Psychotherapy Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tpsr20

Sudden gains and large intersession improvements in internet-based psychodynamic treatment (IPDT) for depressed adolescents

Jakob Mechler , Karin Lindqvist , Fredrik Falkenström , Per Carlbring , Gerhard Andersson & Björn Philips

To cite this article: Jakob Mechler , Karin Lindqvist , Fredrik Falkenström , Per Carlbring , Gerhard Andersson & Björn Philips (2020): Sudden gains and large intersession improvements in internet-based psychodynamic treatment (IPDT) for depressed adolescents, Psychotherapy Research, DOI: 10.1080/10503307.2020.1804084

To link to this article: https://doi.org/10.1080/10503307.2020.1804084

Published online: 17 Aug 2020.

Psychotherapy Research, 2020
https://doi.org/10.1080/10503307.2020.1804084

EMPIRICAL PAPER

Sudden gains and large intersession improvements in internet-based psychodynamic treatment (IPDT) for depressed adolescents

JAKOB MECHLER [1], KARIN LINDQVIST [1], FREDRIK FALKENSTRÖM [2], PER CARLBRING [1], GERHARD ANDERSSON [23], & BJÖRN PHILIPS [1]

[1]*Department of Psychology, Stockholm University, Stockholm, Sweden;* [2]*Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden &* [3]*Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden*

(Received 7 April 2020; revised 20 July 2020; accepted 26 July 2020)

ABSTRACT

Objective: Sudden gains (SGs) have often been found associated with better treatment outcome across different psychiatric disorders. However, no studies have evaluated SGs in internet-based treatment targeting adolescent depression. **Method:** The sample consisted of 66 adolescents diagnosed with major depressive disorder, attending psychodynamic internet-based treatment. Effects of SGs were evaluated at posttreatment and 6-month follow-up. We also evaluated effects of large intersession improvements (LIIs; sudden and relatively large gains, between sessions, without the stability criterion). Effects of SGs and LIIs early in treatment were also investigated. **Results:** A total of 17 patients (25.75%) experienced an SG. The effect of having an SG or early SG was non-significant after treatment ($d = 0.48$) and at follow-up ($d = 0.66$). However, having an LII was related to better outcome after treatment ($d = 0.97$) and at follow-up ($d = 0.76$). Early LIIs were associated with significantly better results at end of treatment ($d = 0.72$). **Conclusions:** The original criteria of SGs might be overly conservative and thus miss important improvements in depression. Relatively large intersession gains, regardless of stability, seem to be predictive of outcome.

Keywords: depression; process research; psychoanalytic/psychodynamic therapy

Clinical or methodological significance of this article: The original sudden gain criteria might be overly conservative and thus lead to important shifts in symptoms being overlooked. Clinicians should be aware that the road to recovery in adolescent depression is often non-linear, rather it is in many cases characterized by relatively large intersession improvements. Some patients might even deteriorate slightly before an actual gain. Patients experiencing large intersession improvements during treatment exhibit significantly better outcomes compared to patients with no such improvements in an internet-based psychodynamic treatment of adolescent depression.

Clinical trials registry: International Standard Randomised Controlled Trial Number (ISRCTN) 16206254; http://www.isrctn.com/ISRCTN16206254.

Understanding processes of change in psychotherapy is essential to further comprehend, develop, and improve treatments. One of these processes is the pattern of change over time. The trajectory of change within the treatment of psychopathology is highly individual. For some, it is a process of gradual gains, but for others, the path is a process of sudden and relatively large improvements between consecutive sessions (Shalom & Aderka, 2020). These large and rapid changes have been named *sudden gains* (SGs; Tang & DeRubeis, 1999). SGs may be related to critical events in therapy as well as to outcome. Tang and DeRubeis (1999) devised a list of criteria for what qualifies as

*Correspondence concerning this article should be addressed to Björn Philips, Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden. Email: bjorn.philips@psychology.su.se

an SG: (a) the reduction in symptoms must be large in absolute terms, (b) it must be a reduction of at least 25% compared to the symptom rating made before the SG, and (c) the reduction must be stable (e.g., the mean symptom level must be significantly higher in the three sessions before the SG compared to the mean of the three sessions that came after). In their original paper, Tang and DeRubeis (1999) found that SGs were both common (occurring in more than half of all treatment responders) and that they constituted a large proportion of all changes (accounting for more than 50% of the treatment's total effect). SGs were also found to be robust, with only a 17% reversal rate, and related to a more positive treatment outcome regarding both symptom levels and recovery, as well as less relapse after treatment (Tang & DeRubeis, 1999). If SGs are a critical part of psychotherapy improvement, the concept is highly relevant to research further, and there is a need to investigate predictors and SG facilitators.

In recent years, research on SGs has increased dramatically. Meta-analytic results (Shalom & Aderka, 2020) indicate that SGs occur in between 14.3% and 62.2% of patients in psychotherapy and that they have a moderate positive effect on outcome at end of treatment (Hedges' $g = 0.68$) and follow-up (Hedges' $g = 0.61$). Earlier research implied that SGs are particularly potent in CBT-based interventions (Aderka et al., 2012), but the most recent meta-analysis indicates a more general effect (Shalom & Aderka, 2020). The between group effect size of SGs vs. no SGs in CBT was medium to large (Hedges' $g = 0.72$), while the between group effect size in non-CBT interventions was medium (Hedges' $g = 0.57$). SGs also predicted secondary outcome measures, albeit to a much smaller extent (mean effect size $g = 0.38$). After correcting for potential bias, the effect size was reduced to $g = 0.32$ but remained statistically significant. Thus, SGs occur and seem to be associated with a range of outcomes across different treatment modalities.

To date, there are only four studies that specifically investigate SGs in psychodynamic psychotherapy (PDT; Brockmeyer et al., 2019; Present et al., 2007; Shalom et al., 2018; Tang et al., 2002), and one study that examine change mechanisms related to SGs (Andrusyna et al., 2006). Shalom et al. (2018) found that SGs predicted symptom reduction in a sample of 106 mixed-disorder participants receiving PDT. Brockmeyer et al. (2019) found that SGs in both CBT and PDT for anorexia were related to better outcomes at post-treatment as well as follow-up. Patients with SGs during the first third of treatment exhibited better outcomes during treatment and a trend of better outcomes during follow-up than patients with SGs later in treatment. Present et al. (2007) noted no relation between SGs and outcomes in PDT for generalized anxiety disorder. Tang et al. (2002), reported that SGs were almost as prevalent in PDT as in CBT when treating depression. Although predictive of outcomes in both treatments, SGs in CBT predicted outcome to a higher degree, had a lower rate of symptom reversal, and predicted outcomes six months after treatment, which was not the case in PDT.

In general, most studies on SGs have been conducted on adult samples, but according to analyses by Shalom and Aderka (2020), effects of SGs do not seem to differ between adults, children, and adolescents. Gaynor et al. (2003) is to our knowledge the only study investigating the importance of SGs in the treatment of adolescent depression. They found that individual CBT and nondirective supportive therapy (NST) led to a higher frequency of SGs compared to systemic behavioral family therapy (SBFT). In total, 39% of the partaking adolescents experienced SGs (CBT: 50%; SBFT: 26%; NST: 39%). SGs were related to better outcome on both self-reported and observer-rated measures of depression at the end of treatment. Interestingly, in participants not experiencing SGs, CBT led to superior outcome compared to SBFT and NST. The authors suggest that this might be due to processes unique to CBT driving positive change in the group of patients not experiencing an SG.

SGs in Internet-Based Treatments

A common way to deliver psychological treatment is to use modern information technology and the internet. Today numerous trials support the efficacy of internet-delivered treatment for a wide array of psychiatric conditions (Andersson, 2016). Hedman et al. (2014) found that SGs significantly predicted outcome in a treatment study on health anxiety using internet-based CBT. This was true both at treatment termination and at six and twelve month follow-ups. Patients exhibiting SGs were also significantly less symptomatic, at post-treatment and at follow-ups, compared to so-called gradual gainers. Recently, these findings have been corroborated and extended with longer follow-ups in patient populations suffering from body dysmorphic disorder (Bjureberg et al., 2019) and obsessive compulsive disorder (Hamdeh et al., 2019). To our knowledge, no studies exist on SGs in internet-based treatments targeting adolescent psychopathology.

Mechanisms of SGs

Numerous studies have investigated potential mechanisms of SGs. Tang and DeRubeis (1999) identified

that changes in cognitions preceded SGs and therefore proposed this as a mechanism of SGs in CBT. They also proposed three stages: the preparation stage in which the therapists use CBT techniques, leading to the second stage where changes in cognition result in a subsequent sudden gain in symptoms. Finally, in stage three, this gain creates an upward spiral, preserving changes made and laying the foundation for further improvement, possibly through an interaction between cognitive changes and a strengthened working alliance leading to further cognitive changes and symptomatic improvement. Later studies have shown mixed results regarding cognitive changes and SGs, where some have corroborated these findings (Tang et al., 2005; Vincent & Norton, 2019), while others have failed to do so (e.g., Kelly et al., 2005). Abel et al. (2016) found that CBT-therapists' case conceptualization skills as well as patient hopefulness were related to SGs in patients suffering from treatment-resistant depression. Interestingly, one study on SGs in PDT suggests there may be different mechanisms driving SGs depending on psychotherapy method. Andrusyna et al. (2006) discovered that cognitive changes were unrelated to SGs in PDT. Instead, they found that the level of accuracy in therapist interpretations during pre-gain sessions was significantly related to whether patients experienced an SG or not. They also found that the therapeutic alliance was predictive of SGs on a trend level. These results seem to imply that the mechanisms of SGs might differ between different treatment modalities. The role of the alliance in SGs has been further explored in a recent study by Zilcha-Mano et al. (2019). According to their results, SGs seem to lead to subsequent strengthening of the alliance, which in turn predict improvements in life satisfaction and psychological functioning. Thus, their findings imply that the alliance may be of importance in sustaining the effects of SGs over time. To our knowledge, no studies exist examining mechanisms of SGs in internet-based treatments.

SGs have also been investigated in placebo-controlled pharmacological trials. These studies seem to suggest that the frequency of SGs are similar among patients receiving antidepressants and placebo (e.g. Vittengl et al., 2005; Zilcha-Mano et al., 2018). Gaynor et al. (2003) found no significant difference between the number of SGs in CBT or NST. These findings challenge the notion that SGs are unique for psychotherapy and a result of modality-specific psychotherapeutic techniques. Gaynor et al. (2003) argue that, since consistent mediators of SGs have been hard to establish, it is unlikely that the lack of difference between different treatment modalities stem from unique, treatment-specific mechanisms.

The Role of Emotion Regulation

Research suggests that emotion dysregulation is associated with both symptoms of anxiety and depression in adolescents (Schäfer et al., 2017). Results from Gonçalves et al. (2019) suggest that emotion regulation deficits in early adolescence are associated with depressive symptoms both cross-sectionally and over time. It has also been found that increasing the capacity of adaptive emotion regulation acts as a mechanism of change across different psychiatric disorders and through different treatment modalities (Berking et al., 2019; Bjureberg et al., 2018; Wirtz et al., 2014). An increased capacity for affect regulation has also been described as a key focus in affect-focused psychodynamic psychotherapy (Frederickson et al., 2018) and furthermore seem to work as a time-varying predictor of subsequent changes in depression in the internet-based psychodynamic treatment used in this study (Mechler et al., 2020). A study of psychotherapy for women with alcohol use disorder showed that women with low confidence in their capacity to regulate emotions without using alcohol were more inclined to experience SGs after sessions targeting emotion regulation in relation to anxiety and depression (Holzhauer et al., 2017). However, to our knowledge no studies exist examining whether increases in the capacity for emotion regulation drives sudden gains in psychotherapy. As an enhanced capacity for emotion regulation has been found mediating outcome in many different treatment modalities and treatment formats we wanted to explore if gains in emotion regulation might precede SGs.

Different Definitions of SGs

The field of SGs has been criticized for suffering from tautology, rather than constituting a therapeutic factor. For SGs to influence outcomes, one would expect they lead to a higher improvement rate after the gain. However, Koffmann (2019) did not find a larger improvement after the gain. In contrast, he identified that better outcomes for patients receiving an SG was accounted for largely by the gain itself—essentially meaning that "patients that have better outcomes have better outcomes." This has led to a discussion regarding whether SGs are more important than other types of gains, such as gradual gains or other gains not fulfilling all three criteria by Tang and DeRubeis (1999). While some studies have found SGs to be a better predictor of outcome than so called gradual gains (Bjureberg et al., 2019[; Hamdeh et al., 2019; Hedman et al., 2014) the question about what the best definition of an SG is remains, as research implies that

studies adhering to the original criteria actually have slightly smaller effects than studies using slightly changed definitions (Shalom & Aderka, 2020).

In fact, discussions regarding SG criteria have been present since the concept was introduced by Tang and DeRubeis (1999). The first criterion "The gain must be large in absolute terms" was initially defined as an improvement on the Beck Depression Inventory (BDI; Beck et al., 1961) of ≥ 7 points. Tang and DeRubeis (1999) themselves noted that this criterion was a somewhat arbitrary cutoff score. As such, the criterion has been problematic, since it is instrument-dependent. Therefore, different strategies have been employed to solve these problems. Some researchers have defined criterion one using reliable change index (Jacobson & Truax, 1991; Stiles et al., 2003), while others have used validated conversions between BDI and other measures (e.g., Hunnicutt-Ferguson et al., 2012). The second criterion: "The gain must be large in relative terms" is defined as a 25% drop in symptoms. This too has been criticized for being "overly relative," meaning the sensitivity for 25% drops is largely dependent on the scale. Hardy et al. (2005) noted that this criterion assumes that symptom scales are on a ratio level, while in fact they are, at best, at an interval level. Furthermore, Tang et al. (2005) discovered this criterion was of little contribution in identifying gains, suggesting it could possibly be omitted. However, Koffmann (2019) discovered that 25% reductions in symptoms was the single criterion most predictive of outcome.

The third criterion of stability originally requires the mean score of the three pre-gain sessions to be larger than the mean score of the three post-gain sessions, shown with a two-sample *t*-test. This strict definition prevents any SGs from being classified during the first or last two sessions; therefore, it is often modified by decreasing the number of measurement points required (Stiles et al., 2003). Also, the number of sessions chosen for the stability criterion is arbitrary (Vittengl et al., 2015). Another critique against the original third criterion is that by definition, it violates a fundamental assumption of the two-sample *t*-test, since repeated measurements made by the same individual almost always are correlated. This has therefore been reworded into more appropriate statistical terms by Andrusyna et al. (2006) as follows: The mean difference between the scores of the three sessions before the gain and the three sessions after the gain must be at least 2.78 times greater than the pooled standard deviations of these two sets of sessions. This modification, however, is only in wording and not in the calculations, while other researchers have gone further and modified criterion three as well as discussed its usefulness. For example, Kelly et al. (2005) substituted this criterion with the requirement that SGs be 50% larger than the standard deviation of the individual's scores across all sessions.

Koffmann (2019) assessed the predictive power of different types of intersession improvements, finding that 25% improvements (SG's second criterion) predicted outcome to a higher degree than full SGs fulfilling all three criteria. Using only criterion one, Reliable Change Index (RCI), also predicted outcome, but to a lower degree than both 25% improvements and full SGs. Koffmann (2019) therefore argued that the empirical basis for including stability in the definition might be lacking, as it lowers SG's predictive power and limits potentially interesting symptomatic variability before and after SGs.

The aim of the present study was to investigate the presence and importance of SGs, with and without the criterion of symptom stability, in affect-focused, internet-based psychodynamic treatment for adolescents suffering from depression (Lindqvist et al., 2020). This research is important because existing literature mostly focuses on adults (Shalom & Aderka, 2020). To our knowledge, there are no studies investigating SGs in internet-based psychodynamic treatments, and this is only the second study investigating SGs in the treatment of adolescent depression (Gaynor et al., 2003). Given Koffmann's (2019) findings, that including symptom stability in the SG criteria lessened the predictive power of intersession improvements, we also aimed to compare the association between outcome and SGs versus large intersession improvements (LIIs). The latter is defined as improvements fulfilling criteria one and two, as described by Tang and DeRubeis (1999), meaning large SGs between consecutive sessions but without the stability criterion. Further, we aim to examine whether significant pre-gain changes in improved emotion regulation appears as a mechanism of SGs.

Methods

Overview

The clinical trial was undertaken by Stockholm University in collaboration with Linköping University. The ISRCTN (International Standard Randomised Controlled Trial Number) registration ID is 16206254. The trial was approved by the Regional Ethics Board of Stockholm, Sweden (number: 2018/2268-31/5). Participants submitted written informed consent via the online treatment platform and received the treatment at no cost. Parental consent was not mandatory as this is not

required by Swedish law from the age of 15. However, all participants under the age of 18 were required to give contact details to one legal guardian.

Design

This study employed a within-group design with repeated measurements in a sample of adolescents suffering from major depression ($N = 66$) that received treatment within a randomized controlled trial (RCT) evaluating the efficacy of an internet-based psychodynamic treatment (Lindqvist et al., 2020). The control group was crossed over to treatment after ten weeks in a control condition. The present study contained all available participants who entered treatment and contributed with data for at least three timepoints, including both participants who immediately entered treatment and those who entered treatment 10 weeks later. Assessments were made before treatment (pre-treatment), weekly during treatment, immediately after treatment (post-treatment), and at six months post-treatment follow-up. During treatment, the primary outcome measure (Quick Inventory of Depressive Symptomatology for Adolescents; QIDS-A17-SR; Bernstein et al., 2010) and a process measure (Difficulties in Emotion Regulation Scale; DERS-16; Bjureberg et al., 2016) were administered weekly. Diagnostic interviews (MINI v. 7.0; Sheehan et al., 1998) were conducted pre-treatment by trained psychologist students at the end of their clinical training or by experienced clinical psychologists, in order to establish a primary depression diagnosis. Overall main treatment results have been published elsewhere (Lindqvist et al., 2020).

Participants

Recruitment was performed during January and February 2019. Adolescents were recruited through social media, as well as via schools, youth centers, and youth mental health care providers. To be included, adolescents had to be 15–18 years old, fulfill diagnosis of unipolar major depressive disorder according to DSM-5 criteria as established by scoring ≥ 10 points on the QIDS-A17-SR (Bernstein et al., 2010), and fulfill criteria for major depressive disorder according to MINI 7.0 (Sheehan et al., 1998). Exclusion criteria included prior suicide attempts and/or expressing substantial suicidality during intake, partaking in other psychological treatments, psychotropic medication not stable for at least three months, other primary diagnosis/diagnoses in need of other treatment, and current fulfillment of any of the following diagnoses: any psychotic disorder, bipolar I/II disorder, antisocial personality disorder, autism-spectrum disorder, or any substance use disorder. Participants demographics are presented in Table I.

Instruments

The primary outcome measure was QIDS-A17-SR, a reliable, self-rated measure of depressive symptoms validated for both adults and adolescents (Bernstein et al., 2010; Rush et al., 2003). The instrument's reliability has been reported in the main outcome study (Lindqvist et al., 2020). Using available data from all time points, an average Cronbach's alpha of $\alpha = .76$ (range: .71–.85) was found, suggesting an acceptable internal consistency. Assessments were made via internet-delivered, self-rated forms pre-treatment, weekly during treatment, and post-treatment. Furthermore, all participants entering

Table I. Demographic data at baseline.

Demographics:	SG ($n = 17$)		Non-SG ($n = 49$)		LII ($n = 34$)		Non-LII ($n = 32$)	
	n/M	%/SD	n/M	%/SD	n/M	%/SD	n/M	%/SD
Female	14	82.4	41	83.7	29	85.3	26	81.3
Gender identity uncertain/other	1	5.9	2	4.1	2	5.9	1	3.1
Age	16.65	1.27	16.61	1.06	16.74	1.19	16.60	1.02
Major depressive disorder[a]	17	100	49	100	34	100	32	100
Any anxiety disorder[a]	12	70.6	27	55.1	22	64.7	17	53.1
PTSD[a]	2	11.8	2	4.1	4	11.8	0	0
Eating disorder[ab]	0	0	3	6.1	1	2.9	2	6.3
QIDS-A17-SR pre-treatment	14.06	5.03	14.73	4.16	14.65	4.84	14.47	3.89
MADRS-S pre-treatment	23.59	9.91	26.61	6.53	25.21	8.18	26.50	6.96
GAD-7 pre-treatment	11.24	5.13	11.8	3.80	11.62	4.26	11.69	4.09

Note. QIDS-A17-SR = Quick inventory of depressive symptomatology adolescent self-rated version; MADRS-S = Montgomery Åsberg Depression Rating Scale–self-rated; GAD-7 = Generalized Anxiety Disorder 7-item scale. No significant between-group differences on any of the data at the baseline. [a]Confirmed by the MINI-International Neuropsychiatric Interview at baseline of the RCT. [b]Bulimia nervosa/Binge-eating disorder.

treatment were assessed six months after treatment termination.

The DERS-16 (Bjureberg et al., 2016) was used as a process measure, administered weekly as well as pre- and post-treatment. Lindqvist et al. (2020) reported a good internal consistency ($\alpha = .89$). Secondary outcome measures used in the present study were the Generalized Anxiety Disorder 7-item scale (GAD-7; Kroenke et al., 2010) and the Montgomery Åsberg Depression Rating Scale self-rated version (MADRS-S; Svanborg & Åsberg, 1994). Lindqvist et al. (2020) found that both measures had acceptable to good internal consistency ($\alpha = .76$ and .83, respectively). Both were administered pre- and post-treatment. GAD-7 was also administered at the six-month follow-up but was not used in the present study. MADRS-S has recently been evaluated for use within adolescent populations (Ntini et al., 2020).

Intervention

The intervention consisted of eight self-help modules delivered weekly over eight weeks on a secure online platform (Vlaescu et al., 2016). Modules consisted of texts, videos, and exercises, which participants sent to their therapists for feedback. To reduce attrition and increase motivation, the intervention also contained a weekly 30-minute text chat session between therapist and participant. Due to limited resources, only the first group of patients ($n = 34$) received chat sessions. The rest of the patients ($n = 32$) received the exact same treatment but without additional synchronous chat sessions.

The IPDT program was developed specifically for this clinical trial (Lindqvist et al., 2020) and based on similar principles as a treatment program targeting adult populations suffering from depression and anxiety (Johansson et al., 2013, 2017; Zwerenz et al., 2017). The aim of IPDT is to decrease emotional avoidance and increase awareness and experience of emotions. The treatment program emphasizes the link between experiential avoidance (through defenses) and symptoms of depression and anxiety. The final part of the program contains material on how to understand pervasive maladaptive relational patterns and communicate affects in close relationships.

Therapists

Project therapists were clinical psychologists ($n = 2$) or clinical psychology students in the last semesters of their psychologist training ($n = 9$). All therapists specialized in PDT during clinical training and thus took courses in PDT theory and practice. Student therapists received a one-day training session by the treatment developers (JM and KL) and treated most participants ($n = 61$). All therapists were supervised weekly in groups of 5–6 for 90 min by an experienced psychotherapist specializing in experiential dynamic psychotherapy.

Definition of SGs With or Without Symptom Stability

SGs were calculated on the QIDS-A17-SR. For criterion one, we used a cutoff of 4 points, in line with previous research on SGs using the QIDS-SR16 (Hunnicutt-Ferguson et al., 2012), where this cutoff was described as equivalent to the original cutoff on the BDI-II. We also adhered to criterion two, meaning that improvement from the pre-gain session had to be 25% or more to warrant classification as an SG. Criterion three stipulates that the gain must represent a stable reduction in contrast to symptom fluctuations. In calculating this criterion, we adhered to a slightly modified version introduced by Tang et al. (2002). We further adapted criterion three according to the work of Stiles et al. (2003). This means we calculated SGs also when there were only two measure points available on one side of the gain (either due to missing data or the SG appearing early or late during treatment). When three measure points existed, all were used. Reversal of an SG was defined as an increase in symptoms corresponding to $\geq 50\%$ of the gain occurring after the period of post-gain stability (3 sessions). When calculating LIIs, we adhered to the exact same criteria except for criterion three, meaning that improvements had to be ≥ 4 points on QIDS-A17-SR and reductions had to be at least 25% of the pre-gain session.

Statistical Analysis

SGs and LIIs were calculated using R 3.5.0 (R Core Team, 2019), and the R package suddengains v. 0.2.1: An R package to identify SGs in longitudinal data (Wiedemann et al., 2020). The same package was used for extracting scores on DERS-16 around the period of each gain.

Statistical analyses were conducted using SPSS version 25 (IBM Corp., Armonk, NY). In line with previous studies on SGs, missing data during the treatment phase was not imputed, since that increases the risk of overestimating SG frequency (Hedman et al., 2014). Analyses were made using a linear mixed effects model framework. Level-1 residuals were assumed independent and identically distributed. At Level-2, random effects were assumed

independent but with different variances. We also tried an unstructured covariance structure for level 2, but as it led to a reduced model fit (increased Akaike's Information Criterion by 2 points), we decided to retain the more parsimonious model. Full Maximum Likelihood Estimation was used in all analyses. Model building started with estimating a basic time model, which included random intercepts as well as fixed and random slopes for Time. Time was coded 0 for pre-treatment, 1–8 for weekly assessments during treatment, and 9 for the post-treatment assessment point. To account for possible non-linearity in the data, a quadratic term for Time (Time × Time) and a cubic term (Time × Time × Time) were also tested but later discarded, as neither reached significance nor improved model fit. SG/LII status (coded 0 for no and 1 for yes) was entered both as a main effect (to test for possible differences between the groups at pre-treatment assessment), as well as in interaction with Time (to test for group differences in change rates over time). We also added group (with or without chat) as a main effect and in interaction with Time to test if group allocation might affect our results. As the interaction term (Group × Time) did not change our results regarding SGs or LIIs this term was dropped from the analysis. To further test for effects from group allocation we compared the number of SGs/LIIs experienced in the two groups by conducting $\mathcal{X}^2$ tests.

Since the treatment period was short, a possible risk was that SG or LII late in treatment would account for outcome effects, since the risk of reversals decrease later in treatment. Therefore, as a validation of the constructs, we also assessed SGs and LIIs early in treatment (up to session four). Effect sizes on all multilevel analyses were calculated using observed standard deviations at baseline (Feingold, 2009).

Analyses on primary outcome at six month follow-up and secondary measures were conducted using ANCOVA, controlling for pretreatment severity on the respective measure. Missing data at post-treatment (GAD-7, $n = 8$, 12.1%; MADRS-S, $n = 9$, 13.6%; QIDS-A17-SR, $n = 4$, 6.1%) were handled through multiple imputation. A total of 50 imputed datasets were created (e. g., Jakobsen et al., 2017). To assess existence of significant changes in emotion regulation between timepoints (e.g., pre-gain sessions) preceding SGs or LIIs, we used a pairwise t-test on DERS-16 total scores.

Results

Table II presents observed values for QIDS-A17-SR and DERS-16 across treatment.

Frequency of Sudden Gains

A total of 17 patients (25.75%) fulfilled all three criteria for one SG each. Twelve of these occurred in the first half of treatment, that is the treatment period's first four weeks. Nine of the early SGs occurred in weeks 2 or 3. The mean SG was 5.18 points (SD = 1.29, range: 4–8) on the QIDS-A17-SR. Out of these 17 patients, 8 (47.06%) had a reversal (≥50% of the gain) before their treatment ended. Figure 1 describes the average QIDS-A17-SR score three sessions before and after the SG. Group allocation (chat or no chat) did not affect the frequency of experienced SGs, $\mathcal{X}^2$ (1, $n = 66$) = .18, $p = .68$.

Frequency of Large Intersession Improvements

When not adhering to criterion three (symptom stability), a total of 34 patients (51.52%) showed at least one LII. Nine participants (13.64%) had multiple LIIs, resulting in 45 occurrences in total. Out of the 34, 14 participants (41.18%) had a reversal (≥50%

Table II. Observed means, standard deviations, and number of observations for outcome and processes over the treatment period.

Measure	Week									
	0	1	2	3	4	5	6	7	8	9
QIDS-A17-SR										
M	14.56	14.48	13.55	12.89	12.13	11.78	11.94	11.60	10.59	9.46
SD	4.37	3.76	4.6	4.52	4.14	4.20	4.31	4.60	5.02	4.98
n	66	66	65	57	55	55	50	55	46	59
DERS-16										
M	55.45	55.14	55.14	53.18	51.33	50.02	48.40	45.73	44.82	41.02
SD	11.75	11.70	12.13	11.66	11.84	12.26	11.62	12.27	13.86	14.21
n	66	66	65	57	55	55	50	55	45	57

Note. QIDS-A17-SR = Quick inventory of depressive symptomatology adolescent self-rated version; DERS-16 = Difficulties in emotion regulation scale brief version.

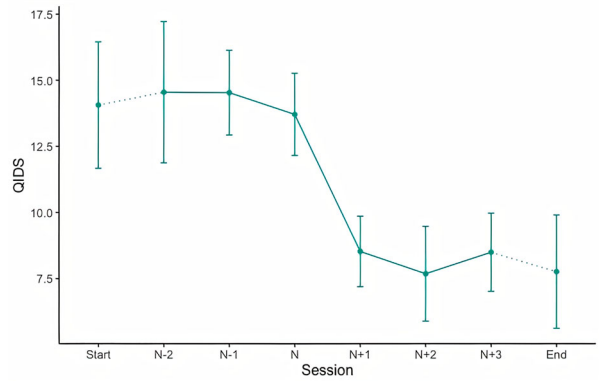

Figure 1. Trajectory of change in depressive symptoms for patients experiencing SGs with 95% confidence intervals for all time points.

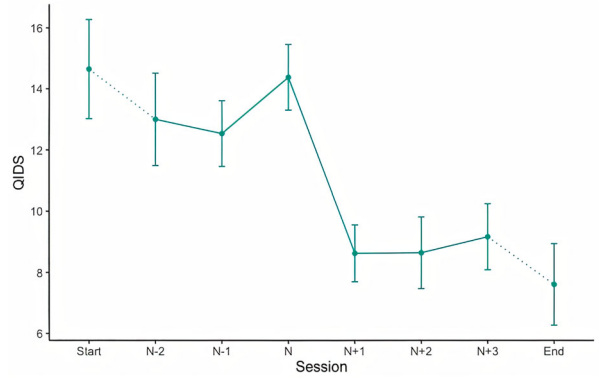

Figure 2. Trajectory of change in depressive symptoms for patients experiencing LIIs with 95% confidence intervals for all time points.

of the intersession gain) before their treatment ended. The mean LII was 5.76 points (SD = 2.13, range: 4–16) on the QIDS-A17-SR. Figure 2 describes the average QIDS-A17-SR score three sessions before and after an LII. The average LII was preceded by a significant worsening of symptoms between the two weeks right before the LII occurred (mean diff = 1.84; $t(44) = 3.0504$; $p < .01$). However, as shown in Figure 2, the drop of symptoms far exceeds a mere return from the worsening of symptoms. Group allocation (chat or no chat) did not affect the frequency of experienced LIIs, $\mathcal{X}^2$ (1, $n = 66$) = .96, $p = .33$.

Association Between SGs, LIIs and Primary Outcome

At post-treatment, mixed effects model analysis did not show a significant effect regarding the interaction effect of group (non-SG vs. SG) and time on QIDS-A17-SR, ($F_{(1,62)} = 3.22$, $p = .08$), indicating a non-significant difference in improvements between subjects with SGs and non-SGs. The between group effect size was $d = 0.48$, 95% CI [−0.06, 1.02].

Regarding LIIs, a significant interaction effect of group (LII vs. non-LII) and time on QIDS-A17-SR ($F_{(1,67)} = 20.03$, $p < .001$) was found, indicating superior improvements in the LII group compared to the non-LII group. The between group effect size, comparing LIIs to non-LIIs was $d = 0.97$, 95% CI [0.54, 1.50].

Early SGs ($n = 12$) showed no significant interaction effect of group (early SG vs. non-early SG) and time on QIDS-A17-SR ($F_{(1,61)} = 1.28$, $p = .26$). The between group effect size was d = 0.35, 95% CI [−0.27, 0.97].

Early LIIs ($n = 22$) showed a significant interaction effect of group (early LII vs. non-early LII) and time on QIDS-A17-SR ($F_{(1,62)} = 8.60$, $p = .01$), indicating superior improvements in the early LII group compared to those who did not have an early LII. The between group effect size was $d = 0.72$, 95% CI [0.23, 1.21].

Difficulties in Emotion Regulation and SGs/LIIs

There were no significant changes in emotion regulation between the two weeks occurring right before the SG or LII. This was true for both SGs (mean diff = −1.71; $t(16) = -0.85$; $p = .41$) and LIIs (mean diff = 0,8; $t(44) = 0.55$; $p = .59$). Figures 3 and 4 describe the average DERS-16 score of sessions before and after an SG or an LII, respectively.

Associations Between SGs, LIIs, and Secondary Outcome

There was no significant effect of SGs on MADRS-S ($F_{(2,63)} = 0.925$, $p = .336$) or on GAD-7 ($F_{(2,63)} = 0.473$, $p = .492$). In contrast, there was a significant

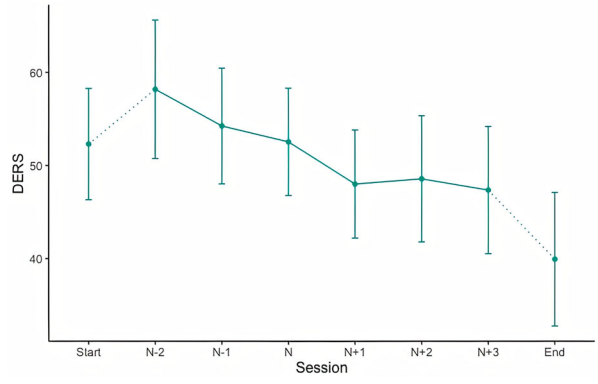

Figure 3. DERS-16 for patients experiencing SGs with 95% confidence intervals for all time points.

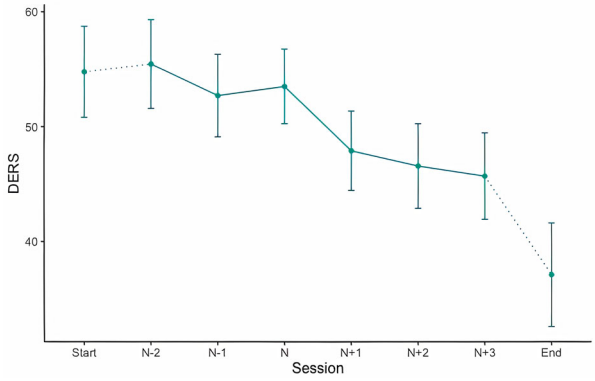

Figure 4. DERS-16 for patients experiencing LIIs with 95% confidence intervals for all time points.

effect of LIIs on MADRS-S ($F_{(2,63)} = 5.626$, $p = .02$, $d = 0.64$, 95% CI [0.111, 1.173]). There was no significant effect of LIIs on GAD-7 ($F_{(2,63)} = 2.93$, $p = .09$).

Association Between SGs, LIIs, and Follow-up

There was no significant effect of SGs on the six month follow-up on QIDS-A17-SR ($F_{(2,63)} = 3.24$, $p = .07$, $d = 0.66$, 95% CI [–0.06, 1.37]). In contrast, there was a significant effect of LIIs on the six month follow-up on QIDS-A17-SR ($F_{(2,63)} = 11.062$, $p = .02$, $d = 0.76$, 95% CI [0.109, 1.414]).

Discussion

The aim of the present study was to investigate the presence and importance of SGs and LIIs in affect-focused, internet-based psychodynamic psychotherapy for adolescents suffering from depression. To our knowledge, this is the first study investigating SG occurrence in IPDT and only the second study concerning SGs in treatment of adolescent depression (Gaynor et al., 2003). Approximately one quarter of all patients entering this treatment experienced an SG. This is somewhat lower than the frequency of 34.65% reported by Shalom and Aderka (2020), but the treatment was also substantially shorter than most treatments in that review. The relatively large amount of significance tests conducted in the present study could lead to spurious findings as it increases the risk for type I error. Given that the sample size was rather small, this needs to be taken into account when interpreting the results.

Patients experiencing SGs during treatment did not present significantly superior outcome compared to patients not experiencing SGs ($p = .08$). The lack of statistical significance could be due to a relatively small sample size, but all the same, results from the present study failed to replicate earlier studies that have found SGs being predictive of outcome. Another explanation might be the relatively high degree of symptom reversal (i.e. an increase in symptoms corresponding to $\geq 50\%$ of the gain occurring after the period of post-gain stability). Compared to earlier studies on SGs and psychodynamic psychotherapy, SG frequency is somewhat lower in the present study, but the reversal rate is similar. Present et al. (2007) and Tang et al. (2002) reported reversal rates of 40% and 46% respectively whilst the reversal rate in the present study was 47%. Shalom and Aderka (2020) found an average reversal rate of 31.49% (range: 7.7–100%), and the frequency of symptom reversal was associated with smaller effects of SGs on outcome. Thus, the reversal rates in the present study may have reduced the impact of the effect of SGs on outcome. However, Storch et al. (2019), investigating SGs in CBT for obsessive compulsive disorder among children and adolescents, reported that reversals were frequent after sudden gains, especially in the presence of depression, but that reversals did not predict poor prognosis. It should be noted that from a methodological perspective, the definition of what constitutes a reversal is somewhat problematic. A 50% reversal of gains might very well be below the threshold for reliable change (depending on the size of the SG in absolute scores). Thus, these fluctuations might result from measurement error rather than actual symptom reversals. Still, the fact that symptom reversals were found to be a moderator in Shalom and Aderka (2020) suggests they indeed do capture a clinically relevant phenomenon. Further research should investigate the role of reversals in the treatment of children and adolescents, preferably with a larger sample.

Gaynor et al. (2003) reported an average SG rate of 39% (and 50% in individual CBT), but to be included participants had to have participated in at least eight sessions. That is equivalent to the entire length of treatment in the present study. Unlike the results from the present study, Gaynor et al. (2003) found SGs to be associated with superior outcome at treatment termination. However, participants that did not experience SGs fared better in CBT compared to NST and SFBT. Contrary to the findings of Tang and DeRubeis (1999), Gaynor et al. (2003) found that outcome in CBT was robust also in the absence of SGs. A speculation is that there were LIIs in the CBT arm that were not captured by the original SG criteria, and that they would have been predictive of outcome. Further research is needed to investigate whether this might be the case.

Shalom and Aderka (2020) found that studies adhering to the original criteria as stated by Tang and DeRubeis (1999) were associated with smaller effects than studies using altered criteria for identifying SGs. The present study did not use the original criteria but rather a slightly adapted version introduced by Stiles et al. (2003), allowing for SGs to be identified also when only two points of measure were available before or after the gain. Using the original criteria would have considerably reduced the number of SGs, as the treatment length was only 8 sessions (rendering a total of 10 timepoints). Even so, our results suggest that further changes to the criteria might be meaningful. These findings are in line with previous research from Koffmann (2019), suggesting that removing the stability criterion leads to better outcome predictions than using the original SG definition. In the present study, LIIs seem to be a better predictor of outcome both at post-treatment and follow-up than SGs. Both early LIIs and LIIs during the entire treatment were found to be significantly related to better outcome at post-treatment and at follow-up. This was also true, although with a smaller effect-size, when using a secondary depression measure (MADRS-S). The results imply that large shifts in depressive symptoms might in themselves be important predictors of long-term change, regardless if the improvement is stable or not during treatment. However, neither LIIs nor SGs had a significant effect on comorbid anxiety as measured by GAD-7.

Shalom et al. (2018) found that SGs were predicted by symptom variability. Results from the present study could be used to further argue this point, at least regarding LIIs. The deterioration preceding the LIIs suggest a substantial symptom variability, where the average patient deteriorate significantly just before the actual gain. This is in line with Koffmann's (2019) arguing that the third SG criterion might prevent us from noticing potentially interesting symptom fluctuations before and after the gain.

Significant changes in emotion regulation did not precede SGs or LIIs. This could be an indication that there are other mechanisms behind both kinds of large intersession changes, that were not investigated in the current study. From a methodological standpoint, both depression and emotion regulation were measured weekly, and it is possible that causal effects between the two were not captured in these relatively wide time intervals. Hence, future studies could investigate this with more frequent assessments. The present study did not examine if the presence of certain treatment elements predicted SGs. Whether certain elements exist within therapist and patient interactions in IPDT that predict SGs or LLIs is an area for future research.

When implementing treatments in an internet-based format, there is always the question regarding how comparable the treatment is to face-to-face treatment and whether the treatment mechanisms are the same. Furthermore, this treatment, spanning over 8 weeks, is considerably shorter than most other PDT treatments and treatments in other studies investigating the role of intersession improvements as predictors of outcome. This study is important as it shows that SGs and LIIs occur in a short, internet-based treatment as well, and that LIIs are predictive of outcome.

Thus, results from this study are a few pieces in the puzzle when it comes to understanding the process and working mechanisms of IPDT for depressed adolescents, informing future treatment guidelines as well as further research. The ability to predict good and poor outcomes, in order to adapt and enhance treatments, is an ongoing quest in clinical psychology. We know that it is difficult for therapists to assess and predict treatment trajectories in their patients, often over-estimating improvements (Walfish et al., 2012) and failing to recognize non-response (Hatfield et al., 2010). This has led to a growing interest in feedback-informed treatment, implementing systems for clinicians to monitor their patient's progress during treatment and making informed treatment decisions. This study gives support to the notion that monitoring patients' trajectories can help identify likely responders and non-responders during treatment and thus might have the potential to improve outcomes in IPDT for depression in adolescents.

For clinicians, this study indicates that patients showing large intersession improvements are more likely to have favorable outcomes in IPDT, even if the improvement is followed by a reversal. In that sense, a score on a depression outcome instrument is best interpreted in the context of earlier trajectory in order to predict treatment response. A high score being stable over several sessions might, according to these results, be prognostically worse than a high score being preceded by improvement, even if that improvement is reversed or followed by symptom fluctuations. Future research should try to identify possible mechanisms driving LIIs in IPDT, in order to further our understanding of working mechanisms and make treatments more effective. This research should preferably encompass both specific and common factors to investigate whether there are treatment-specific elements driving these changes or rather generic, pan-theoretical elements such as alliance. In a wider perspective, there is a need for studies identifying treatment and patient factors predicting response and non-response in IPDT. This way, clinicians can make more informed treatment

decisions and adapt treatments according to patient needs, and we may also be able to make more informed decisions on treatment choice – furthering our understanding of what works for whom.

Strengths and Limitations

One strength of this study was the use of weekly assessments of both the outcome and process measures, both validated for adolescents. However, it is possible that more frequent measurements could have captured causal effects not seen in this study. Future studies could try implementing more frequent measures, even though that may be more demanding for participants. A related limitation is the attrition on the weekly measures, meaning that we may have underestimated the actual frequency of SGs and LIIs. Overall, the study sample was relatively small, and the results on SGs and LIIs that were non-significant could be due to low power. At the same time, despite the small sample size, we found robust effects of LIIs both early and throughout treatment, indicating that LIIs are more robust predictors of treatment outcome than SGs. As previously mentioned, the large amount of significance tests increases the risk for type I error, meaning that some (or all) of the significant results could be the result of chance. At the same time, all findings are in the same direction: LIIs predict depression outcome both at post-treatment and at follow-up, SGs do not. Even so, the findings from the present paper must be considered exploratory and in need of replication. Another limitation is that patients did not receive the exact same treatment, since one group did not have chat sessions. Nevertheless, the main treatment consisted of the modules and corresponding therapist feedback, with the modules being identical and the feedback highly regulated. We also investigated if this affected the effect of SGs and LIIs on outcome by controlling for group allocation in the model, but this did not change the results.

Conclusions

Results from this study highlight the need to continue investigating modifications of the original SG-criteria, as our results are in line with previous research suggesting that removing the stability criterion leads to better outcome predictions. A clinical implication is that therapists should be aware that the improvement in adolescent depression treatment is often not linear. Instead, recovery for many patients is more of a "bumpy road." This is underlined by the fact that many patients experiencing LIIs actually deteriorated just before the actual gain, and relatively large symptom improvements between consecutive treatment weeks, regardless of symptom stability, were predictive of short- and long-term treatment effects.

Acknowledgements

We thank Robert Johansson, PhD, for comments on an earlier draft of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Kavli Trust [grant number 32/18].

ORCID

Jakob Mechler http://orcid.org/0000-0002-1296-3432
Karin Lindqvist http://orcid.org/0000-0001-6325-4380
Fredrik Falkenström http://orcid.org/0000-0002-2486-6859
Per Carlbring http://orcid.org/0000-0002-2172-8813
Gerhard Andersson http://orcid.org/0000-0003-4753-6745
Björn Philips http://orcid.org/0000-0003-4313-1011

References

Abel, A., Hayes, A. M., Henley, W., & Kuyken, W. (2016). Sudden gains in cognitive-Behavior therapy for treatment-resistant depression: Processes of change. *Journal of Consulting and Clinical Psychology*, *84*(8), 726–737. https://doi.org/10.1037/ccp0000101

Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *80*(1), 93–101. https://doi.org/10.1037/a0026455

Andersson, G. (2016). Internet-Delivered psychological treatments. *Annual Review of Clinical Psychology*, *12*(1), 157–179. https://doi.org/10.1146/annurev-clinpsy-021815-093006

Andrusyna, T. P., Luborsky, L., Pham, T., & Tang, T. Z. (2006). The mechanisms of sudden gains in supportive-expressive therapy for depression. *Psychotherapy Research*, *16*(5), 526–536. https://doi.org/10.1080/10503300600591379

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for Measuring depression. *Archives of General Psychiatry*, *4*(6), 561. https://doi.org/10.1001/archpsyc.1961.01710120031004

Berking, M., Eichler, E., Luhmann, M., Diedrich, A., Hiller, W., & Rief, W. (2019). Affect regulation training reduces symptom severity in depression – A randomized controlled trial. *PLoS ONE*, *14*(8), 1–25. https://doi.org/10.1371/journal.pone.0220436

Bernstein, I., Rush, A. J., Trivedi, M., Hughes, C., Macleod, L., Witte, B., … Emslie, G. (2010). Psychometric Properties of the Quick inventory of depressive symptomatology in adolescents. *NIH Public Access*, *19*(4), 185–194. https://doi.org/10.1007/s10461-012-0143-z.Provider-patient

Bjureberg, J., Enander, J., Andersson, E., Ivanov, V. Z., Rück, C., & Fernández de la Cruz, L. (2019). Sudden gains in internet-based cognitive Behavior therapy for body dysmorphic disorder. *Behavior Therapy*. Advance online publication. https://doi.org/10.1016/j.beth.2019.11.002

Bjureberg, J., Ljótsson, B., Tull, M. T., Hedman, E., Sahlin, H., Lundh, L.-G., … Gratz, K. L. (2016). Development and validation of a brief version of the Difficulties in emotion regulation scale: The DERS-16. *Journal of Psychopathology and Behavioral Assessment*, *38*(2), 284–296. https://doi.org/10.1007/s10862-015-9514-x

Bjureberg, J., Sahlin, H., Hedman-Lagerlöf, E., Gratz, K. L., Tull, M. T., Jokinen, J., & Ljótsson, B. (2018). Extending research on emotion regulation individual therapy for adolescents (ERITA) with nonsuicidal self-injury disorder: Open pilot trial and mediation analysis of a novel online version. *BMC Psychiatry*, *18*(1), 1–13. https://doi.org/10.1186/s12888-018-1885-6

Brockmeyer, T., Michalek, S., Zipfel, S., Wild, B., Resmark, G., Teufel, M., & Friederich, H. C. (2019). Sudden gains in cognitive Behavioural therapy and Focal psychodynamic therapy for anorexia nervosa: Findings from the ANTOP study. *Psychotherapy and Psychosomatics*, *88*, 241–243. https://doi.org/10.1159/000499118

Feingold, A. (2009). Trials in the same Metric as for Classical analysis. *Psychological Methods*, *14*(1), 43–53. https://doi.org/10.1037/a0014699.Effect

Frederickson, J. J., Messina, I., & Grecucci, A. (2018). Dysregulated anxiety and dysregulating defenses: Toward an emotion regulation informed dynamic psychotherapy. *Frontiers in Psychology*, *9*(NOV), 1–9. https://doi.org/10.3389/fpsyg.2018.02054

Gaynor, S. T., Weersing, V. R., Kolko, D. J., Birmaher, B., Heo, J., & Brent, D. A. (2003). The prevalence and impact of large sudden improvements during adolescent therapy for depression: A comparison across cognitive-behavioral, family, and supportive therapy. *Journal of Consulting and Clinical Psychology*, *71*(2), 386–393. https://doi.org/10.1037/0022-006X.71.2.386

Gonçalves, S. F., Chaplin, T. M., Turpyn, C. C., Niehaus, C. E., Curby, T. W., Sinha, R., & Ansell, E. B. (2019). Difficulties in emotion regulation predict depressive symptom trajectory from early to Middle adolescence. *Child Psychiatry and Human Development*, *50*(4), 618–630. https://doi.org/10.1007/s10578-019-00867-8

Hamdeh, A. A., Bjureberg, J., Lenhard, F., Hedman-Lagerlöf, E., Flygare, O., Lundström, L., … Andersson, E. (2019). Sudden gains in internet-based cognitive behavior therapy for obsessive-compulsive disorder. *Journal of Obsessive-Compulsive and Related Disorders*, *21*(December 2018), 75–81. https://doi.org/10.1016/j.jocrd.2018.12.005

Hardy, G. E., Stiles, W. B., Cahill, J., Ispan, C., Macaskill, N., & Barkham, M. (2005). Sudden gains in cognitive therapy for depression: A replication and extension. *Journal of Consulting and Clinical Psychology*, *73*(1), 59–67. https://doi.org/10.1037/0022-006X.73.1.59

Hatfield, D., McCullough, L., Frantz, S. H. B., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology and Psychotherapy*, *17*(1), 25–32. https://doi.org/10.1002/cpp.656

Hedman, E., Lekander, M., Ljótsson, B., Lindefors, N., Rück, C., Hofmann, S. G., … Schulz, S. M. (2014). Sudden gains in internet-based cognitive Behaviour therapy for Severe health anxiety. *Behaviour Research and Therapy*, *54*(1), 22–29. https://doi.org/10.1038/jid.2014.371

Holzhauer, C. G., Epstein, E. E., Hayaki, J., Marinchak, J. S., McCrady, B. S., & Cook, S. M. (2017). Moderators of sudden gains after sessions addressing emotion regulation among women in treatment for alcohol use. *Journal of Substance Abuse Treatment*, *83*, 1–9. https://doi.org/10.1016/j.jsat.2017.09.014

Hunnicutt-Ferguson, K., Hoxha, D., & Gollan, J. (2012). Exploring sudden gains in behavioral activation therapy for major depressive disorder. *Behaviour Research and Therapy*, *50*(3), 223–230. https://doi.org/10.1016/j.brat.2012.01.005

Jacobson, N. S., & Truax, P. (1991). Clinical significance : A statistical Approach to Denning meaningful change in psychotherapy research. *J Consult Clin Psychol.*, *59*(1), 12–19. https://doi.org/10.1037//0022-006x.59.1.12

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials - A practical guide with flowcharts. *BMC Medical Research Methodology*, *17*(1), 1–10. https://doi.org/10.1186/s12874-017-0442-1

Johansson, R., Björklund, M., Hornborg, C., Karlsson, S., Hesser, H., Ljótsson, B., … Andersson, G. (2013). Affect-focused psychodynamic psychotherapy for depression and anxiety through the internet: A randomized controlled trial. *PeerJ*, *1*, e102. https://doi.org/10.7717/peerj.102

Johansson, R., Ljótsson, B., Frederick, R. J., Hesslow, T., Jansson, A., Jonsson, L., … Andersson, G. (2017). Internet-based affect-focused psychodynamic therapy for social anxiety disorder: A randomized controlled trial with 2-year follow-up. *Psychotherapy*, *54*(4), 351–360. https://doi.org/10.1037/pst0000147

Kelly, M. A. R., Roberts, J. E., & Ciesla, J. A. (2005). Sudden gains in cognitive behavioral treatment for depression: When do they occur and do they matter? *Behaviour Research and Therapy*, *43*(6), 703–714. https://doi.org/10.1016/j.brat.2004.06.002

Koffmann, A. (2019). Intersession improvement and outcome in psychotherapy. *Clinical Psychology and Psychotherapy*, *26*(1), 146–155. https://doi.org/10.1002/cpp.2337

Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Löwe, B. (2010). The patient health Questionnaire Somatic, anxiety, and depressive symptom scales: A systematic review. *General Hospital Psychiatry*, *32*(4), 345–359. https://doi.org/10.1016/j.genhosppsych.2010.03.006

Lindqvist, K., Mechler, J., Carlbring, P., Lilliengren, P., Falkenstöm, F., Andersson, G., & Björn, P. (2020). Affect-focused psychodynamic internet-based therapy for adolescent depression: A randomized controlled trial. *Journal of Medical Internet Research*, *22*(3), 1–15. https://doi.org/10.2196/18047

Mechler, J., Lindqvist, K., Falkenström, F., Carlbring, P., Andersson, G., & Philips, B. (2020). Emotion regulation as a time-Invariant and time-varying Covariate Predicts outcome in an internet-based psychodynamic treatment targeting adolescent depression. *Frontiers in Psychiatry*, *11*, 1–10. https://doi.org/10.3389/fpsyt.2020.00671

Ntini, I., Vadlin, S., Olofsdotter, S., Ramklint, M., Nilsson, K. W., Engström, I., & Sonnby, K. (2020). The Montgomery and Åsberg depression Rating scale–self-assessment for use in adolescents: An evaluation of psychometric and diagnostic

accuracy. *Nordic Journal of Psychiatry*, *74*(6), 1–8. https://doi.org/10.1080/08039488.2020.1733077

Present, J., Crits-Christoph, P., Connolly Gibbons, M. B., Hearon, B., Ring-Kurtz, S., Worley, M., & Gallop, R. (2007). Sudden gains in the treatment of Generalized anxiety disorder. *Journal of Clinical Psychology*, *66*(4), 430–441. https://doi.org/10.1002/jclp

R Core Team. (2019). *R: A language and environment for statistical computing*. http://www.r-project.org/.

Rush, a. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., … Keller, M. B. (2003). The 16-item Quick inventory of depressive. *Biological Psychiatry*, *54*(5), 573–583. https://doi.org/10.1016/S0006-3223(03)01866-8

Schäfer, JÖ, Naumann, E., Holmes, E. A., Tuschen-Caffier, B., & Samson, A. C. (2017). Emotion regulation strategies in depressive and anxiety symptoms in youth: A meta-analytic review. *Journal of Youth and Adolescence*, *46*(2), 261–276. https://doi.org/10.1007/s10964-016-0585-0

Shalom, J. G., & Aderka, I. M. (2020). A meta-analysis of sudden gains in psychotherapy: Outcome and moderators. *Clinical Psychology Review*, *76*, 1–16. https://doi.org/10.1016/j.cpr.2020.101827

Shalom, J. G., Gilboa-Schechtman, E., Atzil-Slonim, D., Bar-Kalifa, E., Hasson-Ohayon, I., Oppen, P. V., … Aderka, I. M. (2018). Intraindividual variability in symptoms Consistently Predicts sudden gains: An Examination of three independent datasets. *Journal of Consulting and Clinical Psychology*, *86*(11), 892–902. https://doi.org/10.1037/ccp0000344

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., … Dunbar, G. C. (1998). The mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, *59*(*Suppl 2*), 22–33. quiz 34-57. http://www.ncbi.nlm.nih.gov/pubmed/9881538.

Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., … Hardy, G. E. (2003). Early sudden gains in psychotherapy under routine clinic conditions: Practice-based evidence. *Journal of Consulting and Clinical Psychology*, *71*(1), 14–21. https://doi.org/10.1037/0022-006X.71.1.14

Storch, E. A., McGuire, J. F., Schneider, S. C., Small, B. J., Murphy, T. K., Wilhelm, S., & Geller, D. A. (2019). Sudden gains in cognitive behavioral therapy among children and adolescents with obsessive compulsive disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *64*(February 2018), 92–98. https://doi.org/10.1016/j.jbtep.2019.03.003

Svanborg, P., & Åsberg, M. (1994). A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating scale. *Acta Psychiatrica Scandinavica*, *89*(1), 21–28. https://doi.org/10.1111/j.1600-0447.1994.tb01480.x

Tang, T. Z., Beberman, R., DeRubeis, R. J., & Pham, T. (2005). Cognitive changes, critical sessions, and sudden gains in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, *73*(1), 168–172. https://doi.org/10.1037/0022-006X.73.1.168

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, *67*(6), 894–904. https://doi.org/10.1037/0022-006X.73.1.168

Tang, T. Z., Luborsky, L., & Andrusyna, T. (2002). Sudden gains in recovering from depression: Are they also found in psychotherapies other than cognitive-behavioral therapy? *Journal of Consulting and Clinical Psychology*, *70*(2), 444–447. https://doi.org/10.1037/0022-006X.70.2.444

Vincent, L., & Norton, P. J. (2019). Predictors and consequences of sudden gains in transdiagnostic cognitive-behavioural therapy for anxiety disorders. *Cognitive Behaviour Therapy*, *48*(4), 265–284. https://doi.org/10.1080/16506073.2018.1513557

Vittengl, J. R., Clark, L. A., & Jarrett, R. B. (2005). Validity of sudden gains in acute phase treatment of depression. *Journal of Consulting and Clinical Psychology*, *73*(1), 173–182. https://doi.org/10.1037/0022-006X.73.1.173

Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2015). Detecting sudden gains during treatment of major depressive disorder: Cautions from a Monte Carlo analysis. *Current Psychiatry Reviews*, *11*(1), 19–31. https://doi.org/10.2174/1573400510666140929195441

Vlaescu, G., Alasjö, A., Miloff, A., Carlbring, P., & Andersson, G. (2016). Features and functionality of the Iterapi platform for internet-based psychological treatment. *Internet Interventions*, *6*(December), 107–114. https://doi.org/10.1016/j.invent.2016.09.006

Walfish, S., McAlister, B., O'donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports*, *110*(2), 639–644. https://doi.org/10.2466/02.07.17.PR0.110.2.639-644

Wiedemann, M., Thew, G. R., Stott, R., & Ehlers, A. (2020). Suddengains: An R package to identify sudden gains in longitudinal data. *PLoS ONE*, *15*(3), 1–13. https://doi.org/10.1371/journal.pone.0230276

Wirtz, C. M., Radkovsky, A., Ebert, D. D., & Berking, M. (2014). Successful application of adaptive emotion regulation skills predicts the subsequent reduction of depressive symptom severity but neither the reduction of anxiety nor the reduction of general distress during the treatment of major depressive disorder. *PLoS ONE*, *9*(10), 1–12. https://doi.org/10.1371/journal.pone.0108288

Zilcha-Mano, S., Errázuriz, P., Yaffe-Herbst, L., German, R. E., & DeRubeis, R. J. (2019). Are there any robust predictors of "sudden gainers," and how is sustained improvement in treatment outcome achieved following a gain? *Journal of Consulting and Clinical Psychology*, *87*(6), 491–500. https://doi.org/10.1037/ccp0000401

Zilcha-Mano, S., Roose, S. P., Brown, P. J., & Rutherford, B. R. (2018). Abrupt symptom improvements in Antidepressant clinical trials. *The Journal of Clinical Psychiatry*, *80*(1), https://doi.org/10.4088/JCP.18m12353

Zwerenz, R., Becker, J., Johansson, R., Frederick, R. J., Andersson, G., & Beutel, M. E. (2017). Transdiagnostic, psychodynamic Web-based self-help intervention following Inpatient psychotherapy: Results of a Feasibility study and randomized controlled trial. *JMIR Mental Health*, *4*(4), e41. https://doi.org/10.2196/mental.7889