hypothesis derived from exposure research in anxiety disorders posits that habituation to negative emotion and distress on psychological and biophysiological processing levels is responsible for the positive effects of BE. From a theoretical perspective, repeated and prolonged exposure to the conditioned stimulus ''seeing one's own body'' (CS) is assumed to induce decreases in the conditioned negative reaction (CR) by preventing negative reinforcement, e.g., avoidance (Benito et al., 2018; Craske et al., 2014). Indeed, there is evidence for a reduction of self-reported negative affect between and within exposure sessions (e.g., Trentowska et al., 2017). While these findings are supported by some studies assessing physiological parameters (e.g. emotional arousal measured by means of voice stress analysis; Baur et al., 2020), other findings, e.g. from

studies assessing heart rate as a physiological measure of change in distress during BE, are more ambiguous (Trentowska et al., 2017; Vocks et al., 2007). One reason for this inconsistency might be that BE elicits a multitude of emotions in individuals with BID (e.g., Naumann et al., 2013). For instance, in individuals with EDs, disgust has been shown to play a more important role than anxiety (e.g., von Spreckelsen et al., 2018). Moreover, disgust seems more resistant to psychological and physiological habituation processes in other disorders (Olatunji et al., 2009), and is influenced more likely by counterconditioning (e.g., Engelhard et al., 2014). Recently, potential working mechanisms of exposure (in anxiety research) have been overhauled by the so-called inhibitory learning approach. Accordingly, the working mechanism of exposure lies in the development and strengthening of nonthreat associations in memory during exposure (e.g., Craske et al., 2008; Foa & McLean, 2016).

Thus, within an exposure framework of BE, three potential working mechanisms have been suggested: habituation, counterconditioning, and inhibitory learning. While all three approaches are based on an exposure rationale, each offers a distinct and differential therapeutic application of BE in a clinical context. Treatment manuals postulating habituation as a working mechanism recommend that patients mainly focus on their negatively valenced body parts over an extended period of time in order to activate negative affect, which consequently can be reduced (Vocks et al., 2018). Treatment manuals based on the counterconditioning mechanisms should aim to change the unwanted reaction (negative affect) when confronted with the stimulus (body). Thus, they might suggest to rather focus on positively valenced body parts, coupled with an instruction to do something positive for/with one's body (e.g., use body lotion) or, to focus on negatively valenced body parts while instructing to elicit positive thoughts about the body and/or remember what the body already has achieved (e.g, Vocks et al., 2018). And lastly, treatment manuals using inhibitory learning as a rationale would aim to use as many different exposure exercises as possible in order to maximize the possibilities to create nonthreat associations.

Another theoretical rationale of BE is based on the idea of attention bias modification. The hypothesis was derived from data demonstrating a negative attentional bias to subjectively unattractive body parts when confronted with one's own body in individuals with EDs (e.g., Bauer et al., 2017). It was hypothesized that a change in this dysfunctional attention pattern might alter the associated negative affect. Some studies have demonstrated that a focus on positively valenced body parts leads to an improvement on measures of body image (Glashouwer et al., 2016; Krohmer et al., authors' unpublished data; Smeets et al., 2011), and some (Krohmer et al., authors' unpublished data) but not all (Glashouwer et al., 2016) have reported a concurrent change in attention patterns. However, one study did not find differential effects between a negative and a positive focus condition on body dissatisfaction, body-related checking, body concerns, and negative mood from pre- to post-BE (e.g., Jansen et al., 2016). This contradicts the idea

of attention bias modification as the only working mechanism of BE. Following this rationale, corresponding therapeutic BE approaches asked patients to focus on positively valenced body parts only (Jansen et al., 2016; Vocks et al., 2018) or to state their emotional connotations of the respective body parts while distributing their attention evenly (Svaldi & Tuschen-Caffier, 2018).

A third theoretical rationale of BE is based on the hypothesis of reduction of body perception distortion in individuals with EDs. Most individuals with EDs overestimate the dimensions of their own body (e.g., Mohr et al., 2016; Volpe et al., 2018). Furthermore, there is some (Norris, 1984), but also contrasting (Lewer et al., 2017; Vocks et al., 2007) evidence that distorted body perception might change over the course of BE. More recently, a systematic review suggested that the construct of distorted perception may be misleading as the distortion may rather stem from a dysfunctional cognitive-evaluative component of body image than from perceptual deficits (Mölbert et al., 2017). Following this rationale, one would advise an even distribution pattern and the use of non-judgmental language (Hildebrandt et al., 2012) during BE.

A fourth theoretical rationale suggests that central dysfunctional cognitions (e.g., interpretation and memory biases, e.g., Korn et al., 2020) of BID are changed through (implicit) cognitive restructuring in the course of BE. Such cognitive restructuring can be achieved by inducing cognitive dissonance (e.g., between the dysfunctional belief "My stomach looks fat" and the behavior of describing the stomach neutrally), which may in turn reduce body-related negative schemata (Williamson et al., 2004). In addition to the above-mentioned induction of cognitive dissonance and cognitive restructuring, therapeutic approaches of BE derived from this hypothesis instruct patients to either focus on positively valenced body parts or to focus on all body parts evenly, while describing their body positively or neutrally (i.e., with the therapist present; e.g., Jansen et al., 2016; Klimek et al., 2016; Luethcke et al., 2011).

All of these aspects are noteworthy, as BE seems, in general, a promising tool to address body image disturbances in clinical and non-clinical populations (Alleva et al., 2015), even though with only small effect sizes as a stand-alone technique in the latter. Accordingly, there is a need to refine the theoretical rationale as well as (experimental) research on working mechanisms in order to improve the technique and potentially individualize it in the future to maximize outcome in specific patients.

Suggested Foci in Future Research

It is important for future research to focus on factors that determine its positive effects. In the following, we describe variables that require systematic examination.

Where Should One Look During BE?

As briefly reviewed above, depending on the theoretical rationale, BE approaches differ in whether patients are instructed to focus selectively on positively or negatively valenced body parts, or evenly on all body parts. Given that these foci might elicit emotions that may or may not be necessary to reach the intervention goal, it is essential to understand individual needs and differences. In one study, interventions with a focus on exclusively positive or negative body parts successfully reduced body dissatisfaction, body-related checking, body concerns, and negative mood in women with high levels of body dissatisfaction (Jansen et al., 2016). Moreover, the negative focus condition yielded a stronger decrease in body-related avoidance behavior over the follow-up period. For comparison studies, we propose to consider another effective form of BE, which comprises instructions to focus on all body parts from head to toe, successively, in order to correct distorted body perception and alter viewing patterns. Furthermore, we suggest testing a form in which body parts are clustered by their indication of weight gain or status (e.g., thighs, bottom, stomach vs. knees, ankles, forearms), instead of by their subjective valence. This might be of particular interest if the hypothesized working mechanism is dissolution of the conditioned association, as it allows for exposure to the most fear-inducing body parts, given that fear of weight gain is a central concept of individuals with EDs (e.g., Rodgers et al., 2018).

How Should Verbalization Be Instructed During BE?

Another large difference between previous studies lies in the type of body-related descriptions provided by participants, i.e. whether they purely describe their body, or the associated emotions and cognitions, or both. While a negatively toned description might strengthen the experience of BE (in the sense of a stronger habituation effect), subsequently leading to a more effective dissolution of negative body-related affect, a mainly positive or neutral, non-judgmental description might strengthen the decrease in negative affect by correcting distorted perception, thus altering dysfunctional attention processes or cognitive dissonance processes (rather like inhibitory learning). So far, only two studies have compared different forms of instructed verbalizations. In the first study, the authors compared two neutral versions of BE to a cognitive dissonance version in which participants were instructed to describe body parts using positive verbalizations. While all three forms led to improvements on measures of ED and body image, only the cognitive dissonance version of BE yielded an increase in body satisfaction (Luethcke et al., 2011). In the second study, a positive and a negative full-body verbalization condition were compared in healthy individuals. Both interventions yielded improvements in emotional arousal and body satisfaction between sessions. However, within sessions, the negative but not the positive verbalization condition led to a decrease in positive affect and body satisfaction and an increase in negative affect (Tanck et al., authors'

unpublished data). To further disentangle different forms of verbalisation, we propose to compare a neutral description of what patients see, and a description of positive or negative aspects of each body part in future studies. Thereby, while manipulating the form of verbalization, the attentional focus should be controlled (e.g., by asking patients to describe every part of their body from head to toe).

Is a Therapist Needed in BE?

To the best of our knowledge, there are no studies comparing BE with and without a therapist present. Such investigations would be highly relevant, as the presence of a therapist could impact the effectiveness of the intervention, particularly when considering cognitive dissonance as a working mechanism. Comparative studies have looked at differences in the effectiveness of guided vs. unguided BE (Díaz-Ferrer et al., 2015; Díaz-Ferrer et al., 2017; Moreno-Domínguez et al., 2012). For example, women with body dissatisfaction and subclinical EDs underwent either an *unguided* version, in which they freely explored self-chosen body parts and were instructed to verbalize associated emotions and cognitions, or a *guided* version, in which they focused on all body parts and had to describe them using neutral words. Both conditions were found to be effective in reducing BID, with a slight superiority of the unguided condition. However, heart rate and skin conductance observed within sessions indicated that the two techniques might act through different mechanisms (Díaz-Ferrer et al., 2017), with a stronger increase in both indicators in the unguided condition. Notably, the conditions in the comparison studies varied not only with respect to therapists' active guidance during BE, but also regarding the body parts which were focused on and the way in which body parts were described. Thus, in order to understand the impact of therapist presence and guidance during BE, future research should compare guided and unguided versions of BE while controlling for focus and type of verbalization.

How Much BE Is Needed?

The ideal intensity of BE remains unclear. On the one hand, intensity can be captured as frequency of sessions. In anxiety disorder research, the frequency of exposure is assumed to be a major factor in treatment effectiveness (Wolitzky-Taylor et al., 2008). In EDs, several findings highlight that therapeutic effects might occur mostly between rather than within sessions (e.g., Hilbert et al., 2002). Thus, multiple sessions are necessary, which is further underlined by the finding that short-term exposure leads to an activation and deterioration of body satisfaction and negative affect (Veale et al., 2016). Findings from studies investigating the effects of different numbers of sessions are important, because they may, for instance, allay clinicians' fears of overwhelming the patient when delivering multiple BE sessions.

On the other hand, intensity can also be captured as duration of single sessions, thus the length of a BE therapy session, BE sessions over a whole day, or exposure until a reduction in anxiety to a certain predefined extent is realized. In intensive exposure ("flooding"), aversive stimuli are presented at the highest level of intensity, while gradual exposure follows a stepwise approach starting at a low level of intensity. Previous research in the area of obsessive-compulsive disorder suggests that intensive exposure might lead to a stronger short-term reduction of anxiety symptoms. By contrast, gradual exposure might be more helpful for reducing emotions that habituate more slowly, such as disgust (Olatunji et al., 2009). More recent studies in the area of anxiety disorders advocate for variability in the exposure hierarchy in order to maximize inhibitory learning (e.g., Knowles & Olatunji, 2019). Future research should test whether variations in intensity impact BE effects on BID. Besides frequency and duration of sessions, potentially relevant moderating variables in the context of intensity of BE may relate to the setting (e.g., mirror size, light, distance to mirror) or clothing (everyday vs. tight clothes vs. underwear).

Who Benefits or Does not Benefit From BE?

Evidence of differential effectiveness of BE in specific groups is limited by the low diversity of the groups researched so far. Men have been overlooked in body image research, including BE interventions (Burlew & Shurts, 2013), and BE in individuals with comorbidities remains to be investigated. Additionally, as Alleva et al. (2015) highlighted, individuals of middle to older age have also been neglected in past BE research.

Furthermore, for body dysmorphic disorder, another mental illness with the core symptom of BID, BE (mirror retraining), also represents an essential part of the CBT protocol (e.g., Wilhelm et al., 2013). However, to date, no study has examined the effectiveness of this technique detached from the overall CBT treatment. Further research into the effectiveness of BE in mental disorders potentially associated with BID, namely borderline personality disorder, posttraumatic stress disorder, or social anxiety disorder (Dyer et al., 2013; Dyer et al., 2015) is also lacking.

Lastly, a comprehensive evaluation of BE effectiveness should also include the systematic assessment of side effects, adverse events, or predictors of non-responders, and a subsequent trade-off between positive effects and negative aspects for single patient groups. As looking at oneself in a mirror can lead to significant distress and a worsening of negative affect (Veale et al., 2016; Walker et al., 2012; Windheim et al., 2011), BE might destabilize some patients. Eventually, extending research to subgroups will help to formulate diagnosis- and patient group-specific treatment guidelines, which will move us closer to establishing individualized evidence-based treatments.

What Might Further Influence the Efficacy of BE?

Several potential moderators may be worthy of further investigation, because they may have confounded previous research results. Moderating factors may also influence practitioner's decision to implement BE. Given the scarcity of previous research, we are not able to quantify the impact of, for example, current weight, gender- and weight-match between patient and therapist, current status of treatment, chronicity of symptoms, level of habitual checking and avoidance, and the delivery in groups vs. alone on the effectiveness of BE. We suggest that all of these factors should be assessed in future studies to provide information regarding their impact on BE effects and on clinician's decision to implement BE.

Tools for Evaluating BE Mechanisms and Efficacy

Past studies varied regarding outcome and process variables. To understand the differential effectiveness of BE on various levels of experience, a comprehensive set of process and outcome measures needs to be considered. First, we suggest that different facets of body image should be assessed in order to capture processes and outcomes on all levels of BID (i.e. perceptual, cognitive-affective, and behavioral). Second, we advocate for the adoption of a multi-method approach encompassing self- and expert-report measures, as well as objective measures in order to elucidate mechanisms of BE on as many processing levels as possible. The former might include self- and external report measures on body dissatisfaction and disorder-specific symptomatology. The latter might consist of psychobiological indicators of emotional activation indexing fear- and anxiety-related differences in the autonomic nervous system, e.g., such as fear-potentiated startle and heart rate, but also indices of attention allocation and information processing as well as the very recent approach of vocal arousal.

Conclusion

Despite findings regarding the effectiveness of BE in intervention studies, it is still largely unknown which version works best for whom. Thus, first, lab-based experimental studies need to be conducted to isolate the effect of potential working mechanisms and test their impact within the different proposed forms of BE on BID outcomes (Glashouwer et al., 2020). Current studies from our workgroups target this research gaps by setting out to differentiate attention foci and verbalization forms measuring self-reported, peripherphysiological, and eye-tracking outcomes. Findings from these and other studies can then inform theory-based and empirically based models on key processes, and can advance refined etiological models of BID. In the future, interventions based on these models can then be tested in larger randomized controlled trials including additional analyses of moderators to identify which specific BE procedure is maximally

successful (or unsuccessful) for a specific patient subsample. Of further relevance, research needs to prove that the positive effects of BE outweigh the fact that this technique can be strenuous for patients, as they are confronted with the very thing they fear the most.

References

Alleva, J. M., Sheeran, P., Webb, T. L., Martijn, C., & Miles, E. (2015). A meta-analytic review of stand-alone interventions to improve body image. *PLoS One, 10*(9), Article e0139177. https://doi.org/10.1371/journal.pone.0139177

Bauer, A., Schneider, S., Waldorf, M., Braks, K., Huber, T. J., Adolph, D., & Vocks, S. (2017). Selective visual attention towards oneself and associated state body satisfaction: An eye-tracking study in adolescents with different types of eating disorders. *Journal of Abnormal Child Psychology, 45*(8), 1647-1661. https://doi.org/10.1007/s10802-017-0263-z

Baur, J., Krohmer, K., Naumann, E., Tuschen-Caffier, B., & Svaldi, J. (2020). Vocal arousal: A physiological correlate of body distress in women with overweight and obesity. *Eating and Weight Disorders, 25*(5), 1161-1169. https://doi.org/10.1007/s40519-019-00744-2

Benito, K. G., Machan, J., Freeman, J. B., Garcia, A. M., Walther, M., Frank, H., . . . Franklin, M. (2018). Measuring fear change within exposures: Functionally-defined habituation predicts outcome in three randomized controlled trials for pediatric OCD. *Journal of Consulting and Clinical Psychology, 86*(7), 615-630. https://doi.org/10.1037/ccp0000315

Burlew, L. D., & Shurts, W. M. (2013). Men and body image: Current issues and counseling implications. *Journal of Counseling and Development, 91*(4), 428-435. https://doi.org/10.1002/j.1556-6676.2013.00114.x

Craske, M. G., Kircanski, K., Zelikowsky, M., Mystkowski, J., Chowdhury, N., & Baker, A. (2008). Optimizing inhibitory learning during exposure therapy. *Behaviour Research and Therapy, 46*(1), 5-27. https://doi.org/10.1016/j.brat.2007.10.003

PsychOpen GOLD

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

Delinsky, S. S., & Wilson, G. T. (2006). Mirror exposure for the treatment of body image disturbance. *International Journal of Eating Disorders, 39*(2), 108-116. https://doi.org/10.1002/eat.20207

Díaz-Ferrer, S., Rodríguez-Ruiz, S., Ortega-Roldán, B., Moreno-Domínguez, S., & Fernández-Santaella, M. C. (2015). Testing the efficacy of pure versus guided mirror exposure in women withbulimia nervosa: A combination of neuroendocrine and psycho-logical indices. *Journal of Behavior Therapy and Experimental Psychiatry, 48*, 1-8. https://doi.org/10.1016/j.jbtep.2015.01.003

Díaz-Ferrer, S., Rodríguez-Ruiz, S., Ortega-Roldán, B., Mata-Martín, J. L., & Carmen Fernández-Santaella, M. (2017). Psychophysiological changes during pure vs guided mirror exposure therapies in women with high body dissatisfaction: What are they learning about their bodies? *European Eating Disorders Review, 25*(6), 562-569. https://doi.org/10.1002/erv.2546

Dyer, A. S., Borgmann, E., Kleindienst, N., Feldmann, R. E., Jr., Vocks, S., & Bohus, M. (2013). Body image in patients with posttraumatic stress disorder after childhood sexual abuse and co-occurring eating disorder. *Psychopathology, 46*(3), 186-191. https://doi.org/10.1159/000341590

Dyer, A. S., Feldmann, R. E., Jr., & Borgmann, E. (2015). Body-related emotions in posttraumatic stress disorder following childhood sexual abuse. *Journal of Child Sexual Abuse, 24*(6), 627-640. https://doi.org/10.1080/10538712.2015.1057666

Engelhard, I. M., Leer, A., Lange, E., & Olatunji, B. O. (2014). Shaking that icky feeling: Effects of extinction and counterconditioning on disgust-related evaluative learning. *Behavior Therapy, 45*(5), 708-719. https://doi.org/10.1016/j.beth.2014.04.003

Foa, E. B., & McLean, C. P. (2016). The efficacy of exposure therapy for anxiety-related disorders and its underlying mechanisms: The case of OCD and PTSD. *Annual Review of Clinical Psychology, 12*, 1-28. https://doi.org/10.1146/annurev-clinpsy-021815-093533

Glashouwer, K. A., Brockmeyer, T., Cardi, V., Jansen, A., Murray, S. B., Blechert, J., . . . Werthmann, J. (2020). Time to make a change: A call for more experimental research on key mechanisms on in anorexia nervosa. *European Eating Disorders Review, 28*(4), 361-367. https://doi.org/10.1002/erv.2754

Glashouwer, K. A., Jonker, N. C., Thomassen, K., & de Jong, P. J. (2016). Take a look at the bright side: Effects of positive body exposure on selective visual attention in women with high body dissatisfaction. *Behaviour Research and Therapy, 83*, 19-25. https://doi.org/10.1016/j.brat.2016.05.006

Glashouwer, K. A., van der Veer, R. M., Adipatria, F., de Jong, P. J., & Vocks, S. (2019). The role of body image disturbance in the onset, maintenance, and relapse of anorexia nervosa: A systematic review. *Clinical Psychology Review, 74*, Article 101771. https://doi.org/10.1016/j.cpr.2019.101771

PsychOpen GOLD

Griffen, T. C., Naumann, E., & Hildebrandt, T. (2018). Mirror exposure therapy for body image disturbances and eating disorders: A review. *Clinical Psychology Review, 65*, 163-174. https://doi.org/10.1016/j.cpr.2018.08.006

Hilbert, A., Tuschen-Caffier, B., & Vögele, C. (2002). Effects of prolonged and repeated body image exposure in binge-eating disorder. *Journal of Psychosomatic Research, 52*(3), 137-144. https://doi.org/10.1016/S0022-3999(01)00314-2

Hildebrandt, T., Loeb, K., Troupe, S., & Delinsky, S. (2012). Adjunctive mirror exposure for eating disorders: A randomized controlled pilot study. *Behaviour Research and Therapy, 50*(12), 797-804. https://doi.org/10.1016/j.brat.2012.09.004

Jansen, A., Voorwinde, V., Hoebink, Y., Rekkers, M., Martijn, C., & Mulkens, S. (2016). Mirror exposure to increase body satisfaction: Should we guide the focus of attention towards positively or negatively evaluated body parts? *Journal of Behavior Therapy and Experimental Psychiatry, 50*, 90-96. https://doi.org/10.1016/j.jbtep.2015.06.002

Klimek, P., Grotzinger, A., & Hildebrandt, T. (2016). Using acceptance to improve body image among individuals with eating disorders. In A. F. Haynos, E. M. Forman, M. L. Butryn, & J. Lillis (Eds.), *Mindfulness & acceptance for treating eating disorders & weight concerns: Evidence based interventions* (pp. 121-142). Oakland, CA, USA: New Harbinger Publications.

Knowles, K. A., & Olatunji, B. O. (2019). Enhancing inhibitory learning: The utility of variability in exposure. *Cognitive and Behavioral Practice, 26*(1), 186-200. https://doi.org/10.1016/j.cbpra.2017.12.001

Korn, J., Dietel, F. A., & Hartmann, A. S. (2020). Testing the specificity of interpretation biases in women with eating disorder symptoms: An online experimental assessment. *International Journal of Eating Disorders, 53*(3), 372-382. https://doi.org/10.1002/eat.23201

Lass-Hennemann, J., Tuschen-Caffier, B., & Michael, T. (2018). Expositionsverfahren. In J. Margraf & S. Schneider (Eds.), *Lehrbuch der Verhaltenstherapie* (4th ed., Vol. 1, pp. 411-424). Heidelberg, Germany: Springer.

Lewer, M., Kosfelder, J., Michalak, J., Schroeder, D., Nasrawi, N., & Vocks, S. (2017). Effects of a cognitive-behavioral exposure-based body image therapy for overweight females with binge eating disorder: A pilot study. *Journal of Eating Disorders, 5*(1), Article 43. https://doi.org/10.1186/s40337-017-0174-y

Luethcke, C. A., McDaniel, L., & Becker, C. B. (2011). A comparison of mindfulness, nonjudgmental, and cognitive dissonance-based approaches to mirror exposure. *Body Image, 8*(3), 251-258. https://doi.org/10.1016/j.bodyim.2011.03.006

Mölbert, S. C., Klein, L., Thaler, A., Mohler, B. J., Brozzo, C., Martus, P., . . . Giel, K. E. (2017). Depictive and metric body size estimation in anorexia nervosa and bulimia nervosa: A systematic review and meta-analysis. *Clinical Psychology Review, 57*, 21-31. https://doi.org/10.1016/j.cpr.2017.08.005

Mohr, H. M., Rickmeyer, C., Hummel, D., Ernst, M., & Grabhorn, R. (2016). Altered visual adaptation to body shape in eating disorders: Implications for body image distortion. *Perception, 45*(7), 725-738. https://doi.org/10.1177/0301006616633385

PsychOpen GOLD

Moreno-Domínguez, S., Rodríguez-Ruiz, S., Fernández-Santaella, M. C., Jansen, A., & Tuschen-Caffier, B. (2012). Pure versusguided mirror exposure to reduce body dissatisfaction: A preliminary study with university women. *Body Image, 9*, 285-288. https://doi.org/10.1016/j.bodyim.2011.12.001

Naumann, E., Trentowska, M., & Svaldi, J. (2013). Increased salivation to mirror exposure in women with binge eating disorder. *Appetite, 65*, 103-110. https://doi.org/10.1016/j.appet.2013.01.021

Norris, D. L. (1984). The effects of mirror confrontation on self-estimation of body dimensions in anorexia nervosa, bulimia and two control groups. *Psychological Medicine, 14*(4), 835-842. https://doi.org/10.1017/S0033291700019802

Olatunji, B. O., Wolitzky-Taylor, K. B., Willems, J., Lohr, J. M., & Armstrong, T. (2009). Differential habituation of fear and disgust during repeated exposure to threat-relevant stimuli in contamination-based OCD: An analogue study. *Journal of Anxiety Disorders, 23*(1), 118-123. https://doi.org/10.1016/j.janxdis.2008.04.006

Rodgers, R. F., DuBois, R., Frumkin, M. R., & Robinaugh, D. J. (2018). A network approach to eating disorder symptomatology: Do desire for thinness and fear of gaining weight play unique roles in the network? *Body Image, 27*, 1-9. https://doi.org/10.1016/j.bodyim.2018.07.004

Smeets, E., Jansen, A., & Roefs, A. (2011). Bias for the (un) attractive self: On the role of attention in causing body (dis) satisfaction. *Health Psychology, 30*(3), 360-367. https://doi.org/10.1037/a0022095

Svaldi, J., & Tuschen-Caffier, B. (2018). *Bulimia nervosa*. Göttingen, Germany: Hogrefe.

Trentowska, M., Svaldi, J., Blechert, J., & Tuschen-Caffier, B. (2017). Does habituation really happen? Investigation of psycho-biological responses to body exposure in bulimia nervosa. *Behaviour Research and Therapy, 90*, 111-122. https://doi.org/10.1016/j.brat.2016.12.006

Veale, D., Miles, S., Valiallah, N., Butt, S., Anson, M., Eshkevari, E., . . . Baldock, E. (2016). The effect of self-focused attention and mood on appearance dissatisfaction after mirror-gazing: An experimental study. *Journal of Behavior Therapy and Experimental Psychiatry, 52*, 38-44. https://doi.org/10.1016/j.jbtep.2016.03.002

Vocks, S., Bauer, A., & Legenbauer, T. (2018). *Körperbildtherapie bei Anorexia und Bulimia nervosa: ein kognitiv-verhaltenstherapeutisches Behandlungsprogramm (3. Aufl.)* [Body image therapy in anorexia and bulimia nervosa: A cognitive behavioral treatment program (3rd ed.)]. Göttingen, Germany: Hogrefe.

Vocks, S., Legenbauer, T., Wächter, A., Wucherer, M., & Kosfelder, J. (2007). What happens in the course of body exposure? Emotional, cognitive, and physiological reactions to mirror confrontation in eating disorders. *Journal of Psychosomatic Research, 62*(2), 231-239. https://doi.org/10.1016/j.jpsychores.2006.08.007

Volpe, U., Monteleone, A. M., & Monteleone, P. (2018). Diagnostic classification of eating disorders: The role of body image. In M. Cuzzolaro & S. Fassino (Eds.), *Body image, eating, and weight* (pp. 57-66). Cham, Switzerland: Springer.

PsychOpen GOLD

von Spreckelsen, P., Glashouwer, K. A., Bennik, E. C., Wessel, I., & de Jong, P. J. (2018). Negative body image: Relationships with heightened disgust propensity, disgust sensitivity, and self-directed disgust. *PLoS One, 13*(6), Article e0198532. https://doi.org/10.1371/journal.pone.0198532

Walker, D. C., Murray, A. D., Lavender, J. M., & Anderson, D. A. (2012). The direct effects of manipulating body checking in men. *Body Image, 9*(4), 462-468. https://doi.org/10.1016/j.bodyim.2012.06.001

Wilhelm, S., Phillips, K. A., & Steketee, G. (2013). *A cognitive-behavioral treatment manual for body dysmorphic disorder.* New York, NY, USA: Guilford Press.

Williamson, D. A., White, M. A., York-Crowe, E., & Stewart, T. M. (2004). Cognitive-behavioral theories of eating disorders. *Behavior Modification, 28*(6), 711-738. https://doi.org/10.1177/0145445503259853

Wilson, G. T., Fairburn, C. C., Agras, W. S., Walsh, B. T., & Kraemer, H. (2002). Cognitive-behavioral therapy for bulimia nervosa: Time course and mechanisms of change. *Journal of Consulting and Clinical Psychology, 70*(2), 267-274. https://doi.org/10.1037/0022-006X.70.2.267

Windheim, K., Veale, D., & Anson, M. (2011). Mirror gazing in body dysmorphic disorder and healthy controls: Effects of duration of gazing. *Behaviour Research and Therapy, 49*(9), 555-564. https://doi.org/10.1016/j.brat.2011.05.003

Wolitzky-Taylor, K. B., Horowitz, J. D., Powers, M. B., & Telch, M. J. (2008). Psychological approaches in the treatment of specific phobias: A meta-analysis. *Clinical Psychology Review, 28*(6), 1021-1037. https://doi.org/10.1016/j.cpr.2008.02.007

Scientific Update and Overview

Check for updates

Has the Time Come to Stop Using the "Standardised Mean Difference"?

Pim Cuijpers [1]

[1] *Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health research institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.*

Abstract

Background: Most meta-analyses use the 'standardised mean difference' (effect size) to summarise the outcomes of studies. However, the effect size has important limitations that need to be considered.

Method: After a brief explanation of the standardized mean difference, limitations are discussed and possible solutions in the context of meta-analyses are suggested.

Results: When using the effect size, three major limitations have to be considered. First, the effect size is still a statistical concept and small effect sizes may have considerable clinical meaning while large effect sizes may not. Second, specific assumptions of the effect size may not be correct. Third, and most importantly, it is very difficult to explain what the meaning of the effect size is to non-researchers. As possible solutions, the use of the 'binomial effect size display' and the number-needed-to-treat are discussed. Furthermore, I suggest the use of binary outcomes, which are often easier to understand. However, it is not clear what the best binary outcome is for continuous outcomes.

Conclusion: The effect size is still useful, as long as the limitations are understood and also binary outcomes are given.

Keywords

effect size, standardised mean difference, meta-analysis, outcome studies

It was a historical event for the field of clinical psychology. In his presidential address to the American Educational Research Association in 1976 in San Francisco, Gene Glass

not only coined the term "meta-analysis" but he also introduced the basic ideas of modern meta-analyses (Hunt, 1997). This event is broadly considered as the starting point of modern meta-analyses (Hunt, 1997). Since then this method has conquered the field of clinical psychology and beyond, and meta-analyses have become the standard for integrating the results of multiple studies on the same research question into one estimate of the effects or associations. Meta-analyses are now considered to be the gold standard for estimating the effects of interventions and are at the basis of treatment guidelines for mental health and other problems as well as policy recommendations about treatments.

Glass brought forward two basic ideas that are at the core of modern meta-analyses. The first idea he brought forward was the 'standardized mean difference', or what is often called the 'effect size'. The effect size indicates the difference between two conditions after the intervention in terms of standard deviations instead of actual scores on an outcome instrument. This makes the outcomes 'standardised' and therefore they can be compared across studies. The other basic idea of meta-analyses that Glass brought forward was that these standardised outcomes can be pooled across studies, while weighting them based on the size of the samples. This pooling of the standardised outcomes results in one overall estimate of the true effect size across multiple studies.

It is now 45 years ago that these two basic ideas were introduced. The second idea, the pooling of outcomes according to the size of the study, has hardly been disputed since the introduction by Glass. But the idea of the standardised mean difference has been more controversial over the years. In this paper, I will focus on the standardised mean difference. I will discuss whether this is still the best way of indicating the outcomes of interventions or associations between variables or whether it is better to start using binary outcomes instead. I will call the standardised mean difference the 'effect size' which is in fact not correct (Higgins & Green, 2011), but I will still do that to increase the readability of this paper.

The Effect Size

It was a brilliant idea to indicate the difference between two groups in terms of the standard deviation of the outcome measure, instead of the actual difference in scores between the groups. This not only allows to compare these outcomes across different studies regardless of the outcome instrument used, but it also gives an indication of the size of the effect. Previous research often only indicated whether the difference between two groups was significant or not. However, that is not very informative and does not say anything about the size of the difference. Whether or not a difference is significant depends on the size of the sample, and even a tiny difference becomes significant when the sample size is large enough. The effect size solved this problem, because it goes beyond significance levels and indicates how large the difference is. Cohen suggested

that an effect size of 0.2 should be considered as small, 0.5 as moderate and 0.8 as large (Cohen, 1988).

However, the use of effect sizes also has several important limitations. One important limitation is that it is still a statistical concept. It may indicate the strength of an outcome, but it still cannot say anything about the clinical relevance of the outcome (Cuijpers, Turner, Koole, van Dijke, & Smit, 2014). The clinical relevance of an effect size depends on the content. For example, an effect size of 0.1 would be considered a major breakthrough if years to mortality would be the outcome. And effect size of 0.1 with "knowledge of depression" as outcome, however, would be considered trivial by most people. This means that the categories of small, moderate and large effect sizes, as given by Cohen (1988) may be misleading because the effect size depends too much on what the outcome actually is. It should be noted that this was fully acknowledged by Cohen.

One solution to the problem that the effect size is a statistical concept, could be the use of the 'Minimal clinically important difference' (MCID; McGlothlin & Lewis, 2014). The MCID is the smallest difference in score considered clinically worthwhile by the patient and it captures both the magnitude of improvement and the value the patient places on that improvement. For example, it was found in one study that a reduction of 17.5% from baseline to post-test on the BDI-II can be considered as the Minimal clinically important difference (Button et al., 2015). Currently, it is also possible to convert different measures into one common metric (e.g., Wahl et al., 2014), making the use of the effect size no longer needed.

The effect size has other problems. For example, it assumes that different outcome scales are linear transformation of each other and the standard deviation units are indeed the same across all studies (Cummings, 2011). These assumptions do not necessarily need to be true in all situations. Furthermore, the effect size may be influenced by how narrow the inclusion criteria are (Cummings, 2011). If a trial only includes participants with a narrow severity range at baseline, it can be expected that the distribution of the severity at post-test is also relatively narrow. If patients with a broader severity range are included, the distribution of severity will be broader as well. This implies that if two trials, one with a narrow severity range and one with a broad severity range, show the same absolute difference (points on a severity scale), the effect size can still vary widely, because the distribution differs across the two studies.

What Does the Effect Size Mean?

The most important problem of the effect size is, however, that it is so difficult to explain what it exactly means to non-scientists. Imagine a patient who considers to accept a treatment and asks the clinician what the chances are to get better after treatment. The clinician will have to say something like "if you get the treatment you will score 0.5 standard deviation lower on the outcome measure than not receiving the intervention".

Of course a patient has no clue for what this actually means, and many clinicians also find it hard to understand what it means.

There are some solutions to this problem. One older solution is to transform the effect size into the 'binomial effect size display' (BESD) (Rosenthal & Rubin, 1982). The BESD reduces an outcome to a simple dichotomy (for example whether a score is below or above the mean on the outcome instrument) and indicates the difference between the two treatment groups (e.g., therapy and control) in percentages of participants who score below (or above) the mean (Randolph & Edmondson, 2005). For example, an effect size of 0.2 indicates a difference of 0.10 in the proportion of participants reaching this threshold. One could say that such a value of the BESD means that 45% of the control group and 55% of the treatment group had reached the threshold of 'success'. However, this is still a relative outcome and can in no way be interpreted as if 55% of the participants will score below the mean of the outcome measure.

Another way to make the effect size easier to interpret is to transform it into the number-needed-to-treat (NNT). The NNT indicates the number of patients that have to be treated in order to have one more positive outcome than no treatment (or an alternative treatment) (Laupacis, Sackett, & Roberts, 1988). There are several ways to transform the effect size into the NNT (da Costa et al., 2012; Furukawa & Leucht, 2011), but all are based on the normal distribution of the outcome measure and a cut-off on this normal distribution for a 'positive outcome'. However, again it is not clear what this 'positive outcome' exactly is and the NNT still does not answer the question of the patient what the chances are to get better after treatment. Transforming the effect size into the NNT is, however, done by many meta-analyses to make the outcomes easier to interpret from a clinical point of view.

Moving to Binary Outcomes?

Binary outcomes are easier to understand than effect sizes. For example, in a trial the researchers can calculate the proportion of participants that respond (for example defined as a 50% reduction in symptoms from baseline to post-test) in the treatment and control group. They can also calculate the proportion of participants who recover completely (for example by scoring below a cut-off on a symptom measure), who reliably improved, or who reliably deteriorated, or dropped out from treatment. These binary outcomes can answer the question of the imaginary patient that we presented earlier very well. The patient will hear an exact chance of getting better after the treatment compared to no treatment.

For example, we recently conducted a meta-analysis of psychotherapies for depression (Cuijpers, Karyotaki, de Wit, & Ebert, 2020) and found that the effect size for psychotherapy versus control conditions was g = 0.72, 95% CI [0.67, -0.78]. That is a considerable effect according to the criteria of Cohen. But what does it really mean?

What is the chance of getting better for a patient receiving therapy compared to the chance in the control conditions? In another recent meta-analysis, we calculated the exact proportions of response (50% reduction of symptoms between baseline and post-test) (Cuijpers, Karyotaki, Ciharova, Miguel, Noma, & Furukawa, 2021) for psychotherapies with at least 10 trials for which the response rate was reported or could be estimated using a validated method (Furukawa, Cipriani, Barbui, Brambilla, & Watanabe, 2005). We found that the response rate for psychotherapies was 41% (using the most conservative estimate), while the response rate was 17% in the usual care groups. This is definitely more informative for patients and clinicians than the effect size. It shows for example that about 60% of patients do not respond after therapy and that the proportion of patients responding to usual care is really very low. The effect size gives no indication at all for such outcomes. It just says that the effect are "large", but this "hides" in a way that the majority of patients still don't respond to treatment.

Disadvantages of Binary Outcomes

So does this solve the problem? Should we all move away from the effect size and instead use binary outcomes? Unfortunately, binary outcomes also have problems. Maybe the most important problem is that outcomes may be best considered as a continuous phenomenon and not as a binary outcome. One can use binary outcomes that are informative, such as response or remission, but that does not solve the problem that in principle outcomes are still continuous. Another problem is that in individual trials binary outcomes have less statistical power to find significant differences between treatment and comparison conditions. Furthermore, there is no way to decide what the best binary outcome is. In many trials on psychological treatments the Reliable Change Index (RCI) is used (Jacobson & Truax, 1991), a psychometric criterion used to evaluate whether the change between baseline and post-test is considered statistically significant (the difference between baseline and post-test means divided by the standard error of the difference between the two scores is greater than 1.96, conservatively assuming a Cronbach's alpha of 0.75) (Jacobson & Truax, 1991). Other studies use the response (50% reduction in symptoms from baseline to post-test) or remission (scoring below a cut-off on a rating scale indicating the return to 'normal' functioning) as the main outcome. There is no way to decide what the most important binary outcome is and that may therefore vary widely across studies, making meta-analyses of these outcomes more complicated. But it also makes the answer to the question of the patient more complicated. It can be said what the chance of getting better is, but what getting better actually is, is not so clear.

Another problem with reporting the chance of getting better in the treatment and control conditions is that these chances can be very well reported for individual trials, but pooling them in meta-analyses may be problematic. The problem with exact percen-

tages is that when you pool them, the heterogeneity of the outcome is often very high. Heterogeneity indicates the variability in the outcomes of the included studies in a meta-analysis. If heterogeneity is too high that means that the outcomes are too different from each other to allow pooling. And that is typically the case when proportions are pooled. But on the other hand, these outcomes are so important for patients and clinicians, that one could make the case to pool anyway, but always say that the outcomes can vary considerably.

Usually, binary outcomes in meta-analyses are not reported in terms of absolute percentages, because of the high heterogeneity. In most cases binary outcomes are reported in terms of relative outcomes, such as the Relative Risk (*RR*) or the Odds Ratio (*OR*). The *OR* indicates the odds of getting better in the treatment group compared to the control group. This is also difficult to interpret, because it is not immediately clear what the odds are and it can be argued that the *OR* should be avoided as well because it is not clear what it means (Higgins & Green, 2011). The *RR* is easier to interpret. An *RR* of 1.40 for example indicates that the chance of getting better is 40% higher in the treatment group than in the control group. Sometimes the NNT is also used. The NNT is actually the inverse of the Risk Difference (*RD*). So if 60% get better in the treatment group and only 40% in the control group, the *RD* is 20% and the NNT is 5 (1/0.20).

But all relative outcomes do not answer the question of the patients what the chances are of getting better after the treatment. In order to answer that, it cannot be avoided to give the actual chances.

Conclusion

So should we stop using the effect size and instead move to reporting the proportions of participants who improve in the treatment and the control group? I don't think that is needed. Many studies already give the effect size and one or more binary outcomes. That is probably the best solution.

But we should avoid to obscure outcomes by just saying that a treatment is effective and the effect size is large, moderate or small. Such a statement can mean many different things. A large effect size can still indicate that many people don't get better, and a small effect size can be a major breakthrough. It is important to add in trials but also in meta-analyses what the effect sizes exactly mean in terms of relative but also absolute binary outcomes.

References

Button, K. S., Kounali, D., Thomas, L., Wiles, N. J., Peters, T. J., Welton, N. J., Ades, A. E., & Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory – II according to the patient's perspective. *Psychological Medicine, 45*, 3269-3279. https://doi.org/10.1017/S0033291715001270

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ, USA: Erlbaum.

Cuijpers, P., Karyotaki, E., Ciharova, M., Miguel, C., Noma, H., & Furukawa, T. A. (2021). The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatrica Scandinavica, 144*(3), 288-299. https://doi.org/10.1111/acps.13335

Cuijpers, P., Karyotaki, E., de Wit, L., & Ebert, D. D. (2020). The effects of fifteen evidence-supported therapies for adult depression: A meta-analytic review. *Psychotherapy Research, 30*, 279-293. https://doi.org/10.1080/10503307.2019.1649732

Cuijpers, P., Turner, E. H., Koole, S. L., van Dijke, A., & Smit, F. (2014). What is the threshold for a clinically relevant effect? The case of major depressive disorders. *Depression and Anxiety, 31*, 374-378. https://doi.org/10.1002/da.22249

Cummings, P. (2011). Arguments for and against standardized mean differences (effect sizes). *Archives of Pediatrics & Adolescent Medicine, 165*, 592-596. https://doi.org/10.1001/archpediatrics.2011.97

da Costa, B. R., Rutjes, A. W. S., Johnston, B. C., Reichenbach, S., Nüesch, E., Tonia, T., . . . Jüni, P. (2012). Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: Meta-epidemiological study. *International Journal of Epidemiology, 41*, 1445-1459. https://doi.org/10.1093/ije/dys124

Furukawa, T. A., Cipriani, A., Barbui, C., Brambilla, P., & Watanabe, N. (2005). Imputing response rates from means and standard deviations in meta-analyses. *International Clinical Psychopharmacology, 20*, 49-52.

Furukawa, T. A., & Leucht, S. (2011). How to obtain NNT from Cohen's *d*: Comparison of two methods. *PLoS One, 6*(4), Article e19070. https://doi.org/10.1371/journal.pone.0019070

PsychOpen GOLD

Higgins, J. P. T., & Green, S. (Eds.). *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0, updated March 2011). The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org

Hunt, M. (1997). *How science takes stock: The story of meta-analysis.* New York, NY, USA: Russel Sage Foundation.

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*, 12-19. https://doi.org/10.1037/0022-006X.59.1.12

Laupacis, A., Sackett, D. L., & Roberts, R. S. (1988). An assessment of clinically useful measures of the consequences of treatment. *The New England Journal of Medicine, 318*, 1728-1733. https://doi.org/10.1056/NEJM198806303182605

McGlothlin, A. E., & Lewis, R. J. (2014). Minimal clinically important difference: Defining what really matters to patients. *Journal of the American Medical Association, 312*, 1342-1343. https://doi.org/10.1001/jama.2014.13128

Randolph, J. J., & Edmondson, R. S. (2005). Using the Binomial Effect Size Display (BESD) to present the magnitude of effect sizes to the evaluation audience. *Practical Assessment, Research & Evaluation, 10*, Article 14. https://doi.org/10.7275/zqwr-mx46

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Education & Psychology, 74*, 166-169. https://doi.org/10.1037/0022-0663.74.2.166

Wahl, I., Löwe, B., Bjorner, J. B., Fischer, F., Langs, G., Voderholzerg, U., Aita, S. A., Bergemann, N., Brähler, E., & Rose, M. (2014). Standardization of depression measurement: A common metric was developed for 11 self-report depression measures. *Journal of Clinical Epidemiology, 67*, 73-86. https://doi.org/10.1016/j.jclinepi.2013.04.019