Psychological Medicine (2014), **44**, 2099–2112. © Cambridge University Press 2013
doi:10.1017/S0033291713002134

ORIGINAL ARTICLE

Randomized controlled trial and uncontrolled 9-month follow-up of an adjunctive emotion regulation group therapy for deliberate self-harm among women with borderline personality disorder

K. L. Gratz[1]*, M. T. Tull[1] and R. Levy[2]

[1] *Department of Psychiatry and Human Behavior, University of Mississippi Medical Center, Jackson, MS, USA*
[2] *T. Denny Sanford School of Social and Family Dynamics, Arizona State University, Tempe, AZ, USA*

Background. Despite the clinical importance of deliberate self-harm (DSH; also referred to as non-suicidal self-injury) within borderline personality disorder (BPD), empirically supported treatments for this behavior among individuals with BPD are difficult to implement in many clinical settings. To address this limitation, a 14-week, adjunctive emotion regulation group therapy (ERGT) for DSH among women with BPD was developed. The current study examined the efficacy of this ERGT in a randomized controlled trial (RCT) and the durability of treatment gains over a 9-month uncontrolled follow-up period.

Method. Female out-patients with BPD and recent recurrent DSH were randomly assigned to receive this ERGT in addition to their ongoing out-patient therapy immediately (n=31) or after 14 weeks (n=30). Measures of DSH and other self-destructive behaviors, psychiatric symptoms, adaptive functioning and the proposed mechanisms of change (emotion dysregulation/avoidance) were administered pre- and post-treatment or -waitlist (to assess treatment efficacy), and 3 and 9 months post-treatment (to assess durability of treatment gains).

Results. Intent-to-treat (ITT) analyses (n=61) revealed significant effects of this ERGT on DSH and other self-destructive behaviors, emotion dysregulation, BPD symptoms, depression and stress symptoms, and quality of life. Analyses of all participants who began ERGT (across treatment and waitlist conditions; n=51) revealed significant improvements from pre- to post-treatment on all outcomes, additional significant improvements from post-treatment to 9-month follow-up for DSH, emotion dysregulation/avoidance, BPD symptoms and quality of life, and no significant changes from post-treatment to 9-month follow-up on the other measures.

Conclusions. The results support the efficacy of this ERGT and the durability of treatment gains.

Received 12 December 2012; Revised 21 July 2013; Accepted 27 July 2013; First published online 28 August 2013

Key words: Borderline personality, emotion regulation, group, self-harm, self-injury, treatment.

Introduction

Deliberate self-harm (DSH; also referred to as non-suicidal self-injury), defined as the deliberate, direct, self-inflicted destruction of body tissue without suicidal intent and for purposes not socially sanctioned (e.g. cutting, burning, severe scratching; Gratz, 2001; ISSS, 2007), is a clinically important behavior commonly associated with borderline personality disorder (BPD) and implicated in the high levels of health-care utilization among individuals with BPD (Linehan, 1993; Zanarini, 2009). Despite the clinical relevance of this behavior, there are few empirically supported treatments for DSH within BPD. Short-term treatments for DSH in general have not been found to be effective for patients with BPD, and may lead to an increase in the repetition of DSH among individuals with BPD (Tyrer *et al.* 2004). Moreover, the two treatments with demonstrated efficacy in the treatment of DSH among patients with BPD, dialectical behavior therapy (DBT; Linehan, 1993) and mentalization-based treatment (Bateman & Fonagy, 2004), are difficult to implement in traditional clinical settings (due to their duration and intensity), and are not readily available in many communities (Zanarini, 2009). Thus, there is a need for shorter, less intensive, and more clinically feasible interventions that directly target DSH among individuals with BPD, particularly adjunctive treatments that may augment the therapies provided by clinicians in the community (Zanarini, 2009).

* Address for correspondence: Dr K. L. Gratz, Department of Psychiatry and Human Behavior, University of Mississippi Medical Center, 2500 North State Street, Jackson, MS 39216 USA.
(Email: KLGratz@aol.com)

To address this need, Gratz & Gunderson (2006) developed an adjunctive emotion regulation group therapy (ERGT) for DSH among women with BPD, designed to augment the usual treatments provided in the community by directly targeting both DSH and its underlying mechanism. Specifically, based on the theory that DSH stems from emotion dysregulation (Linehan, 1993; Gratz & Gunderson, 2006; see also Chapman *et al.* 2006; Nock, 2009; Selby & Joiner, 2009), this ERGT was developed with the expectation that teaching self-harming women with BPD more adaptive ways of responding to and regulating their emotions would reduce their DSH.

Two studies have provided support for the utility of this ERGT in the treatment of DSH within BPD. The first, a small randomized controlled trial (RCT), found that the addition of this ERGT to participants' ongoing out-patient therapy (treatment as usual, TAU) had positive effects on DSH, emotion dysregulation, experiential avoidance, BPD symptoms, and symptoms of depression, anxiety and stress (Gratz & Gunderson, 2006). Moreover, participants in the treatment condition evidenced significant changes over time on all outcome measures and reached normative levels of functioning on most. The second, an open trial examining the utility of this ERGT within a more diverse and underserved setting, found significant improvements from pre- to post-treatment in DSH, emotion dysregulation, experiential avoidance, BPD symptoms, depression, anxiety, and stress symptoms, and social and vocational impairment (Gratz & Tull, 2011).

The present study examined the efficacy of this ERGT in a larger RCT, and the durability of treatment gains over a 9-month uncontrolled follow-up period. Female out-patients with BPD and recent recurrent DSH were randomly assigned to receive this ERGT in addition to their ongoing out-patient therapy immediately (ERGT+TAU) or after 14 weeks (TAU waitlist). The addition of this ERGT to patients' ongoing treatment was expected to have positive effects on DSH and other self-destructive behaviors, emotion dysregulation and experiential avoidance, psychiatric symptoms and adaptive functioning.

Method

Sample

Participants were obtained through referrals by clinicians and self-referrals in response to advertisements for an 'emotion regulation skills group for women with self-harm' posted online and throughout the community. Inclusion criteria included: (*a*) threshold or subthreshold[1]† diagnosis of BPD (given evidence that even subthreshold BPD is clinically meaningful; Clifton & Pilkonis, 2007); (*b*) a history of repeated DSH, with at least one episode in the past 6 months; (*c*) having an individual therapist, psychiatrist or case manager; and (*d*) being a woman aged 18–60 years. To increase generalizability of the sample, exclusion criteria included only diagnoses of a primary psychotic disorder, bipolar I disorder and current (past month) substance dependence. Participants meeting inclusion and exclusion criteria were matched on four prognostic variables [emotion dysregulation, number of lifetime incidents of DSH, global assessment of functioning (GAF) scores, and age] and randomly assigned by the principal investigator (PI) to either the ERGT+TAU (n=31) or TAU waitlist (n=30) condition using a stratified randomization procedure. Demographic, clinical and diagnostic data on participants in each condition are presented in Table 1.

Measures

The following instruments were administered during the initial assessment to screen participants and collect baseline clinical and diagnostic data: (*a*) the Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV; Zanarini *et al.* 1996); (*b*) the Structured Clinical Interview for DSM-IV Axis I Disorders – Patient Edition (SCID-I/P; First *et al.* 1996); (*c*) a modified version of the Lifetime Parasuicide Count (Linehan & Comtois, 1996) to assess lifetime suicidal behaviors; (*d*) an interview version of the Deliberate Self-Harm Inventory (DSHI; Gratz, 2001) to assess lifetime DSH; and (*e*) the Treatment History Interview (THI; Linehan & Heard, 1987) to assess past-year psychiatric treatment.

The Credibility/Expectancy Scales (Borkovec & Nau, 1972) were administered at the end of the first group session to assess the perceived credibility of ERGT and patients' expectancies regarding its benefits. Evidence for the reliability and predictive validity of this measure has been provided (Devilly & Borkovec, 2000).

The following measures were administered pre- and post-treatment or -waitlist to assess treatment efficacy, and at 3 and 9 months post-treatment (for participants in both conditions who received ERGT) to assess maintenance of treatment gains over time.

Measures of DSH and other self-destructive behaviors

The DSHI (Gratz, 2001) is a 17-item self-report questionnaire that assesses various aspects of DSH

† The notes appear after the main text.

Table 1 *Pretreatment demographic, clinical and diagnostic data for the intent-to-treat (ITT) sample*

	ERGT+TAU (n=31)	TAU waitlist (n=30)
Demographic characteristics		
Age (years), mean±S.D.	33.3±11.0	33.0±10.9
Racial/ethnic minority, % (n)	16.1 (5)	26.7 (8)
Lesbian/bisexual/questioning, % (n)	12.9 (4)	13.7 (4)
Marital status, % (n)		
Single	51.7 (16)	56.7 (17)
Married	25.8 (8)	13.3 (4)
Separated/divorced	22.6 (7)	30.0 (9)
Highest educational attainment, % (n)		
Less than high school	6.5 (2)	6.7 (2)
High school graduate	54.8 (17)	73.3 (22)
College graduate	25.8 (8)	16.7 (5)
Income, % (n)		
<US$20 000	38.7 (12)	57.1 (16)
US$20 000–US$59 999	32.3 (10)	32.1 (9)
>US$60 000	29.0 (9)	10.7 (3)
Clinical characteristics		
Number of BPD criteria (DIPD-IV), mean±S.D.	6.5±1.6)	6.0±1.5)
% meeting full criteria for BPD, % (n)	90.3 (28)	86.7 (26)
Suicide attempt in lifetime, % (n), range	58.1 (18), 0–16	66.7 (20), 0–5
Suicide attempt past year, % (n), range	16.1 (5), 0–2	20.0 (6), 0–2
DSH frequency in past 3 months, mean±S.D.	35.5±68.4	28.4±39.4
In-patient hospitalization past year, % (n)	12.9 (4)	26.7 (8)
Total hours/week of ongoing therapy, mean±S.D.	1.2±1.4	2.5±2.6
Hours/week individual therapy, mean±S.D.	0.7±0.4	1.0±0.8
Hours/week group therapy, mean±S.D.	0.4±1.3	0.6±1.8
Number psychiatric medications, mean±S.D.	1.9±1.7	2.1±1.2
Number of months with individual therapist, mean±S.D.	15.5±19.3	14.9±25.4
GAF score, mean±S.D.	43.4±24.6	40.5±19.8
Diagnostic data		
Lifetime Axis I disorders, % (n)		
Mood disorder	80.6 (25)	86.7 (26)
Substance use disorder	54.8 (17)	60.0 (18)
Anxiety disorder	74.2 (23)	86.7 (26)
PTSD	48.4 (15)	63.3 (19)
Eating disorder	36.7 (11)	42.9 (12)
Current Axis I disorders, % (n)		
Mood disorder	41.9 (13)	60.0 (18)
Substance use disorder	0.0 (0)	3.3 (1)
Anxiety disorder	54.8 (17)	70.0 (21)
PTSD	29.0 (9)	43.3 (13)
Eating disorder	16.7 (5)	10.7 (3)
Axis II co-morbidity, % (n)	40.0 (12)	53.3 (16)
Cluster A PD	6.7 (2)	10.0 (3)
Cluster B PD (other than BPD)	13.3 (4)	20.0 (6)
Cluster C PD	36.7 (11)	43.3 (13)

ERGT, Emotion regulation group therapy; TAU, treatment as usual; BPD, borderline personality disorder; DIPD-IV, Diagnostic Interview for DSM-IV Personality Disorders; DSH, deliberate self-harm; PD, personality disorder; GAF, Global Assessment of Functioning; PTSD, post-traumatic stress disorder; S.D., standard deviation.

None of the demographic, clinical or diagnostic variables in this table differed significantly between conditions (t's<1.50, χ^2s<3.49, p's>0.15), with the exception of total hours/week of ongoing therapy.

(defined as the deliberate, direct self-destruction of body tissue without suicidal intent) over specified time periods, including frequency and type of DSH behavior (e.g. cutting, burning, carving, bone breaking). The DSHI demonstrates adequate test–retest reliability and construct, discriminant and convergent validity among diverse non-clinical and patient samples (Gratz, 2001; Fliege *et al.* 2006). For this study (see also Gratz & Gunderson, 2006), a continuous variable measuring frequency of DSH over the specified time period (e.g. in the 3.5 months before the study, since the last assessment) was created by summing participants' scores on the frequency questions for each item (α=0.75).

The Self-Harm Inventory (SHI; Sansone *et al.* 1998) is a 22-item, self-report measure that assesses the presence and frequency of various self-destructive behaviors, including substance abuse, disordered eating behaviors, risky sexual behavior and suicidal behaviors. The SHI has demonstrated good convergent and predictive validity (Sansone *et al.* 1998). This study used a modified version of the SHI to assess past-month frequency of self-destructive behaviors (α=0.67).

Measures of psychiatric symptoms

The Zanarini Rating Scale for Borderline Personality Disorder (ZAN-BPD; Zanarini, 2003) is a clinician-administered instrument for assessing change in BPD symptoms over time. The ZAN-BPD demonstrates good reliability and convergent and discriminant validity (Zanarini, 2003), and was used to provide an interviewer-based assessment of past-week BPD symptom severity (α=0.81). Interviews were conducted by clinical assessors trained to reliability with the PI [intraclass correlation coefficient (ICC)=0.92].

The Borderline Evaluation of Severity over Time (BEST; Pfohl *et al.* 2009) is a 15-item, self-report measure of past-month BPD symptom severity. The BEST demonstrates adequate reliability and good convergent and discriminant validity (Pfohl *et al.* 2009). Internal consistency in this sample was good (α=0.84).

The Beck Depression Inventory – Second Edition (BDI-II; Beck *et al.* 1996) is a widely used, psychometrically sound, self-report measure of depression symptom severity. Items were summed to obtain a total depression severity score (α=0.92).

The Depression Anxiety Stress Scales (DASS; Lovibond & Lovibond, 1995) is a 21-item self-report measure that provides separate scores for depression, anxiety and stress. The DASS has demonstrated good test–retest reliability and construct and discriminant validity (Lovibond & Lovibond, 1995; Roemer, 2001) and was used to assess general psychiatric symptom severity (α's=0.87–0.93).

Measures of adaptive functioning

The BPD-related composite of the Inventory of Interpersonal Problems (IIP-BPD; Lejuez *et al.* 2003) is an 18-item self-report measure of interpersonal problems relevant to BPD, including interpersonal sensitivity and aggression. The IIP-BPD has been found to demonstrate good convergent validity and specificity for BPD (Lejuez *et al.* 2003). Internal consistency in this sample was good (α=0.86).

The Sheehan Disability Scale (SDS; Sheehan, 1983) is a widely used, three-item, self-report measure of social and vocational impairment due to psychological symptoms. The SDS demonstrates adequate reliability and construct, convergent and discriminant validity across various clinical populations (Hambrick *et al.* 2004; Diefenbach *et al.* 2007). Items were summed to obtain a total score of social and vocational impairment (α=0.86).

The Quality of Life Inventory (QOLI; Frisch *et al.* 1992) is a 32-item self-report measure based on an empirically validated model of life satisfaction that conceptualizes satisfaction as the sum of satisfactions in areas of life that are important to an individual. Sixteen areas of life are assessed in terms of degree of importance and level of satisfaction. The QOLI has good convergent, divergent and predictive validity (Frisch *et al.* 1992). Scores range from −6 to +6, with higher positive scores indicating greater quality of life (α=0.82).

Measures of emotion dysregulation and experiential avoidance

The Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004) is a 36-item self-report measure that assesses individuals' typical levels of emotion dysregulation across six domains: emotional non-acceptance, difficulties controlling impulsive behaviors and engaging in goal-directed behaviors when distressed, limited access to effective emotion regulation strategies, and lack of emotional awareness and clarity. The DERS has good test–retest reliability and construct and predictive validity (Gratz & Roemer, 2004; Gratz & Tull, 2010). Internal consistency in this sample was good (α's=0.77–0.93).

The Acceptance and Action Questionnaire (AAQ; Hayes *et al.* 2004) is a nine-item, self-report measure of experiential avoidance, or the tendency to avoid unwanted internal experiences (particularly emotions). The AAQ demonstrates adequate convergent, discriminant and concurrent validity (Hayes *et al.* 2004).

Higher scores indicate greater experience avoidance (α=0.75).

Procedures

All methods received prior approval by the institution's Institutional Review Board. After providing written informed consent, participants completed the initial assessment interview, conducted by clinical assessors trained to reliability with the PI ($\kappa \geqslant 0.80$). All initial assessment interviews were reviewed by the PI; diagnostic discrepancies were found in <10% of cases. In these instances, areas of disagreement were discussed and a consensus was reached.

For participants meeting inclusion and exclusion criteria, random assignment to the treatment or waitlist condition occurred as soon as enough participants had been screened; therefore, the time between the initial assessment and randomization differed between participants, ranging from <1 week to approximately 4 months (mean=29 days). Five treatment cohorts were recruited between June 2009 and December 2010. Pre-treatment and pre-waitlist assessments were completed within 1 week prior to the start of the group therapy for participants in the treatment condition; post-treatment and post-waitlist assessments were completed within 1 week following the end of the group therapy. The post-waitlist assessment served as the pre-treatment assessment for participants in the waitlist condition, with their post-treatment assessment occurring within 1 week following the end of their group therapy. For all participants, follow-up assessments were completed 3 and 9 months (i.e. 14 and 38 weeks) following completion of the post-treatment assessment. All assessments were conducted by trained assessors masked to participant condition.

Treatment

ERGT

This 14-week ERGT is based on the conceptualization of emotion regulation as a multidimensional construct involving: (*a*) the awareness, understanding and acceptance of emotions; (*b*) the ability to engage in goal-directed behaviors and inhibit impulsive behaviors when experiencing negative emotions; (*c*) the use of situationally appropriate strategies to modulate the intensity or duration of emotions (rather than to eliminate emotions entirely); and (*d*) the willingness to experience negative emotions as part of pursuing meaningful activities in life (Gratz & Roemer, 2004). ERGT draws from two acceptance-based behavioral therapies, acceptance and commitment therapy (Hayes *et al.* 1999) and DBT (Linehan, 1993), and emphasizes the following themes: (*a*) the potentially paradoxical effects of emotional avoidance, (*b*) the emotion-regulating consequences of emotional acceptance/willingness and (*c*) the importance of controlling behaviors when emotions are present, rather than controlling emotions themselves. A detailed manual has been developed and a full description of the specific topics addressed each week is available elsewhere (Gratz & Gunderson, 2006). Groups meet weekly for 90 min over 14 weeks and are limited to six patients per group.

TAU

All participants were required to have an individual clinician, and all continued with their ongoing outpatient treatment over the course of the RCT. Participants had been meeting with their individual clinicians for an average of 15 months (S.D.=22.3 months, range <1 month to 9 years) prior to the start of the study, with 72% reporting a duration of $\geqslant$2 months. Few participants (18%) received group therapy outside of ERGT, and 54% received <1 h of individual therapy each week. Further information on participants' TAU is provided in Table 1. With regard to the individual clinicians of the study participants, 18% were in private practice and the others worked in a community mental health center (51%), college counseling center (10%) or hospital (20%). With regard to their training, 72% had a master's degree, 20% were clinical psychologists and 7% were psychiatrists. Most participants (>70%) were receiving supportive or dynamic individual therapy (according to the THI and discussions with the clinicians); however, 19% were receiving cognitive behavioral therapy (although none were receiving DBT).

Group therapists and treatment integrity

Two doctoral-level therapists were trained by the treatment developer (K.L.G.) to lead the groups. The initial training [including direct or indirect (by videotape) observation of K.L.G. administering the treatment, role plays and weekly supervision] lasted approximately 4 months, with ongoing supervision provided by the PI throughout the RCT. Depending on therapist availability, the groups had one or two leaders.

The PI reviewed all group sessions for adherence and competence (with 25% rated by an independent trained rater with good reliability; κ=0.90 for adherence ratings and ICC=0.86 for competence ratings). An adherence checklist (adapted from Roemer & Orsillo, 2007) was developed, that lists 10 elements encouraged (although not required) in each session (e.g. emphasizing the functionality of emotions, promoting emotional acceptance, emphasizing behavioral *versus* emotional control) and four elements forbidden

(e.g. emphasizing emotional control, emphasizing the need to change cognitions). All elements are rated for each session, despite differing content each week (Roemer & Orsillo, 2007). Project therapists were adherent to the protocol, with an average of 8.1±1.1 of the encouraged elements discussed in each group and only one minor non-protocol event recorded. Competence across six key domains (e.g. therapeutic stance, promoting emotional acceptance, increasing awareness of the functions of DSH) was rated on a 0–2 scale (0=poor/unacceptable; 1=adequate/acceptable; 2=good), with a score of 9 considered acceptable. The average competence rating of the project therapists was 11.3±0.9 (mean item-level rating=1.9±0.1).

Results

Ninety-one women completed the initial assessment, and 61 were randomized to the treatment (n=31) or waitlist (n=30) condition. There were no significant between-group differences in the time between the initial assessment and randomization to condition (t=0.67, p>0.50). Figure 1 shows the flowchart of patient enrollment and disposition across the study. Twelve participants dropped out of ERGT (five from the treatment and seven from the waitlist condition; χ^2=2.41, p>0.10), resulting in an overall drop-out rate of 23.5%. Ratings of treatment credibility and expectancy were 7.22% and 66% respectively.

One-way (treatment *versus* waitlist) ANOVAs were conducted on pre-treatment scores on outcome measures to determine equivalence across conditions. The results indicate no significant between-group differences (F's <3.82, Cohen's d's <0.50, p's >0.05), with one exception: the treatment condition had significantly lower scores than the waitlist condition on the DERS lack of clarity subscale ($F_{1,59}$=5.89, p<0.05). Comparable findings were obtained when using non-parametric Mann–Whitney tests to compare conditions without assuming normality. Moreover, the results revealed no significant between-group differences in past-year or lifetime DSH frequency (t's <1.50, p's >0.15).

The results of a series of t tests and χ^2 analyses conducted on the demographic, clinical, and diagnostic variables in Table 1 revealed no significant between-group differences on any of these variables (t's <1.50, χ^2s<3.49, p's >0.15), with the exception of hours of overall TAU per week (which was greater for participants in the waitlist *versus* treatment condition; t=2.34, p<0.05). Due to this difference in TAU, the average number of hours of treatment per week did not differ significantly between conditions when including the 1.5 h of treatment time associated with ERGT (ERGT+TAU=2.7; TAU waitlist=2.5; t=0.43, p>0.10).

Finally, providing support for the stability of participants' symptoms from the initial assessment to the pre-treatment assessment, participants' scores on measures of emotion dysregulation and DSH frequency did not change significantly from the initial assessment to the pre-treatment assessment (t's<1.28, p's>0.20). Moreover, the impact of time between the initial and pre-treatment assessments on changes in participants' scores on these measures across this time period was negligible (F's<0.28, η_p^2s<0.01, p's>0.60).

RCT analyses

Latent growth models were used to examine treatment effects, with a linear growth structure modeled from pre- to post-values and Condition membership (coded 1 for treatment, 0 for waitlist) modeled as influencing the latent intercept and latent slope. This model is depicted as a path diagram in the supplementary online material. In this model, pre-treatment assessments serve as the reference point and the latent intercept and slope variables represent participant-specific starting points and slopes. The effect of Condition on the Intercept factor captures pre-treatment differences between conditions (comparable to the results of the ANOVAs reported earlier), and the effect of Condition on the Slope factor captures the treatment effect.

We adopted a Bayesian approach to growth modeling (Zhang *et al.* 2007) and fit the models using the Markov chain Monte Carlo routines in Mplus (Muthén & Muthén, 1998–2010) using $N(0,10^{10})$ priors for the intercepts and paths from Condition to the factors, and $G^{-1}(-1,0)$ priors for the error variances. This approach implements a multiple imputation strategy to handle missing data (Enders, 2010), enabling an analysis of the intent-to-treat (ITT) sample. This approach is also advantageous in the analysis of small sample sizes (Lee & Song, 2004) and supports the direct probabilistic statements about parameters. Three chains were run from dispersed start points to diagnose convergence and yield draws from the posterior.

Descriptive statistics for all outcome measures at pre- and post-treatment or -waitlist and the effects of Condition on the Slope factors are provided in Table 2. For each effect, posterior means are given as point estimates and posterior standard deviations are given to represent the uncertainty. Boundaries of the 95% central posterior credibility interval (analogous to 95% confidence intervals but interpretable as direct probabilistic statements) are also presented. The estimate of the effect of Condition was interpreted as significant when the credibility interval did not

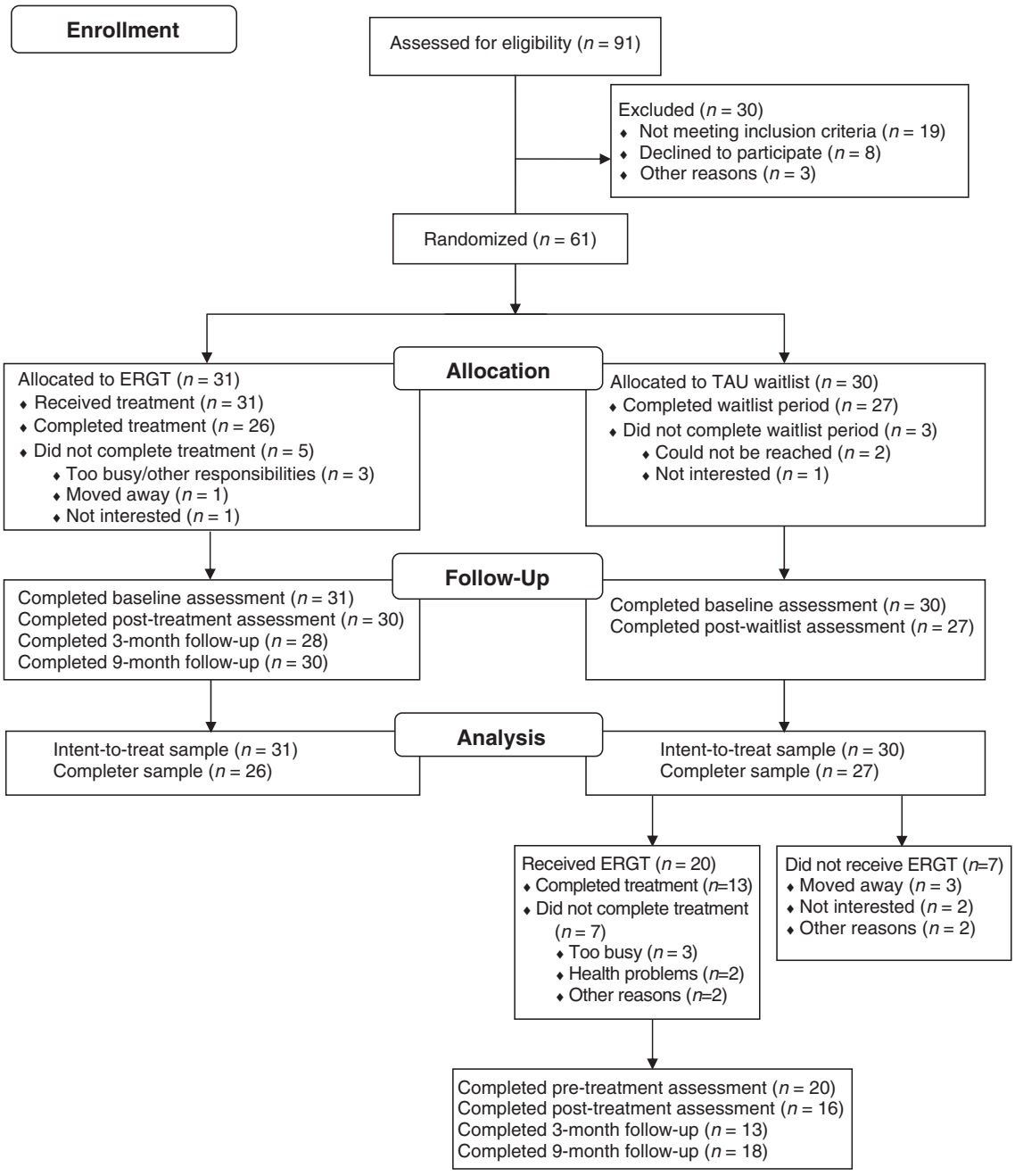

Fig. 1. CONSORT flow diagram of patient enrollment and disposition in the randomized controlled trial (RCT) and across the follow-up period. ERGT, Emotion regulation group therapy.

contain zero. The results revealed significant effects (accompanied by medium to large effect sizes) of ERGT on DSH and other self-destructive behaviors, emotion dysregulation (overall and across the dimensions of emotional non-acceptance, difficulties engaging in goal-directed behaviors when distressed, and lack of access to effective regulation strategies), BPD symptoms on the ZAN-BPD, depression and stress symptoms on the DASS, and quality of life. The effects of ERGT on experiential avoidance and interpersonal functioning were medium sized (0.71 and 0.48 respectively) and approached the criterion for significance (and would have reached significance with the use of 90% credibility intervals).

An approach consistent with that proposed by Jacobson & Truax (1991) was used to determine the clinical significance of these treatment effects, requiring that participants (*a*) report reliable improvement and (*b*) reach normative levels of functioning (see also Espie *et al.* 2001; Ogles *et al.* 2001). The results of χ^2

Table 2. *Descriptive statistics and results of latent growth analyses examining treatment effects for all outcomes at pre- and post-treatment or -waitlist for the intent-to-treat (ITT) sample (n=61)*

	ERGT+TAU				TAU waitlist				Latent growth analyses of treatment effects				
	Pre-treatment		Post-treatment		Pre-waitlist		Post-waitlist		Condition on slope				
Outcome	Mean[a]	S.D.[b]	Mean[a]	S.D.[c]	Mean[a]	S.D.[b]	Mean[a]	S.D.[c]	Posterior Mean	Posterior S.D.	2.5%	97.5%	Effect size[d]
DSHI[e]	68.47	159.96	16.67	39.74	23.47	37.19	19.26	24.25					
DSHI-t	1.17	0.74	0.69	0.35	1.02	0.57	0.97	0.35	−0.42	0.14	−0.70	−0.15*	−0.64
SHI	42.66	61.08	16.05	38.76	19.51	26.77	29.40	38.76	−36.49	10.13	−53.01	−14.97*	−0.77
DERS	106.81	21.87	95.27	15.60	112.26	25.31	113.62	15.60	−12.91	5.49	−23.18	−2.91*	−0.55
DERS-NA	17.58	6.54	15.70	4.19	17.30	6.59	19.14	4.19	−3.72	1.31	−6.23	−1.25*	−0.57
DERS-IM	15.92	5.79	15.23	3.83	17.26	6.20	18.74	3.83	−2.16	1.13	−4.31	0.04	−0.36
DERS-GO	18.03	4.72	15.76	2.98	17.54	5.47	19.10	2.98	−3.83	0.97	−5.73	−1.87*	−0.75
DERS-AW	18.00	4.63	16.88	3.53	20.27	4.83	17.39	3.53	1.76	1.16	−0.43	3.99	0.37
DERS-ST	24.38	6.92	20.60	4.76	24.76	7.53	26.08	4.76	−5.09	1.29	−7.57	−2.67*	−0.70
DERS-CL	12.45	3.93	11.70	3.36	14.84	3.73	13.70	3.36	0.39	0.99	−1.54	2.38	0.10
AAQ	43.18	7.43	38.11	6.73	42.64	6.79	42.63	6.73	−5.06	2.33	−9.11	0.08	−0.71
ZAN-BPD	11.94	8.05	4.35	4.04	10.59	6.88	12.03	4.04	−9.02	1.48	−11.80	−6.12*	−1.20
BEST	33.32	11.23	27.47	6.59	38.06	10.15	35.88	6.59	−3.67	2.37	−8.26	0.96	−0.34
BDI-II	26.24	13.34	19.98	8.26	32.54	12.67	28.81	8.26	−2.53	2.53	−7.36	2.33	−0.19
DASS-D	20.09	12.61	13.04	5.63	21.39	14.92	21.30	5.63	−6.97	2.77	−11.40	−0.26*	−0.51
DASS-A	15.02	11.72	10.78	5.65	20.23	12.97	20.66	5.65	−4.68	2.89	−10.13	1.14	−0.38
DASS-S	20.03	11.62	14.58	4.16	23.81	11.75	25.34	4.16	−6.99	2.25	−11.27	−2.52*	−0.60
IIP-BPD	1.77	0.65	1.45	0.57	1.91	0.80	1.94	0.57	−0.35	0.18	−0.71	0.01	−0.48
SDS	17.82	8.75	14.25	6.24	18.09	9.44	16.01	6.24	−1.50	2.49	−6.70	3.22	−0.16
QOLI	−0.88	2.10	0.31	1.62	−0.57	2.20	−0.50	1.62	1.11	0.50	0.14	2.10*	0.52

ERGT, Emotion regulation group therapy; TAU, treatment as usual; DSHI-t, Deliberate Self-Harm Inventory – log-transformed score; SHI, Self-Harm Inventory; DERS-NA, Difficulties in Emotion Regulation Scale Non-Acceptance; DERS-IM, DERS Impulse; DERS-GO, DERS Goals; DERS-AW, DERS Awareness; DERS-ST, DERS Strategies; DERS-CL, DERS Clarity; AAQ, Acceptance and Action Questionnaire; ZAN-BPD, Zanarini Rating Scale for Borderline Personality Disorder; BEST, Borderline Evaluation of Severity over Time; BDI-II, Beck Depression Inventory – Second Edition; DASS-D, Depression Anxiety Stress Scales – Depression; DASS-A, DASS – Anxiety; DASS-S, DASS – Stress; IIP-BPD, BPD-related composite of the Inventory of Interpersonal Problems; SDS, Sheehan Disability Scale; QOLI, Quality of Life Inventory; S.D., standard deviation.

[a] Model-implied mean computed from the latent growth model point estimates.

[b] Observed standard deviation.

[c] Point estimate of residual standard deviation.

[d] Effect size estimated as group difference in post-treatment mean improvement divided by pooled standard deviation at pre-treatment.

[e] Observed means and standard deviations for participants who completed assessments.

* $p<0.05$.

analyses examining between-group differences in the percentage of participants meeting none, one or both of these criteria revealed significant differences for DSH, emotion dysregulation, experiential avoidance and all psychiatric symptom measures (see Table 3), with >35% of participants in the treatment condition reporting clinically significant improvements on measures of DSH, experiential avoidance and BPD symptoms, and >60% reaching normative levels of functioning on measures of emotion dysregulation, experiential avoidance, BPD symptoms and interpersonal functioning.

Uncontrolled analyses of maintenance of treatment gains

Latent growth models were used to model the extent to which treatment gains were maintained among the full sample of participants who began ERGT [across treatment and waitlist conditions (n=51); Table 4]. To model different possible trajectories during and after treatment, we used piecewise linear growth models (Chou *et al.* 2004; Hardy & Thiels, 2009), modeling the pre-treatment, post-treatment, 3-month and 9-month follow-up time points. This model (depicted in the supplementary online material) scales time in terms of weeks. To model DSH over time in light of the unequal intervals between assessments, DSH frequencies were scaled to be the frequency per 14 weeks.[2] In this model, the pre-treatment assessment serves as the reference point for the first linear growth process (with the post-treatment assessment occurring 14 weeks later), and a second linear growth process commences at the conclusion of treatment (influencing the 3- and 9-month follow-ups that occur 14 and 38 weeks post-treatment respectively). The latent variables for the intercept and slopes represent participant-specific starting points and slopes. However, with only two time points defining the first slope (pre-treatment and post-treatment), this cannot be modeled as a random effect over participants due to identification restrictions. To fit the model, Slope 1 was treated as constant over participants (i.e. there is a single Slope 1 that applies to all participants). The mean of the first slope factor (Slope 1) captures the linear change during treatment (assumed to apply to all participants due to the identification restriction) and the mean of the second slope factor (Slope 2) captures the average linear change from post-treatment to the 9-month follow-up.

As before, we used a Bayesian approach to growth modeling and fit the models using $N(0, 10^{10})$ priors for latent means, $G^{-1}(-1, 0)$ priors for the error variances and a $W^{-1}(0, -3)$ prior for the latent variable covariance matrix. As shown in Table 4, the results for Slope 1 revealed significant improvements from pre- to post-treatment on all measures (as the posterior means and entirety of the 95% credibility intervals are <0 for all measures other than the QOLI; for the QOLI, the entirety of the 95% credibility interval is >0). Specific lower and upper boundaries of the 95% credibility intervals for Slope 1 for all outcome measures are available in the supplementary online material. The results for Slope 2 revealed further significant improvements from post-treatment to 9-month follow-up for DSH, emotion dysregulation (overall and across three specific dimensions), experiential avoidance, BPD symptoms on the BEST, and quality of life (Table 4). Indeed, changes on these measures from pre-treatment to 9-month follow-up were accompanied by large effect sizes (with the exception of quality of life, which had a medium-sized effect). For the remaining outcome measures, findings that the 95% credibility intervals for Slope 2 contained zero indicate no significant changes from post-treatment to the 9-month follow-up (Table 4).

Data on the clinical significance of these improvements at both post-treatment and 9-month follow-up are presented in Table 3. At post-treatment, >60% of the participants had reached normative levels of functioning on measures of emotion dysregulation, experiential avoidance, BPD symptoms and interpersonal functioning, and >47% reported normative levels of depression and stress symptoms and DSH (i.e. abstinence from DSH). Moreover, >33% of participants reported clinically significant improvements in DSH, experiential avoidance and BPD symptoms. Further improvements were seen during the follow-up period, as >43% of participants reported clinically significant improvements in DSH, emotion dysregulation, experiential avoidance and BPD symptoms at 9 months post-treatment.

Discussion

These results provide further support for the efficacy of this ERGT, revealing positive effects of this treatment on DSH and other self-destructive behaviors, emotion dysregulation, BPD symptoms, depression and stress symptoms, and overall quality of life within a conservative ITT sample. Moreover, findings from the uncontrolled follow-up period provide preliminary support for the durability of treatment gains, as all improvements observed from pre- to post-treatment were maintained or further improved upon at follow-up, including additional significant improvements from post-treatment to the 9-month follow-up for DSH, emotion dysregulation, experiential avoidance, BPD symptoms and quality of life. Providing support for the clinical significance of these improvements, more

Table 3. *Clinical significance of treatment effects in the RCT (n=61) and treatment gains in the uncontrolled follow-up (n=51)*

	Intent-to-treat sample at post-treatment (*n*=61)							Full sample of participants who began ERGT (*n*=51)					
	ERGT+TAU			TAU waitlist				Post-treatment			9-month follow-up		
Outcome	% Reliable improvement[a]	% Normal function[b]	% Both criteria	% Reliable improvement[a]	% Normal function[b]	% Both criteria	χ^2	% Reliable improvement[a]	% Normal function[b]	% Both criteria	% Reliable improvement[a]	% Normal function[b]	% Both criteria
DSH													
DSHI	61.3[c]	45.2[d]	35.5	26.7[c]	20.0[d]	16.7	10.62**	60.8[c]	47.1[d]	33.3	86.3[c]	47.1[e]	43.1
								61.5	*51.3*	*38.5*	*94.9*	*53.8*	*51.3*
Proposed mediators													
DERS	29.0[f]	61.3	25.8	10.0[f]	23.3	0.0	13.91**	29.4[f]	60.8	27.5	52.9[f]	64.7	49.0
								33.3	*69.2*	*30.8*	*53.8*	*69.2*	*48.7*
AAQ	41.9	64.5	35.5	13.3	43.3	13.3	6.08*	45.1	64.7	37.3	51.0	72.5	45.1
								43.6	*66.7*	*38.5*	*56.4*	*79.5*	*51.3*
Psychiatric symptoms													
ZAN-BPD	51.6[f]	77.4	41.9	10.0[f]	26.7	6.7	24.05**	54.9[f]	80.4	45.1	58.8[f]	86.3	49.0
BEST	29.0	71.0	19.4	20.0	53.3	20.0	6.12*	25.5	70.6	19.6	49.0	82.4	47.1
BDI-II	19.4	48.4	9.7	13.3	16.7	10.0	12.04**	27.5	41.2	11.8	45.1	56.9	35.3
DASS-D	29.0	41.9[g]	9.7	3.3	26.7[g]	0.0	8.62*	29.4	49.0[g]	9.8	37.3	54.9[g]	23.5
DASS-A	29.0	38.7[g]	3.2	6.7	16.7[g]	3.3	13.64**	25.5	33.3[g]	3.9	27.5	47.1[g]	13.7
DASS-S	25.8	51.6[g]	19.4	6.7	26.7[g]	6.7	6.46*	29.4	49.0[g]	15.7	33.3	52.9[g]	27.5
Adaptive functioning													
IIP-BPD	22.6	67.7	19.4	13.3	50.0	13.3	2.83	33.3	70.6	29.4	43.1	86.3	41.2
SDS	25.8	25.8[h]	9.7	20.0	13.3[h]	10.0	3.25	27.5	31.4[h]	11.8	37.3	37.3[h]	23.5
QOLI	32.3[f]	35.5	19.4	10.0[f]	26.7	10.0	3.11	37.3[f]	35.3	17.6	43.1[f]	45.1	27.5

RCT, Randomized controlled trial; ERGT, emotion regulation group therapy; TAU, treatment as usual; DSH, Deliberate self-harm; DSHI, Deliberate Self-Harm Inventory; DERS, Difficulties in Emotion Regulation Scale; AAQ, Acceptance and Action Questionnaire; ZAN-BPD, Zanarini Rating Scale for Borderline Personality Disorder; BEST, Borderline Evaluation of Severity over Time; BDI-II, Beck Depression Inventory – Second Edition; DASS-D, Depression Anxiety Stress Scales – Depression; DASS-A, DASS – Anxiety; DASS-S, DASS – Stress; IIP-BPD, BPD-related composite of the Inventory of Interpersonal Problems; SDS, Sheehan Disability Scale; QOLI, Quality of Life Inventory.

Primary findings are based on the full sample of participants in each phase of the study; missing values were estimated using multiple imputation based on the fitted model subject to boundaries of the variables. Italic type indicates percentages for treatment completers only (*n*=39).

[a] Scores that changed by $\geqslant$1 standard deviation (S.D.) of the pre-treatment mean from pre- to post-treatment, unless noted otherwise.

[b] Scores within 1 S.D. of the mean for non-clinical samples, unless noted otherwise.

[c] Greater than 50% reduction in pre-treatment DSH frequency (see Blanchard *et al.* 1990; Espie *et al.* 2001; Ogles *et al.* 2001).

[d] Abstinence from DSH during the last 2 months of treatment.

[e] Abstinence from DSH throughout the entire follow-up period.

[f] Jacobson & Truax's (1991) reliable change index (RCI) calculated.

[g] Scores in the normal range on the DASS.

[h] No significant impairment reported in any area.

*p<0.05, ** p<0.01.

Table 4. *Descriptive statistics and results of piecewise linear growth models examining changes across treatment and follow-up for the full sample of participants who began ERGT (n=51)*

									Slopes from piecewise linear growth models						
	Pre-treatment		Post-treatment		3-month follow-up		9-month follow-up		Mean of Slope 1		Mean of Slope 2			Effect sizes	
Outcome	Mean[a]	S.D.[b]	Mean[a]	S.D.[c]	Mean[a]	S.D.[c]	Mean[a]	S.D.[c]	Posterior Mean	Posterior S.D.	Posterior Mean	Posterior S.D.	95% CI	Post-treatment[d]	9-month follow-up[d]
DSHI[e]	49.68	127.08	16.09	37.35	9.80	37.82	10.98	34.69							
DSHI-t	1.12	0.68	0.65	0.65	0.49	0.45	0.20	0.10	−0.03	0.01*	−0.01	0.00	−0.02 to −0.01*	−0.68	−1.36
SHI	37.35	63.62	16.03	16.03	16.52	21.02	17.36	7.97	−1.52	0.55*	0.04	0.12	−0.21 to 0.28	−0.34	−0.31
DERS	109.05	21.72	94.48	94.48	90.63	16.64	84.03	7.71	−1.04	0.19*	−0.28	0.10	−0.46 to −0.08*	−0.67	−1.15
DERS-NA	18.26	6.33	15.49	15.49	14.69	4.18	13.32	2.11	−0.20	0.05*	−0.06	0.02	−0.10 to −0.01*	−0.44	−0.78
DERS-IM	17.19	5.88	14.90	14.90	14.24	3.76	13.11	1.65	−0.16	0.05*	−0.05	0.02	−0.09 to −0.01*	−0.39	−0.69
DERS-GO	18.49	4.63	15.57	15.57	15.16	3.33	14.46	1.69	−0.21	0.04*	−0.03	0.02	−0.07 to 0.01	−0.63	−0.87
DERS-AW	17.60	4.41	16.17	16.17	15.94	7.98	15.53	9.59	−0.10	0.04*	−0.02	0.02	−0.06 to 0.03	−0.32	−0.47
DERS-ST	24.63	6.72	20.28	20.28	19.03	5.33	16.89	2.22	−0.31	0.07*	−0.09	0.03	−0.14 to −0.04*	−0.65	−1.15
DERS-CL	12.96	4.18	11.98	11.98	11.55	2.13	10.81	1.21	−0.07	0.04*	−0.03	0.01	−0.06 to 0.00	−0.23	−0.52
AAQ	42.37	7.54	37.93	37.93	36.85	5.48	35.00	2.65	−0.32	0.08*	−0.08	0.03	−0.14 to −0.01*	−0.59	−0.98
ZAN-BPD	12.03	7.62	4.51	4.51	4.26	3.31	3.82	2.04	−0.54	0.07*	−0.02	0.02	−0.05 to 0.02	−0.99	−1.08
BEST	33.71	10.55	28.37	28.37	26.61	6.40	23.58	7.86	−0.38	0.09*	−0.13	0.05	−0.22 to −0.03*	−0.51	−0.96
BDI-II	26.41	12.92	18.98	18.98	18.03	10.19	16.40	5.17	−0.53	0.11*	−0.07	0.05	−0.17 to 0.03	−0.58	−0.78
DASS-D	19.74	12.80	12.93	12.93	12.57	10.04	11.95	3.98	−0.49	0.09*	−0.03	0.04	−0.11 to 0.06	−0.53	−0.61
DASS-A	16.70	12.19	13.13	13.13	12.76	8.18	12.11	3.95	−0.26	0.08*	−0.03	0.04	−0.11 to 0.05	−0.29	−0.38
DASS-S	21.36	10.93	15.65	15.65	15.32	7.48	14.74	3.21	−0.41	0.08*	−0.02	0.04	−0.11 to 0.06	−0.52	−0.61
IIP-BPD	1.81	0.71	1.49	1.49	1.39	0.34	1.22	0.28	−0.02	0.01*	−0.01	0.00	−0.01 to 0.00	−0.46	−0.83
SDS	16.80	8.57	13.27	13.27	12.60	6.62	11.45	5.01	−0.25	0.09*	−0.05	0.04	−0.13 to 0.04	−0.41	−0.62
QOLI	−0.68	2.40	0.37	0.37	0.62	1.46	1.05	0.98	0.08	0.02*	0.02	0.01	0.00 to 0.04*	0.44	0.72

ERGT, Emotion regulation group therapy; DSHI, Deliberate Self-Harm Inventory; DSHI-t, DSHI – log-transformed score; SHI, Self-Harm Inventory; DERS, Difficulties in Emotion Regulation Scale; DERS-NA, Difficulties in Emotion Regulation Scale Non-Acceptance; DERS-IM, DERS Impulse; DERS-GO, DERS Goals; DERS-AW, DERS Awareness; DERS-ST, DERS Strategies; DERS-CL, DERS Clarity; AAQ, Acceptance and Action Questionnaire; ZAN-BPD, Zanarini Rating Scale for Borderline Personality Disorder; BEST, Borderline Evaluation of Severity over Time; BDI-II, Beck Depression Inventory – Second Edition; DASS-D, Depression Anxiety Stress Scales – Depression; DASS-A, DASS – Anxiety; DASS-S, DASS – Stress; IIP-BPD, BPD-related composite of the Inventory of Interpersonal Problems; SDS, Sheehan Disability Scale; QOLI, Quality of Life Inventory; CI, credibility interval; S.D., standard deviation.

[a] Model-implied mean computed from the latent growth model point estimates.

[b] Observed standard deviation.

[c] Point estimate of residual standard deviation.

[d] Effect size estimated as the difference between the post-treatment or 9-month mean and the pre-treatment mean divided by the standard deviation at pre-treatment.

[e] Observed means and standard deviations for participants who completed assessments.

* 95% CI does not contain 0 ($p<0.05$).

than 33% of participants reported clinically significant improvements in DSH, experiential avoidance, and BPD symptoms at post-treatment, and more than 43% reported clinically significant improvements in DSH, emotion dysregulation, experiential avoidance, and BPD symptoms at the 9-month follow-up. In addition, more than 60% of participants reported normative levels of functioning on measures of emotion dysregulation, experiential avoidance, BPD symptoms, and interpersonal functioning at both post-treatment and 9-month follow-up, and more than 47% of participants reported abstinence from DSH during the second half of the group therapy and throughout the follow-up period.

Researchers have underscored the need for shorter, less intensive and more clinically feasible interventions for DSH among patients with BPD, with an emphasis on adjunctive treatments that augment the therapy of clinicians in the community (Zanarini, 2009). The findings from this study suggest that this ERGT may be a useful treatment in this regard. Despite not being paired with any particular form of individual therapy (and most participants receiving supportive or dynamic psychotherapy rather than an ERGT-consistent cognitive behavioral therapy), positive treatment effects were found across multiple domains and preliminary evidence suggests that improvements observed over the course of the treatment are maintained. Moreover, this ERGT served as the primary treatment for 43% of participants who met with their individual clinicians only once or twice a month. The finding of positive effects of this treatment within a relatively underserved sample receiving a range of TAU provides further support for its portability and generalizability.

Several limitations warrant discussion. First, although the assessments included both clinician-administered and self-report measures, the exclusive reliance on subjective measures may introduce bias. Future studies would benefit from the additional inclusion of objective measures (e.g. behavioral and/or physiological measures of emotion dysregulation; Thayer & Lane, 2000; Gratz *et al.* 2006). Second, given our exclusive focus on women with BPD, the generalizability of these findings to adolescents and men remains unclear. Third, in the absence of a control group in the analyses examining the maintenance of treatment gains during the follow-up period, it is not possible to determine the effects of ERGT (*versus* the passage of time or some other unknown factor) on the maintenance of treatment gains over time. Thus, the long-term efficacy of this treatment remains unknown. Future research using an active comparison condition is needed to examine the durability of treatment effects over time. Longer follow-up periods are also needed.

Fourth, despite providing evidence for the positive effects of this ERGT on certain dimensions of adaptive functioning, the results of this study revealed limited improvement in social and vocational functioning as a result of this treatment. Further research is needed to explore whether, to what extent, and for whom improvements in various dimensions of adaptive functioning are observed following this brief treatment. How this ERGT compares to existing treatments for DSH among women with and without BPD also warrants investigation.

Finally, additional research is needed to examine the mechanisms of change in this treatment. Although preliminary findings provide support for emotion regulation as a mechanism of change in ERGT (Gratz *et al.* 2012), these findings warrant replication in the current sample. Future research should also examine whether changes in emotion dysregulation mediate the observed improvements in BPD and depression symptoms (both of which are theorized to stem from emotion dysregulation; Gratz & Tull, 2010), as well as other self-destructive behaviors that may serve an emotion-regulating function (e.g. substance abuse and disordered eating behaviors; Safer *et al.* 2009; Gratz & Tull, 2010).

Notes

[1] The vast majority of participants (88.5%; $n = 54$) met full diagnostic criteria for BPD.

[2] One participant was dropped from the analyses of DSH because her 9-month follow-up assessed the frequency of DSH since post-treatment (rather than since the 3-month follow-up as for the other participants).

Supplementary material

For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S0033291713002134.

Acknowledgments

This research was supported by National Institute of Mental Health Grant R34 MH079248, awarded to K. L. Gratz. Portions of these data were previously presented at the First Swedish National Conference on Self-Harm in Stockholm, Sweden in October 2012; the first annual meeting of the North American Society for the Study of Personality Disorders in Boston, MA in April 2013; the annual meeting of the International Society for the Study of Self-Injury in Vancouver, Canada in June 2013; and as part of the National Education Alliance for Borderline Personality Disorder's Teleconference Series in March 2013. We thank Drs J. Gunderson, C. Lejuez and L. Roemer for

consulting on this project, and A. Cain (now Klein), S. A. Moore, N. Weiss, J. Fulton, R. Brooks, A. Martin and M. Soenke for their invaluable work on this project.

Declaration of Interest

None.

References

Bateman A, Fonagy P (2004). Mentalization-based treatment of BPD. *Journal of Personality Disorders* **18**, 36–51.

Beck AT, Steer RA, Brown GK (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation: San Antonio, TX.

Blanchard EB, Appelbaum KA, Radnitz CL, Michultka D, Morrill B, Kirsch C, Hillhouse J, Evans DD, Guarnieri P, Attanasio V, Andrasik F, Jaccard J, Dentinger MP (1990). Placebo controlled evaluation of abbreviated progressive muscle relaxation and of relaxation combined with cognitive therapy in the treatment of tension headache. *Journal of Consulting and Clinical Psychology* **58**, 210–215.

Borkovec TD, Nau SD (1972). Credibility of analogue therapy rationales. *Journal of Behavior Therapy and Experimental Psychiatry* **3**, 257–260.

Chapman AL, Gratz KL, Brown MZ (2006). Solving the puzzle of deliberate self-harm: the experiential avoidance model. *Behaviour Research and Therapy* **44**, 371–394.

Chou C-P, Yang D, Pentz MA, Hser Y-I (2004). Piecewise growth curve modeling approach for longitudinal prevention study. *Computational Statistics and Data Analysis* **46**, 213–225.

Clifton A, Pilkonis PA (2007). Evidence for a single latent class of Diagnostic and Statistical Manual of Mental Disorders borderline personality pathology. *Comprehensive Psychiatry* **48**, 70–78.

Devilly GJ, Borkovec TD (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry* **31**, 73–86.

Diefenbach GJ, Abramowitz JS, Norberg MM, Tolin DF (2007). Changes in quality of life following cognitive-behavioral therapy for obsessive-compulsive disorder. *Behaviour Research and Therapy* **45**, 3060–3068.

Enders CK (2010). *Applied Missing Data Analysis*. Guilford Press: New York.

Espie CA, Inglis SJ, Harvey L (2001). Predicting clinically significance response to cognitive behavior therapy for chronic insomnia in general medical practice: analysis of outcome data at 12 months posttreatment. *Journal of Consulting and Clinical Psychology* **69**, 58–66.

First MB, Spitzer RL, Gibbon M, Williams JBW (1996). *Structured Clinical Interview for DSM-IV Axis I Disorders – Patient Edition (SCID-I/P, Version 2.0)*. Biometrics Research, New York State Psychiatric Institute: New York.

Fliege H, Kocalevent R, Walter OB, Beck S, Gratz KL, Gutierrez P, Klapp BF (2006). Three assessment tools for deliberate self-harm and suicide behavior: evaluation and psychopathological correlates. *Journal of Psychosomatic Research* **61**, 113–121.

Frisch MB, Cornwell J, Villanueva M, Retzlaff PJ (1992). Clinical validation of the Quality of Life Inventory: a measure of life satisfaction of use in treatment planning and outcome assessment. *Psychological Assessment* **4**, 92–101.

Gratz KL (2001). Measurement of deliberate self-harm: preliminary data on the Deliberate Self-Harm Inventory. *Journal of Psychopathology and Behavioral Assessment* **23**, 253–263.

Gratz KL, Gunderson JG (2006). Preliminary data on an acceptance-based emotion regulation group intervention for deliberate self-harm among women with borderline personality disorder. *Behavior Therapy* **37**, 25–35.

Gratz KL, Levy R, Tull MT (2012). Emotion regulation as a mechanism of change in an acceptance-based emotion regulation group therapy for deliberate self-harm among women with borderline personality pathology. *Journal of Cognitive Psychotherapy* **26**, 365–380.

Gratz KL, Roemer L (2004). Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment* **26**, 41–54.

Gratz KL, Rosenthal MZ, Tull MT, Lejuez CW, Gunderson JG (2006). An experimental investigation of emotion dysregulation in borderline personality disorder. *Journal of Abnormal Psychology* **115**, 850–855.

Gratz KL, Tull MT (2010). Emotion regulation as a mechanism of change in acceptance- and mindfulness-based treatments. In *Assessing Mindfulness and Acceptance: Illuminating the Processes of Change* (ed. R. A. Baer), pp. 107–134. New Harbinger Publications: Oakland, CA.

Gratz KL, Tull MT (2011). Extending research on the utility of an adjunctive emotion regulation group therapy for deliberate self-harm among women with borderline personality pathology. *Personality Disorders: Theory, Research, and Treatment* **2**, 316–326.

Hambrick JP, Turk CL, Heimberg RG, Schneier FR, Liebowitz MR (2004). Psychometric properties of disability measures among patients with social anxiety disorder. *Journal of Anxiety Disorders* **18**, 825–839.

Hardy SA, Thiels C (2009). Using latent growth curve modeling in clinical treatment research: an example comparing guided self-change and cognitive behavioral therapy treatments for bulimia nervosa. *International Journal of Clinical and Health Psychology* **9**, 51–71.

Hayes SC, Strosahl KD, Wilson KG (1999). *Acceptance and Commitment Therapy: An Experiential Approach to Behavior Change*. Guilford Press: New York.

Hayes SC, Strosahl K, Wilson KG, Bissett RT, Pistorello J, Toarmino D, Polusny MA, Dykstra TA, Batten SV, Bergan J, Stewart SH, Zvolensky MJ, Eifert GH, Bond FW, Forsyth JP, Karekla M, McCurry SM (2004). Measuring experiential avoidance: a preliminary test of a working model. *The Psychological Record* **54**, 553–578.

International Society for the Study of Self-injury (ISS) (2007). *Definitional Issues Surrounding our Understanding of*

Self-Injury. Conference proceedings from the annual meeting of the International Society for the Study of Self-injury (http://www.isssweb.org/aboutnssi.php).

Jacobson NS, Truax P (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* **59**, 12–19.

Lee SY, Song XY (2004). Evaluation of Bayesian and maximum likelihood approaches in analyzing structural equation models with small samples. *Multivariate Behavioral Research* **39**, 653–686.

Lejuez CW, Daughters SB, Nowak JA, Lynch T, Rosenthal MZ, Kosson D (2003). Examining the Inventory of Interpersonal Problems as a tool for conducting analogue studies of mechanisms underlying borderline personality disorder. *Journal of Behavior Therapy and Experimental Psychiatry* **34**, 313–324.

Linehan MM (1993). *Cognitive–Behavioral Treatment of Borderline Personality Disorder*. Guilford Press: New York.

Linehan MM, Comtois KA (1996). *Lifetime Parasuicide Count*. University of Washington: Seattle, WA.

Linehan MM, Heard HL (1987). *Treatment History Interview*. University of Washington: Seattle, WA.

Lovibond SH, Lovibond PF (1995). *Manual for the Depression Anxiety Stress Scales*, 2nd edn. The Psychology Foundation of Australia: Sydney.

Muthén LK, Muthén BO (1998–2010). *Mplus User's Guide*, 6th edn. Muthén & Muthén: Los Angeles, CA.

Nock MK (2009). Why do people hurt themselves? New insights into the nature and functions of self-injury. *Current Directions in Psychological Science* **18**, 78–83.

Ogles BM, Lunnen KM, Bonesteel K (2001). Clinical significance: history, application, and current practice. *Clinical Psychology Review* **21**, 421–446.

Pfohl B, Blum N, St John D, McCormick B, Allen J, Black DW (2009). Reliability and validity of the Borderline Evaluation of Severity over Time (BEST): a self-rated scale to measure severity and change in persons with borderline personality disorder. *Journal of Personality Disorders* **23**, 281–293.

Roemer L (2001). Measures of anxiety and related constructs. In *Practitioner's Guide to Empirically Based Measures of Anxiety* (ed. M. M. Antony, S. M. Orsillo and L. Roemer), pp. 49–83. Kluwer Academic/Plenum: New York.

Roemer L, Orsillo SM (2007). An open trial of an acceptance-based behavior therapy for generalized anxiety disorder. *Behavior Therapy* **38**, 72–85.

Safer DL, Telch CF, Chen EY (2009). *Dialectical Behavior Therapy for Binge Eating and Bulimia*. Guilford Press: New York.

Sansone RA, Wiederman MW, Sansone LA (1998). The Self-Harm Inventory (SHI): development of a scale for identifying self-destructive behaviors and borderline personality disorder. *Journal of Clinical Psychology* **54**, 973–983.

Selby EA, Joiner TE (2009). Cascades of emotion: the emergence of borderline personality disorder from emotional and behavioral dysregulation. *Review of General Psychology* **13**, 219–229.

Sheehan DV (1983). *The Anxiety Disease*. Charles Scribner and Sons: New York.

Thayer JF, Lane RD (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders* **61**, 201–216.

Tyrer P, Tom B, Byford S, Schmidt U, Jones V, Davidson K, Knapp M, MacLeod A, Catalan J; POPMACT Group (2004). Differential effects of manual assisted cognitive behavior therapy in the treatment of recurrent deliberate self-harm and personality disturbance: the POPMACT Study. *Journal of Personality Disorders* **18**, 102–116.

Zanarini MC (2003). Zanarini Rating Scale for Borderline Personality Disorder (ZAN-BPD): a continuous measure of DSM-IV borderline psychopathology. *Journal of Personality Disorders* **17**, 233–242.

Zanarini MC (2009). Psychotherapy of borderline personality disorder. *Acta Psychiatrica Scandinavica* **120**, 373–377.

Zanarini MC, Frankenburg FR, Sickel AE, Young L (1996). *Diagnostic Interview for DSM-IV Personality Disorders*. McLean Hospital: Boston, MA.

Zhang Z, Hamagami F, Wang L, Nesselroade JR (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development* **31**, 374–383.