Difficulties in Conducting Long Term Follow Ups in Psychotherapy Research—Issues in the Literature and Data From a Randomized Therapy Comparison Study for Posttraumatic Stress Disorder

Julia König, PhD,*† Regina Karl, PhD,*‡ Rita Rosner, PhD,*† and Willi Butollo, PhD*§

Abstract: Studies of psychotherapy for posttraumatic stress disorder (PTSD) typically encompass short follow-up periods, leading to a dearth of information on the long-term course of symptoms after treatment. We summarize existing long-term follow-up studies and highlight the issues making such research difficult. In this context, we report on a 2-year follow-up on a randomized treatment study comparing dialogical exposure therapy and cognitive processing therapy for adults with PTSD after type I trauma. Many of the problems connected to long-term follow-up also pertained to this study. Attrition was high, and the recruited sample was not representative of the study completers. Gains made during therapy were mostly stable, with no significant differences between interventions. At least for an originally successful subset of patients, the gains made during both treatments in our study tended to stay stable 2 years after the end of treatment.

Key Words: Long term follow up, posttraumatic stress disorder, gestalt therapy, cognitive processing therapy

(J Nerv Ment Dis 2018;206: 513-521)

P osttraumatic stress disorder (PTSD) can occur after a person is exposed to a traumatic stressor, such as physical or sexual violence, but also a life-threatening accident or natural disaster. Those with PTSD typically experience intrusive symptoms, such as unwanted memories of the event or nightmares, hyperarousal and numbing (e.g., difficulties falling or staying asleep, increased startle response), and avoidance of stimuli connected with the event. Besides these three symptom clusters, which are listed in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV; American Psychiatric Association, 1994), the new edition of the DSM (American Psychiatric Association, 2013) added a fourth cluster, negative alterations in cognitions and mood. The prevalence of PTSD differs widely between countries, for example, in the United States, it is estimated at about 7% (Kessler et al., 2005), whereas numbers in Europe are lower at about 2% (Alonso et al., 2004). There are several effective psychotherapies for PTSD, with treatments from a cognitive-behavioral therapy (CBT) tradition and eye movement desensitization and reprocessing (EMDR) having garnered the most empirical support to date (Bisson and Andrew, 2007; Watts et al., 2013). Often, efficacy studies have short follow-up (FU) periods of up to 1 year. However, longer FU intervals are important: it is possible that treatments which perform equally well in the short term may have different outcomes in the long term. Treatment effects might "wear off," resulting in relapses, or treatments may start "virtuous cycles," leading to further improvements after

All the authors have left LMU since the study was completed.

Copyright © 2018 Wolters Kluwer Health, Inc. All rights reserved. ISSN: 0022-3018/18/20607-0513

DOI: 10.1097/NMD.00000000000844

the end of active therapy, so it is possible that treatments which are similarly effective in the short term have different outcomes years later.

Although the need for long-term FU studies seems to be widely recognized, such studies are scarce, even though their number has increased in recent years. This article has a threefold aim: we give an overview of long-term FU studies on PTSD samples that we are aware of, we give possible reasons for the dearth of such studies, and we present 2-year FU data from a therapy comparison study as an example of the difficulties attending long-term FUs.

Table 1 provides an overview of studies published in 2000 or later that report FU data collected 18 months or more after psychotherapy for PTSD had ended. FU studies focusing on acute stress disorder (Bryant, 2006; Bryant et al., 2003; Shalev et al., 2016), on children (Deblinger et al., 1999; Gutermann et al., 2017; Vloet et al., 2014), or on outcomes other than PTSD (Ahmadi et al., 2016; Bagley and Young, 1998; Member Centers of the National Consortium of Torture Treatment Programs [NCTTP], 2015) are not included. Neither are studies not reporting posttreatment measurements (Peleikis, 2005; Zimmermann et al., 2007).

As becomes apparent in the table, studies differ widely with respect to population, sample size, length of FU period, and type of original study. Most do not report on an untreated control group, either because there was none to begin with or because waitlist patients were offered treatment after the waiting period was over. This highlights a difficulty inherent in researching long-term therapy outcomes: although it would, from a purely methodological standpoint, be preferable to follow both intervention and control groups for a long time, it would be unethical to withhold treatment from a random subset of patients. In the large majority of studies, symptom scores remained stable compared with posttreatment. The same is true for diagnostic statuswhereas a minority of patients relapsed or remitted over the FU interval, the large majority did not change their diagnostic status. There are three notable exceptions to this pattern of stability over the FU period. In one study on EMDR treatment in veterans (Macklin et al., 2000), participants lost the very modest gains they had made during treatment and were even more symptomatic 5 years later than at baseline. They did not differ from a comparison group. This was not a randomized study; the untreated comparison sample was drawn for the long-term FU analyses from a similar population and not recruited at the same time as the treated patients. A second study examining Iranian and Afghan refugees/asylum seekers in the Netherlands after an intensive day hospital treatment lasting 1 year (Drožáck et al., 2014) found that both the more recent and earlier cohorts, treated 1 and 11 years previously, respectively, reported more symptoms, whereas symptom load was lowest in the cohort assessed 5 years after the end of treatment. When interpreting these findings, it has to be kept in mind that each cohort was only followed up once, so it cannot be assumed that all cohorts had the same symptom trajectory. There may have been other influences at play, such as differences in the way the program was implemented, but also changes in the political climate for refugees, or changes in participants' home countries, all of which might influence

^{*}Psychology Department of the Ludwig Maximilian University (LMU), Munich; †Catholic University of Eichstaett-Ingolstadt, Eichstaett; ‡CIP Clinic; and §Munich Institute for Trauma Therapy MIT, Munich, Germany.

[§]Munich Institute for Trauma Therapy MIT, Munich, Germany.Send reprint requests to Julia König, PhD, Catholic University of Eichstaett-Ingolstadt, Ostenstr. 25, 85072 Eichstaett, Germany. E-mail: julia.koenig@ku.de.

TABLE 1. PTSD Treat	ment Studies With !	FUs of 18 Months a	and More				
Authors	Population (trauma type ^a)	Mean FU Time in Years	n (%) of Sample Retained ^b	Treatment(s)	Design ^c (Control Condition)	Results at Posttreatment ^d	Long-term FU Results ^d
Macklin et al., 2000	Veterans (combat)	S	13 (76)	EMDR	Uncontrolled	Modest decreases in symptoms	Gains made during therapy were lost; no significant differences between treated patients and an untreated comparison group
Blanchard et al., 2004	Civilians (MVAs)	7	39 (68)	CBT, supportive therapy	Randomized	Improvement in both groups, CBT superior	Symptoms largely stable from 3 mo to 2 yr FU, CBT continued to be superior on PTSD measures
Johnson et al., 2004	Vietnam veterans (combat)	9	47 (92)	Inpatient program	Uncontrolled	No significant change in PTSD symptoms	No change from posttreatment to FU
Tarrier and Sommerfield, 2004	Civilians (different types)	Ś	32 (52)	Cognitive therapy, imaginal exposure	Randomized, uncontrolled	Both groups improved similarly and significantly	Treatment effects largely stable, cognitive therapy superior to imaginal exposure
Durham et al., 2005	Civilians (different types)	3.8	34 (38)	EMDR, CBT	Randomized, controlled (WL)	Both treatment groups significantly improved, EMDR slightly better, WL unchanged	No difference between active treatment groups on PTSD symptoms
Creamer et al., 2006	Veterans (combat)	1.75	1508 (68)	Different programs	Uncontrolled	Small but significant improvements	Treatment effects were stable, trend toward further improvement
Högberg et al., 2008	Civilians (occupational)	2.9	17 (85)	EMDR	Uncontrolled	Significant improvement	Results largely stable from posttreatment to FU
Knaevelsrud and Maercker, 2010	Civilians	1.5	34 (83)	Internet-based CBT	Uncontrolled	Significant improvement (large effects)	Treatment gains stable, no significant change from posttreatment
Resick et al., 2012	Female adults (rape)	6	126 (74°)	CPT, PE	Randomized, controlled (WL)	Significant improvement (large effects) in both treatment groups	Treatment gains stable, trend: more further improvement in PE on one PTSD measure
Mithoefer et al., 2013	Treatment-resistant PTSD patients	3.8	16 (84)	MDMA-assisted psychotherapy	Randomized, controlled (placebo)	MDMA significantly more successful, large effect size	No significant change from posttreatment
Drožáck et al., 2014	Refugees (torture)	٦ ^۴	69 (77)	1 Yr of multimodal group therapy	Uncontrolled	Significant decrease in symptoms	Symptom scores continued to decrease up to 5 yr after treatment, then increased, but remained below baseline
Earley et al., 2014	Adults (CSA)	2.3	19 (83)	MBSR	Uncontrolled	Significant decrease in symptoms	Slight increase from posttreatment, but difference to baseline remained significant

514 www.jonmd.com

Lampe et al., 2014	Female adults (CSA)	7	43 (53)	Psychodynamic inpatient treatment	Controlled (WL)	No differences between groups in PTSD symptoms	Improvements with moderate effect sizes (pretreatment to LTFU) were found, but might be due to other factors ^g
CSA, child sexual abus wait list.	se; LTFU, long-term follow-u	p; MVA, motor vehicle	e accident; PE, pro	longed exposure; MDM	A, 3,4-methylenedioxym	ethamphetamine; MBSR, mind	fulness-based stress reduction; WL,
^a If no trauma type is g	iven, the study included patie	ents with various types	s of trauma.				
^b Percentages refer to th	he completer sample if not ot	therwise specified.					
^c The distinction betwe	sen controlled and uncontrol	led designs pertains or	nly to the primary	r study. Generally, the c	ontrol group was offered	treatment later on, so long-te	arm FU results are all within-group
comparisons.							
^d Results given here per	rtain to PTSD symptoms onl	b; results for comorbic	diagnoses or rela	ted problems may differ			
^e Of intent-to-treat sam	ple.						
^f Different FU intervals	represent different cohorts o	of the 1-year-long prog	ram. Intervals rang	ged up to 11 years.			
^g 97% of the FU sampl.	e received further treatment i	in the FU interval. Als	o, the FU sample	was, on a trend level, m	ore likely to have received	I prior treatment at the same h	ospital than FU nonparticipants.

participants' mental health. A third large study from the Australian veterans administration system (Creamer et al., 2006) found modest further improvement during the FU period after a range of diverse initial treatments, but as information on possible further treatment was lacking, this cannot clearly be attributed to those initial treatments.

It is interesting that in the studies where treatment during the FU interval was examined, patients receiving such treatment tended to have worse outcomes (Blanchard et al., 2004; Resick et al., 2012). This is probably because patients are more likely to seek further treatment if they have not profited from the initial course of treatment.

In one treatment comparison study (Tarrier and Sommerfield, 2004), there had been no significant differences between cognitive therapy and imaginal exposure at 12 months posttreatment. However, at the 5-year FU, a significant advantage of cognitive therapy emerged. This result underscores the importance of long-term FU studies. Most studies suffer from small sample sizes, high attrition, or study diverse or integrative treatments.

One reason for the dearth of long-term FUs is the fact that funding periods for psychotherapy studies are generally too short to include them. A possible solution would be to apply for grant money specifically for a FU. This was done for a study on cognitive processing therapy (CPT) and prolonged exposure (Resick et al., 2012) with an average FU period of 6 years and very good retention of participants. It is encouraging that this methodologically very sound study found that the gains made during treatment were stable years later. There was no significant difference in PTSD scores between the two treatment groups, but a slightly higher rate of relapse in the CPT group, which approached statistical significance. This result seems to suggest a slight advantage of exposure work over cognitive work and is in contrast to the results reported by Tarrier and Sommerfield (2004), where cognitive therapy was superior to exposure 5 years after treatment.

To our knowledge, FU data from humanistic approaches to PTSD therapy have not been published so far. We recently compared one such therapy, dialogical exposure therapy (DET) with CPT in a randomized trial including adults with PTSD (Butollo et al., 2016). DET is an integrative treatment with a foundation in gestalt therapy. The self is seen as dialogical, as a dynamic cognitive-emotional state that constantly constitutes itself as a result of interactional experience, both with oneself and with the environment. Therefore, the term "self-processes" is preferred. In PTSD, a traumatic experience overpowers the ability of the self-processes to respond to the situation, causing a disruption in a person's inner dialogues or a breaking-off of contact between different self-processes or aspects of the self. The posttraumatic self, with the experience of worthlessness and submission, becomes dominant, and the contact to pretraumatic self-processes, such as self-worth, is compromised. The goal of therapy is to help reestablish this contact, so the patient can again experience continuity in his or her self-processes and shape interactions with the environment in a productive way. One important technique to help reestablish this contact is empty chair work (Paivio and Greenberg, 1995), different variants of which can be used throughout the treatment. In this form of work, patients are encouraged to interact directly with an imagined interaction partner to experience and express their feelings toward them.

CPT is a predominantly cognitive therapy, even though we used the form including an exposure component, which is now called CPT-A (including trauma account, Resick et al., 2016). Traumatic events are seen as usually discrepant with preexisting cognitive schemas, or beliefs about the world, oneself, and other people. To integrate the event and the schemas, it would be necessary to change the latter to adapt the schema to the new reality. This mechanism is called accommodation. However, patients with PTSD frequently change their schemas too much, resulting in extremely negative and rigid beliefs. This is called overaccommodation. At the same time, the memory or the meaning of the traumatic event is often distorted to fit the existing schemas. This mechanism, called assimilation, often results in self-blame, feelings of guilt, or difficulties accepting the event as having happened (mental undoing). The main focus of CPT is therefore on challenging and changing dysfunctional cognitions caused by assimilation or overaccommodation.

There were no significant differences in outcome between the two treatments at posttreatment or 6 months later. However, given the differences in both underlying theory and the activities constituting the treatments, it is possible that treatments differ in the long term. In this article, we therefore analyze 2-year FU data from a part of this sample. This is the first study examining the long-term course of patients treated with DET, but in light of the existing literature, we hypothesize treatment results in both therapies to be stable. As a secondary aim, we plan to explore differences between the two treatments.

METHODS

Participants

We enrolled consecutive patients who had a primary diagnosis of PTSD after a type I trauma in adulthood and sought treatment in a university-based outpatient clinic. We excluded patients reporting ongoing distress from psychosis or substance dependence or if their symptoms were due to childhood abuse or neglect. Even so, both PTSD severity and general psychological distress were comparable to other therapy study samples (Butollo et al., 2016). The original intent-totreat (ITT) sample consisted of 141 patients, and the completer sample of 122 patients. Here, we report on the 2-year FU assessment. The study protocol was approved by the institution's ethics committee. The longterm FU was not explicitly part of the original study protocol, but patients provided informed consent to be contacted about FU diagnostics before study inclusion.

Measures

PTSD at study intake was assessed with the German version of the International Diagnostic Checklists for *DSM-IV* and *ICD* (IDCL; Hiller et al., 1996), and comorbid *DSM-IV* axis I diagnoses were established using the DIA-X (Wittchen and Pfister, 1997), a computer-assisted version of the Composite International Diagnostic Interview with an excellent reliability (93.3% agreement, $\kappa = 0.76$; Wittchen et al., 1998). The IDCL has also been demonstrated to be reliable, with agreements between $\kappa = 0.67$ and $\kappa = 0.88$ for different interviewers (Hiller et al., 1996).

Both the 6-month and the 2-year FUs consisted of a qualitative interview focusing on the therapy experience and current functioning as well as a questionnaire battery containing most of the instruments used at pre- and posttreatment. The questionnaires were mailed to the patients, and they were asked to return them during the interview. In this article, we report on the quantitative data collected at the 2-year FU and use only the information on further treatment and further traumatic experiences from the interview.

The Impact of Event Scale–Revised (IES-R; Maercker and Schützwohl, 1998), a self-report instrument that measures the intensity of PTSD symptoms, was completed by patients before each session as well as at the assessments. It contains three subscales measuring the symptom clusters of intrusion (Cronbach's $\alpha = .90$), avoidance ($\alpha = .79$), and hyperarousal ($\alpha = .90$). The factor structure of the German translation of the IES-R is comparable to the original English version (Maercker and Schützwohl, 1998).

As a second primary outcome measure that was only used during the assessments, we employed the German version of the Posttraumatic Diagnostic Scale (PDS; Foa et al., 1997), a self-report instrument that can be used for determining PTSD diagnostic status according to *DSM-IV* criteria and symptom severity. Its psychometric properties are good: Griesel et al. (2006) report, with Cronbach's $\alpha = .90$, a high internal consistency for the German translation of the scale and good convergent and discriminant validities, although it has a tendency to overdiagnose. To obtain a diagnosis, typically items, which correspond to the *DSM-IV* (American Psychiatric Association, 1994) symptom criteria, are scored as present or not present and the diagnosis given accordingly, that is, if enough criteria for each symptom cluster are met. However, in a study comparing different scoring methods, the best results were found when a severity cutoff of 18 was used in addition to symptom cluster scoring (Ehring et al., 2007). Therefore, we considered PTSD as present when symptom criteria were met and the severity score was 18 or higher.

We used the global severity index (GSI) of the Brief Symptom Inventory (BSI; Derogatis and Melisaratos, 1983) to measure general psychological distress. The German version of the BSI (Franke, 2002) has good psychometric properties: Geisheim et al. (2002) report a very good internal consistency with Cronbach's $\alpha = .96$ of the GSI and report significant correlations in the moderate to high range with established self-report instruments for different disorders. The GSI was also shown to be sensitive to change.

Interpersonal distress was measured with the inventory of interpersonal problems, circumplex version (IIP-C; Horowitz et al., 1994). It consists of 64 items describing behavior that respondents can have difficulties with or that they may do too much in their interpersonal encounters. The measure has good psychometric properties is sensitive to change, and has been used in psychotherapy research both as an outcome measure and as a predictor of outcome. Horowitz et al. (1988) report high internal consistencies with Cronbach's α values between .82 and .94 for the different subscales and a retest reliability of .98 for the total scale. They are also able to demonstrate that the total score is sensitive to therapeutic change.

Treatments

DET (Butollo and Karl, 2012) is an integrative treatment based on gestalt theory. It has similarities to emotion-focused therapy (Greenberg, 2002; Paivio and Greenberg, 1995; Paivio et al., 2010) and was developed specifically to treat PTSD. It is structured into four phases: safety, stability, confrontation, and integration. The phases overlap, and they can have different emphases and durations depending on the type of trauma and the individual patient's progress. The overall goal of DET is to help patients identify and resolve disruptions in their selfprocesses, which have been caused by the traumatic event and the person's reactions to it. The core component of the therapy is a modified form of empty chair work (Paivio and Greenberg, 1995), where the patient is invited to enter into a direct dialogue with the imagined perpetrator. DET has previously been tested in a group format with Bosnian women with traumatic bereavement (Hagl et al., 2015) and a pilot study (Butollo et al., 2014). More information on DET can be found in our original report (Butollo et al., 2016).

CPT (Resick et al., 2007, 2016) is one of the CBTs for PTSD with the best evidence base (Watts et al., 2013). It is a highly structured, predominantly cognitive therapy protocol targeting dysfunctional cognitions about the traumatic event, oneself, and others. In this study, we used a translated and slightly modified form of CPT (König et al., 2012). In addition to the written exposure component, we included behavioral homework to help patients counteract behavioral avoidance.

Maximum treatment length in both conditions was 24 sessions with an option of up to two "booster sessions" in the 6 months after the end of treatment. As 25 sessions are considered short-term CBT in the German compulsory health care system, we decided to offer the amount of therapy patients would get elsewhere. Typically, sessions were scheduled weekly. Treatment length was flexible however: if a patient and therapist and the supervisor agreed that the therapy was completed, treatment was terminated and the patient still considered a treatment completer.

Therapists

There were two distinct groups of therapists conducting the two treatments. All of the eight DET therapists were master's-level psychologists with clinical experience ranging from 1 to 15 years and ages ranging from 28 to 55 years. One was male at the beginning of the study, two of them were receiving and two had completed training in gestalt therapy, and five were in training to become licensed CBT therapists. Three completed this training during the study. Before conducting treatments in the study, all DET therapists completed a 96-hour curriculum on treating traumatized patients, of which DET was the main focus.

CPT treatment was offered by seven female master's-level psychologists aged 27 to 45 years and with clinical experience of 1 to 10 years at study entry. They were all in training for their CBT licenses, and four completed it during the study. Three therapists completed a 2-day workshop on CPT and a supervision session with Patricia A. Resick. The other four therapists joined the study later and were trained individually. Training included reviewing the treatment manual session by session with the first author, attending supervision several times before taking cases, and treating one closely supervised pilot case, which was not included in analyses.

Therapists in both conditions received supervision on average after every fourth session and between supervision sessions met weekly for case discussions. Supervision was offered in a group format by the third author for CPT and by the fourth author for DET.

Data Analysis

As only treatment completers were asked to participate in the FU, we used the completer sample for all analyses if not otherwise specified, replacing missing data with the last existing data point (last observation carried forward [LOCF]). We considered using imputation methods, but given the very high attrition, felt we could not do so with confidence. Comparisons between subgroups were done with *t*-tests and chi-square tests where appropriate. When comparing very small subgroups, however, such as patients with and without further trauma, we used nonparametric tests.

To determine whether the symptomatic change reported by individual patients was larger than what could be explained by measurement error, we calculated critical differences for each of the four instruments using the formulas provided by Jacobson and Truax (1991). We used the formulas for reliable change and for the standard error of difference between two test scores and substituted internal consistency for the retest reliability (Martinovich et al., 1996). Patients who had achieved a symptom reduction larger than the critical difference were considered reliably improved. Conversely, if there was an increase in scores larger than the critical difference, they were considered reliably worsened. If a positive or negative change in scores was smaller than the critical difference, patients were classed as not reliably changed.

RESULTS

Characteristics of the FU Sample

Unfortunately, because of time and staff constraints, we were not able to contact the full completer sample about participation in the FU. There were three different periods, during which FUs were conducted with those patients who were approximately 2 years posttreatment at the time. We attempted to contact 92 patients or 75.4% of treatment completers but were able to obtain any data from only 39 or 42% of those we attempted to include. As no consistent records were kept about the unsuccessful attempts, we cannot tell how many of the remaining patients declined participation, could not be located, or did not show for their scheduled appointments. According to the staff member responsible for contacting former patients, the reasons given most frequently for declining to participate were time constraints or living too far away, but also unwillingness to answer questions after such a long time about something the former patients considered to be over. The staff member was not able to give an estimate how frequently each reason was given. Three of the participants only took part in the interview portion, leaving us with questionnaire data from 36 patients (39% of those attempted to include, 30% of treatment completers). Of these, 17 had received DET and 19 had received CPT, which is not a significant difference, $\chi^2(1) = 0.479$, p = 0.489. Although the aim was to assess patients 24 months after the end of therapy, there was some variation (mean = 27.55, SD = 7.04).

Information on further traumatic events and further treatment was collected in the qualitative interview. Of the 39 patients with any data, 6 refused to be interviewed. Of the remaining 33 patients, 3 from each intervention group reported having had additional treatment. One CPT patient was in psychotherapy because of the same traumatic event; one CPT patient because she had experienced another traumatic event after the end of therapy, and two patients from the DET and one from the CPT groups had visited mental health professionals for reasons other than PTSD symptoms, namely, diagnostic brain scan, problems with procrastination, and treatment for chronic pain. One DET patient reported receiving counseling from an organization supporting crime victims. Three additional patients, all of whom had received CPT, were planning to seek further psychotherapy, one because of ongoing distress from the same trauma, two for unrelated reasons.

Five DET and three CPT patients or 22% of the FU sample, reported further traumatic experiences in the intervening 2 years. In one CPT case, who was still in treatment, the patient reported further surgeries following from the original trauma as traumatic in themselves. None of the other six patients, who all had experienced losses or witnessed accidental deaths, were planning to seek further treatment.

Among the FU participants, we compared those who had experienced further trauma with those who had not as well as those who had received further treatment with those who had not on the four

	Further Traumatic Event(s)					Further Treatment			
	Yes (n = 7) Median	No (<i>n</i> = 28) Median	Mann–Whitney U	p^{a}	Yes (n = 5) Median	No (<i>n</i> = 25) Median	Mann–Whitney U	p ^a	
IES-R	4.00	12.00	84.00	0.578	18.00	8.00	39.00	0.202	
PDS	7.00	7.50	92.50	0.832	9.00	7.00	38.50	0.192	
BSI	0.30	0.34	93.00	0.959	0.54	0.30	42.00	0.269	
IIP	1.11	0.85	94.50	>0.999	0.98	0.86	60.50	0.924	

TABLE 2. Comparisons of Symptoms at the 2-Year FU Between Participants With and Without Further Trauma and Therapy Respectively

^aExact significance.

	2-Yr I	FU			
	Nonparticipants ($n \le 83$)	Participants ($n \leq 39$)		Test for difference	:
	n (%)	n (%)	df	χ^2	р
Female sex	55 (66.3)	28 (71.8)	1	0.373	0.541
Marital status			4	4.025	0.402
Single	23 (31.5)	6 (21.4)			
In relationship	18 (24.7)	12 (42.9)			
Married	20 (27.4)	6 (21.4)			
Divorced/separated	11 (15.1)	3 (10.7)			
Widowed	1 (1.4)	1 (3.6)			
	Mean (SD)	Mean (SD)	t	df	р
Age at beginning of therapy	35.31 (11.16)	37.14 (12.45)	-0.794	120	0.429
Time of beginning therapy	10-2009 (270d)	12-2009 (290d)	-1.27	120	0.205
Duration of therapy (mo)	6.48 (2.72)	6.14 (2.93)	0.600	119	0.550
No. therapy sessions	22.21 (3.88)	22.40 (4.43)	-0.238	120	0.812
No. booster sessions	0.86 (0.81)	0.87 (0.78)	-0.066	104	0.948
IES-R pre	70.55 (16.55)	61.83 (18.10)	2.591	113	0.011*
IES-R post	31.93 (30.14)	22.15 (25.27)	1.749	118	0.083
IES-R 6-mo FU	31.02 (28.42)	20.74 (22.46)	2.148	92.931	0.034*
PDS pre	30.85 (9.87)	26.64 (9.86)	2.184	75.564	0.032*
PDS post	17.70 (13.46)	11.67 (10.72)	2.655	92.076	0.009*
PDS 6-mo FU	16.59 (13.59)	10.74 (10.63)	2.574	93.468	0.012*
BSI GSI pre	1.54 (0.72)	1.30 (0.69)	1.755	116	0.082
BSI GSI post	0.91 (0.83)	0.58 (0.58)	2.561	102.795	0.012*
BSI GSI 6-mo FU	0.86 (0.77)	0.56 (0.64)	2.285	88.784	0.025*
IIP pre	1.43 (0.57)	1.31 (0.55)	1.034	113	0.303
IIP post	1.12 (0.66)	1.09 (0.66)	0.252	117	0.802
IIP 6-mo FU	1.09 (0.63)	0.93 (0.61)	1.241	118	0.217

TABLE 3. Comparisons Between the Completers With and Without Participation in the 2-Year FU (Completer Sample, LOCF)

Pre, pretreatment; post, posttreatment.

*p < .05.

outcome variables using Mann-Whitney U tests. None of the differences were significant, as shown in Table 2.

As can be seen in Table 3, 2-year FU participants did not differ significantly from nonparticipating treatment completers in terms of

demographic characteristics, time of beginning therapy, duration of therapy, and number of therapy sessions. However, they had been less symptomatic at the three previous assessments, unfortunately making our FU sample unrepresentative even of treatment completers.

|--|

Sample	Measure	Pre	Post	6-mo FU	2-yr FU
Completer sample ($n \le 122$)	IES-R	67.59 (17.51)	28.75 (28.94)	27.68 (26.97)	26.42 (16.63)
	PDS	29.47 (10.03)	15.75 (12.91)	14.70 (12.96)	14.08 (12.80)
	BSI GSI	1.46 (0.69)	0.80 (0.77)	0.76 (0.74)	0.75 (0.72)
	IIP	1.39 (0.57)	1.11 (0.66)	1.04 (0.62)	1.04 (0.62)
DET $(n \le 65)$	IES-R	69.16 (16.09)	36.35 (31.89)	32.31 (29.48)	30.18 (29.26)
	PDS	29.92 (9.98)	17.22 (13.73)	15.06 (13.24)	14.08 (13.06)
	BSI GSI	1.46 (0.67)	0.89 (0.87)	0.79 (0.79)	0.78 (0.76)
	IIP	1.38 (0.56)	1.11 (0.66)	0.98 (0.62)	0.98 (0.62)
CPT ($n \leq 57$)	IES-R	65.75 (19.08)	20.07 (22.42)	22.38 (22.91)	22.13 (22.78)
	PDS	28.96 (10.14)	14.11 (11.83)	14.30 (12.74)	14.09 (12.61)
	BSI GSI	1.46 (0.72)	0.70 (0.63)	0.72 (0.69)	0.72 (0.68)
	IIP	1.41 (0.58)	1.12 (0.66)	1.10 (0.62)	1.11 (0.63)

Pre, pretreatment; post, posttreatment; FU, follow-up.

		Pre-Posttreat	ment	Pr	etreatment-6-1	no FU	Pi	retreatment-2-	yr FU
	DET	СРТ	Between ^a	DET	СРТ	Between	DET	СРТ	Between
IES-R	1.29	2.19	0.73	1.55	2.05	0.37	1.64	2.07	0.26
PDS	0.98	1.35	0.21	1.19	1.27	0.02	1.28	1.30	0.10
BSI GSI	0.73	1.12	0.27	0.91	1.05	0.10	0.95	1.06	0.09
IIP	0.44	0.47	0.04	0.68	0.52	-0.16	0.68	0.50	-0.18

TABLE 5. Effect Sizes Within and Between Intervention Groups in the Completer Sample Using LOCF

FU, follow-up.

^aBetween-treatment effect sizes were calculated with an online calculator (Lenhard and Lenhard, 2016) according to a procedure detailed by Morris (2008). Negative values mean an advantage for DET.

Change Over Time

Scores at all measurement points are given in Table 4. These data show that scores remained stable between the FU measurements in both therapy conditions. As can be seen in Table 5, effect sizes for CPT decreased slightly in the first 6 months but were stable at the 2 years, whereas those for DET tended to increase from pre-post to pre-6 months and then stay stable. Between-treatment effect sizes at the 2-year FU were accordingly generally very small. The largest were for interpersonal problems, where DET patients fared slightly better, and for IES-R, where CPT patients had a small advantage. However, effect sizes are merely descriptive and do not imply that effects are significant.

To examine whether scores changed significantly between the two FU measurements, we only used questionnaire data from the FU sample (n = 36), as in this case, where stability can be seen as a positive outcome, using LOCF would have meant a bias. We conducted separate 2 (time points) × 2 (interventions) repeated-measures analyses of variance for the four measures. Age was included as a covariate because, as in our original sample, CPT patients in the FU sample (mean = 33.26, SD = 10.20) were significantly younger than DET patients (mean = 41.59, SD = 13.17, t[34] = 2.133, p = 0.040). No main or interaction effects including time or intervention group approached significance on any measure, indicating that there was no significant change between the two FU assessments and no differential change or different symptom load between the two interventions.

Results for change from pretreatment to the 2-year FU in the completer sample are given in Table 6. According to the PDS, 47 or 72% of DET patients and 40 or 70% of CPT patients in the FU sample lost their PTSD diagnosis. The difference between the therapies was not significant, $\chi^2(1) = 0.067$, p = 0.795.

DISCUSSION

As expected, the gains made during PTSD treatment were largely stable 2 years later. There was no significant symptomatic change on a

group level between the two FU measurements. On an individual level, the majority of patients showed no reliable change between 3 months and 2 years. Only few patients deteriorated, and none on more than one measure, some showed further improvement. This stability of results is in accordance with previous results for CPT (Resick et al., 2012) and indicates that, like other PTSD treatments, DET can produce lasting symptomatic change. There were no statistically significant differences between intervention groups in the completer sample at the 2-year FU. A small number of patients had experienced further traumatic events in the FU interval. This group did not differ in their symptoms from those patients with no further treatment after the end of the study therapy. Again, this group was not significantly more or less symptomatic than those who had not. Therefore, our results were not due to treatments outside the study.

Between-therapy effect sizes at the 2-year FU were generally small. There was a nonsignificant advantage for CPT on one PTSD measure and for DET on interpersonal problems. The latter can probably be explained by the fact that DET puts more of an emphasis on interpersonal relationships, which includes the therapeutic relationship, but also patients' ability to navigate important relationships in their lives. It would require a larger sample to determine whether CPT is indeed slightly more successful in reducing symptoms of PTSD as we lacked the power to detect a small effect.

It has to be borne in mind that these results were attained from a much reduced study sample and therefore have to be interpreted with great caution. We are able to report data from only about one third of treatment completers. Even taking into account that we only attempted to contact 75% of completers, this is a very high attrition as compared with other studies (see Table 1). Furthermore, patients participating in the 2-year FU had been significantly less symptomatic at the three prior measurements than nonparticipants, making our study sample unrepresentative. This can probably be explained by the fact that patients did not receive remuneration for participation in assessments, resulting in

Outcome	IES	S-R ^a	PE)S ^b	BSI	GSI ^c
	DET	СРТ	DET	СРТ	DET	СРТ
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Worsening	_	—	—	1 (1.8)	2 (3.2)	-
No change	19 (30.6)	9 (17.0)	21 (33.3)	19 (33.9)	27 (43.5)	22 (39.3)
Improvement	43 (69.4)	44 (83.0)	42 (66.7)	36 (64.3)	33 (53.2)	34 (60.7)

^cDifference not significant, $\chi^2(2) = 2.226$, p = 0.335.

more successful patients-who would probably be more satisfied with their treatment-being more likely to provide data than less successful patients who may have been disappointed. The fact that we only included treatment completers in the sample is also not in keeping with current standards, which would require trying to follow up all randomized patients. Another limitation is the fact that two of our original outcome measures, concerning posttraumatic cognitions (PTCI; Foa et al., 1999) and life satisfaction, were not included in the FU test battery. It would have been interesting to see the change especially in the PTCI, where there seemed to be an advantage for CPT at the 6-month assessment. In addition, as an FU analysis, this set of analyses shares the weaknesses of the original design: none of the posttreatment assessment included a diagnostic interview to establish the PTSD diagnosis. Further limitations, such as the lack of an untreated control group and of a measure of depression severity, and the rather strict exclusion criteria, carry more weight in the pre-post analyses than the ones presented here, but they remain limitations nonetheless.

Even with these limitations, our results add to the long-term FU literature. We compared two treatments with very different previous evidence levels. CPT has been studied extensively and has been the subject of a large long-term FU study. Our results that symptom scores decreased between pre- and posttreatment and then stayed stable fit well with previous findings, but do not add much to them. However, ours was not only the first study to examine DET in a randomized design, but, as far as we are aware, we also present the first long-term FU data on any PTSD treatment with a gestalt background. It is possible that results may generalize to other treatments based on gestalt principles and using similar techniques. We find it important to have gestalt and other humanistic therapies included in the canon of treatments studied in rigorous designs so as not to limit research to one or two approaches alone.

CONCLUSION

The analyses presented in this article unfortunately illustrate many of the problems of the long-term FU literature on PTSD, but, as this was the first large study on DET, even information about this limited sample furthers our knowledge. Our results indicate that, at least for an originally successful subset of patients, the gains made during DET, as well as CPT, tend to stay stable a good 2 years after the end of treatment.

DISCLOSURE

This article describes an unfunded study that was carried out by the authors and clinic staff as part of their regular workload at Ludwig Maximilian University in Munich, Germany. Study therapies were paid for by the patients' health insurance, and no other outside funding was received.

The authors declare no conflict of interest.

REFERENCES

- Ahmadi N, Moss L, Simon E, Nemeroff CB, Atre-Vaidya N (2016) Efficacy and longterm clinical outcome of comorbid posttraumatic stress disorder and major depressive disorder after electroconvulsive therapy. *Depress Anxiety*. 33:640–647.
- Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, de Girolamo G, Graaf R, Demyttenaere K, Gasquet I, Haro JM, Katz SJ, Kessler RC, Kovess V, Lépine JP, Ormel J, Polidori G, Russo LJ, Vilagut G, Almansa J, Arbabzadeh-Bouchez S, Autonell J, Bernal M, Buist-Bouwman MA, Codony M, Domingo-Salvany A, Ferrer M, Joo SS, Martínez-Alonso M, Matschinger H, Mazzi F, Morgan Z, Morosini P, Palacín C, Romera B, Taub N, Vollebergh WA; ESEMeD/MHEDEA 2000 Investigators, European Study of the Epidemiology of Mental Disorders (ESEMeD) Project (2004) Prevalence of mental disorders in Europe: Results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta Psychiatr Scand Suppl. 21–27.
- American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders (4th ed). Washington, DC: Author.

- American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders: DSM-5 (5th ed). Washington, DC: Author.
- Bagley C, Young L (1998) Long-term evaluation of group counselling for women with a history of child sexual abuse: Focus on depression, self-esteem, suicidal behaviors and social support. *Social Work with Groups*. 21:63–73. doi:10.1300/ J009v21n03_06.
- Bisson J, Andrew M (2007) Psychological treatment of post-traumatic stress disorder (PTSD). Cochrane Database Syst Rev. CD003388.
- Blanchard EB, Hickling EJ, Malta LS, Freidenberg BM, Canna MA, Kuhn E, Sykes MA, Galovski TA (2004) One- and two-year prospective follow-up of cognitive behavior therapy or supportive psychotherapy. *Behav Res Ther.* 42:745–759.
- Bryant RA (2006) Hypnotherapy and cognitive behaviour therapy of acute stress disorder: A 3-year follow-up. *Behav Res Ther.* 44:1331–1335.
- Bryant RA, Moulds ML, Nixon RV (2003) Cognitive behaviour therapy of acute stress disorder: A four-year follow-up. *Behav Res Ther*. 41:489–494.
- Butollo W, Karl R (2012) Dialogische Traumatherapie. Ein Manual zur Behandlung der posttraumatischen Belastungsstörung (Dialogical Exposure Therapy. A Treatment Manual for Posttraumatic Stress Disorder). Stuttgart, Germany: Klett-Cotta.
- Butollo W, Karl R, König J, Rosner R (2016) A randomized controlled trial of dialogical exposure therapy versus cognitive processing therapy for adult outpatients suffering from PTSD after type I trauma in adulthood. *Psychother Psychosom.* 85: 16–26.
- Butollo W, König J, Karl R, Henkel C, Rosner R (2014) Feasibility and outcome of dialogical exposure therapy for posttraumatic stress disorder. *Psychother Res.* 24: 514–521.
- Creamer M, Elliott P, Forbes D, Biddle D, Hawthorne G (2006) Treatment for combatrelated posttraumatic stress disorder: Two-year follow-up. J Trauma Stress. 19: 675–685.
- Deblinger E, Steer RA, Lippmann J (1999) Two-year follow-up study of cognitive behavioral therapy for sexually abused children suffering post-traumatic stress symptoms. *Child Abuse Negl.* 23:1371–1378.
- Derogatis LR, Melisaratos N (1983) The Brief Symptom Inventory: An introductory report. Psychol Med. 13:595–605.
- Drožáck B, Kamperman AM, Tol WA, Knipscheer JW, Kleber RJ (2014) Seven-year follow-up study of symptoms in asylum seekers and refugees with PTSD treated with trauma-focused groups. J Clin Psychol. 70:376–387.
- Durham RC, Chambers JA, Power KG, Sharp DM, Macdonald RR, Major KA, Dow MGT, Gumley AI (2005) Long-term outcome of cognitive behavioural therapy clinical trials in central Scotland. *Health Technology Assessment*. 9:1–174.
- Earley MD, Chesney MA, Frye J, Greene PA, Berman B, Kimbrough E (2014) Mindfulness intervention for child abuse survivors: A 2.5-year follow-up. *J Clin Psychol.* 70:933–941.
- Ehring T, Kleim B, Clark DM, Foa EB, Ehlers A (2007) Screening for posttraumatic stress disorder: What combination of symptoms predicts best? J Nerv Ment Dis. 195:1004–1012.
- Foa EB, Cashman L, Jaycox L, Perry K (1997) The validation of a self-report measure of posttraumatic stress disorder: The Posttraumatic Diagnostic Scale. *Psychol Assess.* 9:446–451.
- Foa EB, Ehlers A, Clark DM, Tolin DF, Orsillo SM (1999) The Posttraumatic Cognitions Inventory (PTCI): Development and validation. *Psychol Assess.* 11:303–314.
- Franke GH (2002) Brief Symptom Inventory von L. R. Derogatis (Kurzform der SCL-90-R)–Deutsche Version. Göttingen: Beltz Test GmbH.
- Geisheim C, Hahlweg K, Fiegenbaum W, Frank M, Schröder B, Witzleben I. von (2002) Das Brief Symptom Inventory (BSI) als Instrument zur Qualitätssicherung in der Psychotherapie. [The German version of the Brief Symptom Inventory (BSI): Reliability and validity in a sample of outpatient psychotherapy patients.]. *Diagnostica*. 48:28–36.
- Greenberg LS (2002) Emotion-focused therapy: Coaching clients to work through their feelings. Washington, DC: American Psychological Association.
- Griesel D, Wessa M, Flor H (2006) Psychometric qualities of the German version of the Posttraumatic Diagnostic Scale (PTDS). *Psychol Assess.* 18:262–268.

520 www.jonmd.com

- Gutermann J, Schwartzkopff L, Steil R (2017) Meta-analysis of the long-term treatment effects of psychological interventions in youth with PTSD symptoms. *Clin Child Fam Psychol Rev.* 20:422–434.
- Hagl M, Powell M, Rosner R, Butollo W (2015) Dialogical exposure with traumatically bereaved Bosnian women: findings from a controlled trial. *Clin Psychol Psychother*. 22:604–618.
- Hiller W, Zaudig M, Mombour W (1996) *IDCL Internationale Diagnose Checklisten für DSM-IV und ICD-10.* [IDCL International diagnostic check lists for *DSM-IV* and *ICD-10.*]. Göttingen: Hogrefe.
- Högberg G, Pagani M, Sundin O, Soares J, Aberg-Wistedt A, Tärnell B, Hällström T (2008) Treatment of post-traumatic stress disorder with eye movement desensitization and reprocessing: Outcome is stable in 35-month follow-up. *Psychiatry Res.* 159:101–108.
- Horowitz L, Rosenberg SE, Baer BA, Ureño G, Villaseñor VS (1988) Inventory of interpersonal problems: Psychometric properties and clinical applications. J Consult Clin Psychol. 56:885–892.
- Horowitz L, Strauß B, Kordy H (1994) Inventar zur Erfassung Interpersonaler Probleme. [Inventory of Interpersonal Problems.] Weinheim: Beltz Test GmbH.
- Jacobson NS, Truax P (1991) Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 59:12–19.
- Johnson DR, Fontana A, Lubin H, Corn B, Rosenheck R (2004) Long-term course of treatment-seeking Vietnam veterans with posttraumatic stress disorder: Mortality, clinical condition, and life satisfaction. J Nerv Ment Dis. 192:35–41.
- Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE (2005) Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey replication. Arch Gen Psychiatry. 62:593–602.
- Knaevelsrud C, Maercker A (2010) Long-term effects of an internet-based treatment for posttraumatic stress. Cogn Behav Ther. 39:72–77.
- König J, Resick PA, Karl R, Rosner R (2012) Posttraumatische Belastungsstörung. Ein Manual zur Cognitive Processing Therapy. [Posttraumatic stress disorder: A manual for cognitive processing therapy]. Göttingen, Germany: Hogrefe.
- Lampe A, Hofmann P, Gast U, Reddemann L, Schüßler G (2014) Long-term course in female survivors of childhood abuse after psychodynamically oriented, traumaspecific inpatient treatment: A naturalistic two-year follow-up. Z Psychosom Med Psychother. 60:267–282.
- Lenhard W, Lenhard A (2016) Berechnung von Effektstärken. [Calculation of effect sizes]. Dettelbach, Germany: Psychometrica. Available at: https://www.psychometrica.de/ effektstaerke.html. Accessed May 30, 2018.
- Macklin ML, Metzger LJ, Lasko NB, Berry NJ, Orr SP, Pitman RK (2000) Five-year follow-up study of eye movement desensitization and reprocessing therapy for combat-related posttraumatic stress disorder. *Compr Psychiatry*. 41:24–27.
- Maercker A, Schützwohl M (1998) Erfassung von psychischen Belastungsfolgen: Die Impact of Event Skala–revidierte Version (IES-R) [Assessment of psychological sequelae of stress - the Impact of Event Scale–Revised.]. *Diagnostica*. 44:514–521.
- Martinovich Z, Saunders S, Howard K (1996) Some comments on "Assessing clinical significance". *Psychother Res.* 6:124–132.

- Member Centers of the National Consortium of Torture Treatment Programs (NCTTP) (2015) Descriptive, inferential, functional outcome data on 9,025 torture survivors over six years in the United States. *Torture*. 25:34–60.
- Mithoefer MC, Wagner MT, Mithoefer AT, Jerome L, Martin SF, Yazar-Klosinski B, Michel Y, Brewerton TD, Doblin R (2013) Durability of improvement in post-traumatic stress disorder symptoms and absence of harmful effects or drug dependency after 3,4-methylenedioxymethamphetamine-assisted psychotherapy: A prospective long-term follow-up study. J Psychopharmacol. 27:28–39.
- Morris SB (2008) Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. Organizational Research Methods. 11:364–386.
- Paivio SC, Greenberg LS (1995) Resolving "unfinished business": Efficacy of experiential therapy using empty-chair dialogue. J Consult Clin Psychol. 63:419–425.
- Paivio SC, Jarry JL, Chagigiorgis H, Hall I, Ralston M (2010) Efficacy of two versions of emotion-focused therapy for resolving child abuse trauma. *Psychother Res.* 20: 353–366.
- Peleikis DE (2005) Current mental health in women with childhood sexual abuse who had outpatient psychotherapy. *Eur Psychiatry*. 20:260–267.
- Resick PA, Monson CM, Chard KM (2007) Cognitive processing therapy: Veteran/ military version. Washington, DC: Department of Veterans Affairs.
- Resick PA, Monson CM, Chard KM (2016) Cognitive processing therapy for PTSD. A comprehensive manual. New York, NY: The Guilford Press.
- Resick PA, Williams LF, Suvak MK, Monson CM, Gradus JL (2012) Long-term outcomes of cognitive-behavioral treatments for posttraumatic stress disorder among female rape survivors. J Consult Clin Psychol. 80:201–210.
- Shalev AY, Ankri Y, Gilad M, Israeli-Shalev Y, Adessky R, Qian M, Freedman S (2016) Long-term outcome of early interventions to prevent posttraumatic stress disorder. J Clin Psychiatry. 77:e580–e587.
- Tarrier N, Sommerfield C (2004) Treatment of chronic PTSD by cognitive therapy and exposure: 5-Year follow-up. *Behav Ther.* 35:231–246. doi:10.1016/S0005-7894 (04)80037-6.
- Vloet A, Simons M, Vloet TD, Sander M, Herpertz-Dahlmann B, Konrad K (2014) Long-term symptoms and posttraumatic growth in traumatised adolescents: Findings from a specialised outpatient clinic. J Trauma Stress. 27:622–625.
- Watts BV, Schnurr PP, Mayo L, Young-Xu Y, Weeks WB, Friedman MJ (2013) Metaanalysis of the efficacy of treatments for posttraumatic stress disorder. J Clin Psychiatry. 74:e541–e550.
- Wittchen HU, Lachner G, Wunderlich U, Pfister H (1998) Test-retest reliability of the computerized DSM-IV version of the Munich-Composite International Diagnostic Interview (M-CIDI). Soc Psychiatry Psychiatr Epidemiol. 33:568–578.
- Wittchen HU, Pfister H (1997) Diagnostisches Expertensystem f
 ür psychische Störungen (DIA-X). Frankfurt: Swets & Zeitlinger.
- Zimmermann P, Biesold KH, Barre K, Lanczik M (2007) Long-term course of post-traumatic stress disorder (PTSD) in German soldiers: Effects of inpatient eye movement desensitization and reprocessing therapy and specific trauma characteristics in patients with non-combat-related PTSD. *Mil Med.* 172: 456–460.