Published in final edited form as:

Curr Psychiatry Rev. 2015 February 1; 11(1): 19–31.

Detecting Sudden Gains during Treatment of Major Depressive Disorder: Cautions from a Monte Carlo Analysis

Jeffrey R. Vittengl, PhD*,

Department of Psychology, Truman State University

Lee Anna Clark, PhD,

Department of Psychology, University of Notre Dame

Michael E. Thase, MD, and

Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania

Robin B. Jarrett, PhD

Department of Psychiatry, The University of Texas Southwestern Medical Center

Abstract

Sudden gains are relatively large, quick, stable drops in symptom scores during treatment of depression that may (or may not) signal important therapeutic events. We review what is known and unknown currently about the prevalence, causes, and outcomes of sudden gains. We argue that valid identification of sudden gains (vs. random fluctuations in symptoms and gradual gains) is prerequisite to their understanding. In Monte Carlo simulations, three popular criterion sets showed inadequate power to detect sudden gains and many false positives due to (a) testing multiple intervals for sudden gains, (b) finite retest reliability of symptom measures, and (c) failure to account for gradual gains. Sudden gains in published clinical datasets appear similar in form and frequency to false positives in the simulations. We discuss the need to develop psychometrically sound methods to detect sudden gains and to differentiate sudden from random and gradual gains.

*Address correspondence to: Jeffrey R. Vittengl, Department of Psychology, Truman State University, 100 E. Normal St., Kirksville, MO 63501-4221, USA. vittengl@truman.edu. Voice 660-785-6041. FAX 660-785-7128.

Conflict of Interest

Drs. Vittengl and Clark have no financial interest or conflict of interest in the research.

Dr. Thase has no conflicts of interest pertaining to this paper, although he does report the following relationships with companies that develop treatment for depression or provide education pertaining to those treatments: During the past 3 years, Dr. Thase has provided scientific consultation to Alkermes, Allergan, Astra-Zeneca, Bristol-Myers Squibb Company, Dey Pharma, L.P., Eli Lilly & Company, Forest Pharmaceuticals, Inc., Gerson Lehman Group, GlaxoSmithKline, Guidepoint Global, H. Lundbeck A/S, MedAvante, Inc., Merck and Co. Inc., Neuronetics, Inc., Novartis, Otsuka, Ortho-McNeil Pharmaceuticals, PamLab, L.L.C., Pfizer (formerly Wyeth-Ayerst Laboratories), PharmaNeuroboost, Shire US Inc., Sunovion Pharmaceuticals, Inc., Takeda (Lundbeck), Teva and Transcept Pharmaceuticals. Dr. Thase receives grant funding from the Agency for Healthcare Research and Quality, Alkermes, AstraZeneca, Eli Lilly & Company, Forest Laboratories, the National Institute of Mental Health, Otsuka Pharmaceuticals, PharmaNeuroboost, and Roche. He has equity holdings in MedAvante, Inc. and receives royalty income from American Psychiatric Foundation, Inc., Guilford Publications, Herald House, Oxford University Press, and W.W. Norton & Company. His wife is employed by Peloton Advantage, which does business with Pfizer.

Dr. Jarrett's medical center collects the payments from the cognitive therapy she provides to patients. Dr. Jarrett is a paid consultant to the NIMH.

Keywords

depression; psychotherapy; sudden gains; assessment; symptoms

Introduction

Acute-phase cognitive therapy (CT) [1] is an efficacious treatment for major depressive disorder (MDD) [2, 3] but how patients improve in CT is poorly understood [4]. In this article, we consider the phenomenon of “sudden gains” (relatively large, quick, stable drops in symptom scores) [5] that may mark important therapeutic events related to CT’s mechanisms, but that also may reflect gradual gains and random fluctuations in symptoms. In a review of the sudden gains literature and concept, we consider whether published rates of sudden gains substantively exceed Type I error rates from Monte Carlo simulations. We focus our review on studies of MDD, the disorder for which Tang and DeRubeis proposed sudden gains, because the extent to which patterns of sudden gains from episodic MDD generalize to other disorders with different courses (e.g., sexual disorders [6], social phobia [7]), although hypothesized, remains to be demonstrated.

Acute-phase CT [1] often includes 16–20 individual sessions of 50–60 minutes duration. A common schedule in research is 2 sessions per week for 8 weeks followed by 1 session per week for 4 weeks. Typically, early sessions are spent reactivating behaviors (e.g., role functioning, leisure, work) that diminished with depressed mood, and teaching patients about the role of cognition in emotion and behavior. Cognitive therapy aims to reduce depressive symptoms by eliciting thoughts associated with distressing emotions and by teaching patients to evaluate the validity of these thoughts through logical and empirical methods. Patients learn to generate more realistic cognitions when negative thoughts are not supported by hypothesis testing or logical analysis. They also learn to employ problem-solving skills when negative conclusions appear justified empirically.

A substantial research base supports acute-phase CT’s efficacy in treating MDD. On average, patients experience a large reduction in depressive symptom scores from pre- to post-CT, often 1–3 *SD* [2, 3] on such measures as the Beck Depression Inventory (BDI) [8] and Hamilton Rating Scale for Depression (HRSD) [9]. Similarly, about half of patients no longer meet criteria for MDD post-CT [2]. In randomized clinical trials, CT patients’ average post-treatment symptom scores are similar to scores found with interpersonal psychotherapy and pharmacotherapy—about 0.7 *SD* below patients’ scores in non-active control conditions [10].

Establishing a treatment’s efficacy, however, does not clarify its mechanisms [4]. Studying significant events during treatment may advance understanding of how CT works [11]. For example, quick changes (e.g., decreases in symptoms) during psychotherapy are hypothesized to signal the operation of important therapeutic mechanisms. Historically within psychodynamic models and treatment, “insight” has been hypothesized to trigger patients’ improvements [12]. Within a cognitive model and CT, “insight” may be operationalized as learning core therapeutic concepts (e.g., thoughts influence emotion, and thoughts can be changed; adaptive behaviors may change both thoughts and emotions) [13].

In this context, Tang and DeRubeis [5] proposed that depressed patients who experience sudden gains during CT have cognitive changes that facilitate superior short- and long-term outcomes. Tang and DeRubeis [5] defined sudden gains as drops in BDI scores between therapy sessions of at least (a) 7 points, (b) 25%, and (c) $2.78t$, where t is the difference in means of the three scores before and after the gain divided by the standard error of difference in means (i.e., an independent-samples t -test with nominal two-tailed $\alpha = .05$). These criteria were meant to identify importantly large gains and to separate real (true positive) from random (false positive or Type I error) gains.

Tang and DeRubeis [5] indicated that their sudden-gain criteria were partly arbitrary, and most subsequent researchers have used modified criteria. For example, Tang, DeRubeis, *et al.* [14, 15] changed the third criterion from $2.78t$ to $2.78v$, where v is the difference in means divided by the pooled SD (i.e., revamping the third criterion as a variance ratio rather than a formal t -test), perhaps because the t -test is biased due to autocorrelation in this context [16, 17]. In addition, because CT protocols can involve two sessions weekly during the first half of acute-phase treatment, but one session weekly during the last half, Tang and DeRubeis's [5] criteria have the potential to provide inconsistent thresholds for "suddenness" of gains developing over 0.5 versus 1 week. In an attempt to address this issue, Vittengl *et al.* [16] used symptom assessments with a more-consistent time interval (by protocol, 1 week apart) because constant time intervals support valid inferences in traditional time-series analyses [18]. Other published modifications include requiring smaller t ratios [17]; replacing t with maintenance of 50% of the gain for two assessments [19] or with the reliable change index [19, 20, 21]; using symptom measures other than the BDI [22]; and testing gains between all pairs of adjacent assessments by replacing t with the ratio of the drop in symptom scores divided by the SD of the patient's entire series of scores [23].

Despite variations in criteria, the reported frequency (median proportion of patients = 42%, range 33–50%) and magnitude (median change in BDI scores = 10.8, range 9.3–13.3) of sudden gains in CT for unipolar depression have been relatively consistent (see Table 1). Interestingly, sudden gains in CT appear similar to those in other active treatments for depression (median proportion of patients = 41%, range 26–50%; median change in BDI scores = 11.5, range 10.1–13.0) and in non-active comparison groups (median proportion of patients = 43%, range 25–60%; median change in BDI scores = 11.9, range 11.2–12.5; see Table 1), although data are limited to two studies [16, 19]. Further, some patients experience sudden gains before CT starts [19, 24]. Thus, an active treatment producing insights or learning appears may be unnecessary for sudden gains, a possibility we explore in the current analyses.

Most studies find that depressed patients with sudden gains during treatment, compared to patients without sudden gains (often including non-responders), show lower depressive symptoms at the end of treatment. For example, 5 of 8 studies with CT cells showed significantly lower depression among patients with sudden gains (see Table 1). A recent meta-analysis of sudden gains in both anxiety and depression reached a similar conclusion [32]. These findings support the idea that many patients with sudden gains are improved relative to patients without sudden gains, but leave unclear whether it is gains in general of a

given magnitude, or sudden gains in particular, that are important. For example, phenomena that appear to be sudden gains actually may represent gradual gains plus random variance. In the current analyses, we test how often gradual improvement processes produce false-positive sudden gains.

Evidence for long-term benefits and cognitive mechanisms of sudden gains in depression is more limited. Tang and DeRubeis [5] found mixed evidence favoring patients with (vs. without) sudden gains during follow-up (lower BDI scores at 6 and 18, but not 12, months; no difference in relapse rates). Subsequently, two studies showed advantages to patients with sudden gains, two showed no significant difference between patients with and without sudden gains, and one showed a significant disadvantage to patients with sudden gains (see Table 1). In addition, the advantages of sudden gainers were reduced to non-significant levels when only treatment responders were analyzed (i.e., sudden gains correlated with response) in one supportive study [17].

To investigate mechanisms, Tang and DeRubeis [5] coded audiotapes for cognitive changes in CT sessions one versus two sessions before patients' sudden gains. Cognitive changes included identification of errors in thinking, arriving at new beliefs, and bringing thoughts into awareness. To a moderate extent ($d = 0.65$), more cognitive changes were identified in CT sessions immediately before sudden gains. Tang *et al.* [15] replicated this finding in a pooled sample of patients who received CT or Automatic Thought therapy (a component of CT), but cognitive changes were not higher before sudden gains in depressed patients receiving supportive-expressive therapy [29]. Together these findings suggest that cognitive changes may be neither sufficient nor necessary for sudden gains.

Although methods have varied greatly, other research has shown little connection between depressive cognition and sudden gains. Self-evaluations do not appear to improve before the sudden gains of either patients in CT [23] or untreated persons with MDD [27]. In addition, patients with higher pre-treatment negative cognition, who have greater potential for cognitive change during treatment, do not have more sudden gains [16, 17, 23]. Consequently, the mechanisms and types of depressive cognition related to sudden gains require empirical clarification, if they are in fact linked.

In the current study, our Monte Carlo simulations estimated how often random fluctuations in symptoms could produce Type I errors when symptom measures are tested for sudden gains. Further, we estimated power to detect sudden gains of small (1.0 *SD* drop in symptom scores between adjacent assessments), moderate (1.5 *SD* drop), and large (2.0 *SD* drop) sudden gains. Rather than model all published modifications of sudden gain criteria, we addressed three of the more popular criterion sets [5, 19, 23].

At least three processes may influence the psychometric performance (i.e., Type I error / false positive rate and power) of sudden-gains criteria. First, when testing hypotheses (e.g., "this patient had a sudden gain"), each statistical test (e.g., *t*-test for difference in symptom means; comparison of difference in two scores to a reliable change index) should be conducted using an *a priori* alpha level. Most often, researchers choose .05, two-tailed, meaning that "statistically significant" ($p < .05$) results occur by chance alone 5% of the

time (e.g., 2.5% Type I errors in identifying sudden gains plus 2.5% Type I errors in identifying sudden losses = 5%). But the Type I error rate is inflated beyond 5% when conducting multiple statistical tests [33]. For example, if a patient completes 16 symptom assessments, 11 (means of three scores before vs. after the gain) to 15 (between each pair of adjacent scores in the series) separate intervals might be tested for sudden gains. Although the tests are partly overlapping, Type I errors may well exceed the nominal rate of 2.5% for sudden gains. The potential influence of series length on power is less clear, but other processes being equal, longer series of scores contain more information and may increase power to detect sudden gains. We tested the impact of series length on Type I error and power by manipulating the number of repeated symptom assessments from 10 to 20, representing typical numbers of sessions of CT for depression.

Second, the retest reliability of symptom measures may influence the Type I errors and power in testing for sudden gains. Measures with lower retest reliability show more random fluctuations in scores over time [20]. Consequently, the Type I error rate may be higher, and power lower, when symptom measures with lower retest reliability (and greater random fluctuations) are tested for sudden gains. On the other hand, retest reliability can be viewed as autocorrelation in a patient's series of repeated symptom scores. Positive autocorrelation is common in patients' symptom trajectories and occurs when symptom scores closer in time (e.g., weeks 3 and 4 in CT) are on average more similar than scores farther apart (e.g., weeks 3 and 11). Autocorrelation violates the independence-of-observations assumption in statistical tests, including the *t*-test used to identify sudden gains [18]. Greater positive autocorrelation makes traditional critical values too small (e.g., a value larger than $t(4) = 2.78$ is required for $p < .05$) [34]. Consequently, positive autocorrelation may increase both the Type I error rate and power because more putative sudden gains will be detected overall. We investigated the relation of reliability/autocorrelation to Type I errors and power by varying the retest correlations of symptom scores in our simulations.

Finally, gradual reductions in a patient population's symptoms (e.g., due to an efficacious treatment) may make identification of sudden gains more challenging [35]. Patients' symptom reductions with medication and psychotherapy are often not abrupt or sudden but gradual, including linear (small, steady gains throughout treatment) and curvilinear (moderate gains earlier followed by progressively smaller gains later in treatment) patterns [36, 37]. When random fluctuations in symptom scores due to a measure's imperfect reliability occur around underlying gradual gains, Type I errors in testing for sudden gains may be more common. For example, Thomas and Persons [30] found moderate to high rates of false-positive sudden gains (using sudden-gain criteria from [5]) among simulated cases with gradual linear (20–66% of cases) and curvilinear (19–68% of cases) symptom reductions. More false positive were associated with larger gradual gains [30]. In the current simulations, we extended analyses of sudden gains by considering no-change and gradually improving populations, three sets of sudden gains criteria, varying symptom measure reliability, varying lengths of series of symptom scores, and both Type I errors (false positives) and statistical power.

Method

Simulated Symptom Score Datasets

We used summary statistics (means, variances, correlations) of the BDI from CT clinical trials to seed our Monte Carlo simulations. Raw patient data were not used in this report. Pooling published means and *SDs* from datasets used in an influential text documenting CT's efficacy [2] plus two more recent clinical trials [38, 39], we found a pre-CT mean of 30 on the BDI (marking moderate to severe depression [40]), *SD* = 8, and a post-CT mean of 11. Simulations used *M* = 30 at the first assessment (*cf.* first CT session) and *SD* = 8 at all assessments. Our simulations manipulated the number of BDI scores, including series of 10, 15, and 20 assessments, representative of the sudden-gains literature.

We also manipulated changes in population means across repeated assessments (see examples in Figure 1). First, we simulated populations with no change in symptom scores (*i.e.*, *M* = 30 at all assessments). This no-change population can be expressed symbolically as,

$$BDI_k = 30 + e,$$

where *k* refers to a particular assessment in the series (*i.e.*, assessments numbered 1, 2, 3, etc.) and *e* represents individual patients' variance in scores around the population mean (a function of the population *SD* = 8 described above and retest correlations described below).

Second, we simulated populations with gradual linear decreases in symptoms of 1 BDI point per assessment (*Ms* = 30, 29, 28, etc.), which can be expressed as,

$$BDI_k = 30 - 1.00 * (k - 1) + e.$$

Over a 20-session CT protocol, for example, this linear decrease matched our estimates from pooled clinical trials of mean BDI scores dropping from 30 at the first assessment to 11 at the final (20th) assessment.

Third, we simulated populations with gradual curvilinear decreases in symptoms,

$$BDI_k = 30 - 6.34 * \ln(k) + e.$$

This log-linear function describes moderate decreases in symptoms between early assessments slowing to small decreases between later assessments, such as often describe change in psychotherapy [41, 42]. For example, this equation yields population means of 30.0, 25.6, 23.0, and 21.2 for the first four assessments in a series. We selected the coefficient of 6.34 so that the log-linear function yields an overall decrease in symptom means identical to the linear function across 20 sessions (*i.e.*, decrease from 30 to 11).

Fourth, we simulated populations with sudden gains of varying magnitudes. Except for sudden gains, the populations had unchanging mean BDI scores. With a pre-gain BDI score of 30 (the mean pre-treatment score in clinical trials, as discussed above), the smallest between-assessment change that could meet Tang and DeRubeis's [5] sudden gain criteria is 8 BDI points. Because the pooled clinical trial *SD* is also 8, sudden gains of 8 points can be conceptualized as 1.0 *SD*. We doubled the "small" sudden gain of 1.0 *SD* (8 BDI points) to 2.0 *SD* (16 BDI points) to define "large" sudden gains. We labeled the mid-point between small and large sudden gains, 1.5 *SD* or 12 BDI points, as "moderate" sudden gains, which is similar to the median sudden gain in the literature (see Introduction). In our simulations, we placed small, moderate, and large sudden gains randomly among the intervals that sudden-gains criteria searched (see Sudden-Gains Criterion Sets section). That is,

$$\text{BDI}_k = 30 - g_{k>t} + e,$$

where g refers to the sudden gain of 8, 12, or 16 BDI points, and t refers to the assessment immediately before which the gain occurred. For example, in a population with 10 assessments, a sudden gain of 12 points ($g = 12$) might occur after the fourth assessment ($t = 4$) yielding mean BDI scores of 30, 30, 30, 30, 18, 18, 18, 18, 18, 18. Scores for individual patients in the population would vary randomly around the means of 30 and 18.

Finally, we manipulated the retest reliability/autocorrelation of symptom scores. We examined correlation matrices from two large acute-phase CT trials, each with 14 observations collected roughly weekly ($N = 127$ from [43] as described by [16]; $N = 362$ from [44] as described by [45, 46]). The median-lag-1 retest correlation (i.e., 1 assessment apart) was .81 and decreased gradually to .23 at lag 13 (i.e., .05 per assessment, on average). Although longer series likely have lower retest correlations at higher lags, mood measures like the BDI often contain "trait" components that produce retest correlations > 0 over long intervals [47]. Further, BDI retest correlations vary notably among samples [40]. With these issues in mind, we simulated populations with lag-1 correlations of .90, .80, and .70 that decreased by .05 per assessment to a minimum of .10. For example, in a population with retest $r = .90$, lag 1 correlations between adjacent assessments (e.g., between assessments 1 and 2, 2 and 3, 3 and 4) would be .90; lag 2 correlations between assessments two observations apart (e.g., between assessments 1 and 3, 2 and 4, 3 and 5) would be .85, and so on.

In sum, we manipulated series length (10, 15, or 20 BDI assessments), change in population means (no change, gradual linear change, gradual log-linear change, or sudden gains of 8, 12, or 16 BDI points), and retest reliability of the BDI (lag-1 correlations of .70, .80, and .90) in a fully crossed design (all combinations of these population characteristics represented). We programmed our simulations using published computer routines [48]. Searching for sudden gains can be viewed as type of idiographic time-series analysis [16, 18] because the unit of analysis is an individual patient's series of repeated depressive symptom scores. In this context, we drew 10,000 samples (i.e., individual patients' series of 10–20 symptom scores) for each scenario. We drew simulated patient data randomly from multivariate normal populations and rounded simulated BDI scores to integers within the

measure's actual range (0–63). We applied three sets of sudden-gains criteria to each simulated population.

Sudden-Gains Criterion Sets

Tang and DeRubeis [5]—In this criterion set, sudden gains are defined as drops in BDI scores between adjacent assessments of at least (a) 7 points, (b) 25% of the pre-gain score, and (c) $2.78t$, where t is the difference in means of the three scores before and after the gain divided by the standard error of difference in means (i.e., an independent-samples t -test with nominal two-tailed $\alpha = .05$). For gains immediately after the second assessment, and for gains immediately before the second-to-last assessment, criterion c is modified to $3.18t$ using only five symptom scores in the t ratio. Consequently, the Tang and DeRubeis [5] criteria indicate that a total of $k-3$ intervals are tested for sudden gains (e.g., 7 intervals among a series of 10 BDI scores).

Gaynor et al. [19]—Sudden gains are defined as by [5] plus an additional test immediately after the first assessment. Here, sudden gains are drops in BDI scores of at least (a) 7 points, (b) 25% of the pre-gain score, and (c) maintained by at least 50% for the next two assessments. For example, scores of 30, 20, 19, 21 contain a sudden gain, but scores of 30, 20, 19, 26 do not because the drop of 10 points between the first and second assessments is reduced to 4 points (only 40% of the gain) at the second following assessment (BDI = 26). The Gaynor *et al.* [19] criteria indicate that a total of $k-2$ intervals are tested for sudden gains.

Kelly et al. [23]—Sudden gains are defined as drops in BDI scores between adjacent assessments of least (a) 7 points, (b) 25% of the pre-gain score, and (c) $1.5s$, where s is the ratio of the drop in scores to the SD of the patient's entire series of BDI scores. The Kelly *et al.* [23] criteria indicate all $k-1$ intervals are tested for sudden gains.

Results

We estimated the Type I error rate (false positives) and power (true positives) in identifying sudden gains by applying the three sets of criteria just discussed to simulated datasets representing a wide range of clinically plausible scenarios (each using 10,000 cases without sudden gains). As shown in Table 2, Type I errors were common in all scenarios but varied by four processes (median = 43%, range 5–76%).

First, although false-positive sudden gains were common even in populations with no changes in symptom scores, populations with gradual decreases in symptom scores generally showed more Type I errors (see Table 2). Collapsing across other variables, the median Type I error rate was 17% (range 5–71%) in no-change, 47% (range 18–71%) in gradual linear change, and 55% (range 24–76%) in gradual log-linear change populations. Thus, all scenarios showed considerably more than 2.5% Type I errors, the proportion that researchers and clinicians might expect based on our field's tradition of testing hypotheses at $p < .05$, two-tailed. Although purely random fluctuations in symptoms often produced the “sudden gain” pattern, layering random fluctuations on gradually decreasing symptoms (e.g., due to an effective treatment) greatly increased the rate of Type I errors.

Second, longer series of symptom scores generally produced more Type I errors (see Table 2). Collapsing across other variables, the median Type I error rate was 32% (range 5–63%) in 10-assessment, 48% (range 8–69%) in 15-assessment, and 50% (range 11–76%) in 20-assessment populations. The increase in Type I errors with increasing series length is unsurprising because none of the sudden-gains criterion sets have Type I error control when testing multiple intervals. For example, patients' series of 10, 15, and 20 BDI scores contain 9, 14, and 19 intervals that can be tested for sudden gains. The problem of multiple statistical tests may not be solved by deciding to test only one interval for a sudden gain (e.g., the interval during which a patient experienced the largest drop in BDI scores, or the interval with the most stable drop in scores) because an informed decision about which interval to test may require much of the same information used to define sudden gains with the current criterion sets.

It is noteworthy that the Kelly *et al.* [23] criteria set showed some small reversals of the general pattern of increasing Type I errors with increasing series length. The Kelly *et al.* criteria showed these reversals only in gradually changing populations (no reversals in no-change populations), presumably because within-series variation was increased, and the ratio of symptom change to within-series variation decreased (see criterion *c* in the Method section), due to the gradual decreases in symptom scores. Nonetheless, the Type I error rates for the Kelly *et al.* criteria in gradually changing populations were very high (median = 48%, range 18–64%).

Third, lower retest reliabilities in symptom scores generally produced more Type I errors (see Table 2). Collapsing across other variables, the median Type I error rate was 20% (range 5–56%) in .90 reliability, 48% (range 9–73%) in .80 reliability, and 58% (range 12–76%) in .70 reliability populations. Populations with symptom measures of lower reliability have greater random fluctuations in scores from assessment to assessment, more often matching the “sudden-gain” pattern. Although higher autocorrelation/reliability inflates the *t* ratio [34] used in the Tang and DeRubeis [5] and Gaynor *et al.* [19] criterion sets, and so makes meeting sudden gains criteria more likely, lower reliability/autocorrelation was associated with increased Type I errors in all three sudden-gain criterion sets, including Kelly *et al.* [23] criteria that do not use the *t* ratio. Greater variation from assessment to assessment (e.g., bigger drops in scores) due to lower reliability appeared dominant in producing the sudden-gain pattern.

Fourth, the Type I error rate varied by the sudden-gains criteria applied (see Table 2). Collapsing across other variables, the median Type I error rate was 33% (range 5–68%) for Tang and DeRubeis [5], 41% (range 6–76%) for Gaynor *et al.* [19], and 48% (range 11–71%) for Kelly *et al.*'s [23] criteria. Given the finding that testing more intervals increased the Type I error rate, one interpretation is that the difference among *k*-3 (Tang and DeRubeis), *k*-2 (Gaynor *et al.*), and *k*-1 (Kelly *et al.*) intervals searched for sudden gains may largely account for variation in Type I errors among criterion sets. Nonetheless, all showed frequent Type I errors.

To understand the size, stability, and timing of the false-positive sudden gains, we extracted medians for several indices across the 81 modeled scenarios shown in Table 2. The grand

median sudden-gain size was 10 BDI points (range 8–13) or 1.3 *SD* (range 1.0–1.6), representing 67% (range 36–100%) of the total change from first to last symptom scores in the data series. Relatively few false-positive sudden gains reversed by the last observation (median = 9%, range 3–44%), with reversals defined as upticks in symptom scores of at least half the size of preceding sudden gains [5]. False-positive sudden gains occurred sooner in shorter data series: In the 10, 15, and 20-observation series, sudden gains occurred after the median 5th (range 2–6), 8th (range 4–10), and 11th (range 5–13) observations, respectively. False-positive sudden gains also occurred sooner in log-linear decreasing series (median = 5, range 3–11) than in linear (median = 9, range 5–13) and no-change series (median = 8, range 4–13) because the log-linear populations demonstrate larger (but still gradual) gains earlier.

To visualize the random fluctuations of scores identified as sudden gains, we created plots parallel to those in the sudden gains literature [5]. For example, Figure 2 shows mean symptom scores for Type I errors resulting from application of Tang and DeRubeis's [5] criteria to two populations: (a) no change in symptoms, 10 BDI assessments, retest $r = .90$; and (b) gradual linear decrease in symptoms, 15 BDI assessments, retest $r = .80$. The second population arguably is common clinically, and here the average decreases in symptoms in the “sudden-gain” interval (about 10 BDI points) and pattern shown are very similar to putative sudden gains in published clinical datasets.

As shown in Table 3, power to detect sudden gains usually was inadequate ($< 80\%$; median = 72%, range 40–93%). It is important to recall features of the simulated populations used to estimate power: There were no systematic changes in symptoms other than those due to sudden gains (i.e., power was not limited due to confusion of gradual for sudden gains); sudden gains were placed randomly only with the ranges of assessments searched by each criterion set (i.e., power was not limited by sudden gains occurring in untested intervals); each scenario included 10,000 cases drawn from populations with small (8 BDI points = 1.0 *SD* drop in symptoms between adjacent assessments), moderate (12 BDI points = 1.5 *SD*) sudden gains, and large (16 BDI points = 2.0 *SD*) sudden gains; and the sudden gains were stable (i.e., the 8-, 12-, or 16-point drop was absent at all observations before, and present at all observations after, the sudden gain). If viewed as standardized mean difference effect sizes, the “small” ($d = 1.0$), “medium” ($d = 1.5$), and “large” ($d = 2.0$) sudden gains all mark substantial drops in symptoms.

Three variables had only small effects on power. First, median power differed little among the three criterion sets, 72% (range 40–93%) for Tang and DeRubeis [5], 74% (range 43–93%) for Gaynor *et al.* [19], and 72% (range 43–79%) for Kelly *et al.*'s [23] sudden-gains criteria. Second, power increased slightly from series of 10 (median = 68%, range 40–86%) to 15 (median = 72%, range 44–91%) to 20 (median = 75%, range 43–93%) symptom scores. Third, median power differed little by symptom scores' retest reliability, 74% (range 40–90%), 70% (range 41–91%), and 73% (range 41–93%) for retest $r = .70$, .80, and .90, respectively.

The counterintuitive finding of lowest median power at retest $r = .80$ was due to the Kelly *et al.* [23] criteria. Whereas for increasing reliabilities of .70, .80, and .90, the Tang and

DeRubeis [5] criteria (68%, 72%, 78%) and Gaynor *et al.* [19] criteria (70%, 74%, 79%) showed increasing power medians, respectively, the Kelly *et al.* criteria showed decreasing power (75%, 69%, 66%). We suspect that this unusual feature of the Kelly *et al.* criteria again is due to within-series variation increasing, and the ratio of symptom change to within-series variation decreasing (see criterion *c* in the Method section), with lower retest reliability of symptom scores.

The size of the simulated sudden gain was the most important predictor of power (see Table 3). Power to detect small (median = 50%, range 40–75%) and moderate (median = 72%, range 60–81%) sudden gains usually was inadequate (< 80%), whereas power to detect large sudden gains sometimes was acceptable (median = 83%, range 75–93%), especially when the BDI had moderate to high reliability and the Tang and DeRubeis [5] or Gaynor *et al.* [19] criteria were applied. However, even under these most favorable conditions, the power to detect large sudden gains (median = 89%) must be viewed in the context of the high Type I error rate (median = 33%; see Table 2).

Discussion

Frequent measurement of depressive symptoms during treatment of MDD is now a recommended standard of care [49], but interpretation of idiographic patterns of change is an evolving concern [46, 50]. We found that sudden gains reported in the MDD literature are not obviously different in frequency or pattern from random fluctuation in symptoms (Type I errors / false positive sudden gains) in Monte Carlo simulations. For example, our review of the literature suggested that roughly of 42% of depressed patients in published studies showed sudden gains, and our median Type I error rate across a wide range of simulated populations was 43%. In fact, none of the 81 simulations yielded a false-positive rate as low as expected (2.5%) based on the convention of testing hypotheses at $p < .05$, two-tailed. Further, visual inspection of the symptom patterns in our simulations showed that Type I errors often mimicked reported sudden gains in published clinical data [5]. Consequently, we speculate that the literature is likely contaminated with false positives due to lack of psychometrically adequate criteria to differentiate sudden gains from random fluctuations and gradual gains.

Our simulations identified several variables that better sudden-gains criteria should account for to control the Type I /false positive rate. Symptom measures with lower retest reliability, patients with longer series of symptom scores (and more intervals tested for sudden gains), and gradual reductions in symptoms (e.g., due to an effective treatment) generally produced more Type I errors. Consequently, we encourage researchers to develop sudden-gains criteria that account for the number of intervals tested for sudden gains, reliability of their measures, and gradual progress over the course of treatment (as opposed to no improvement) as the null hypothesis against which sudden gains are tested. Because measurement situations may vary widely (e.g., number of psychotherapy sessions; retest reliability in a particular patient population and specific assessment interval), researchers may need to develop and test sudden-gains criteria locally rather than relying on universal criteria.

If researchers simply attempt to modify existing sudden-gains criteria [5, 19, 23], they may find an untenable trade-off between controlling the Type I error rate and reducing power to detect sudden gains. Even with high Type I error rates indicating liberal criteria, our simulations showed that power generally was inadequate (< 80%) to detect small and moderate sudden gains and only sometimes adequate (about 60% of the time) for detecting large sudden gains. If researchers control the Type I error rate by tuning existing sudden-gains criteria to be more conservative (e.g., Bonferroni corrections to maintain the nominal alpha level when testing multiple intervals; increasing the required *t* ratio for detection of sudden gains), power to detect sudden gains may suffer further [51]. We note that sudden-gains researchers often use data series (e.g., 5 or 6 assessments for a *t*-test) much shorter than recommended minimums for time-series analyses with adequate statistical power (e.g.,

50 assessments) [18, 52]. High Type I error rates and low power appear to be widespread problems for the time-series analytic techniques used in single-case treatment studies [53]. In sum, it may be unrealistic to expect better for sudden-gains tests with short data series.

In this context, researchers may want to explore additional information and technologies to augment sudden-gains tests. An obvious source of increased information would be to include more symptom scores in sudden-gains tests (e.g., increase from 2–3 to 6–8 assessments before and after the gain). Of course, increasing the number of symptom assessments will increase the clinical assessment burden and decrease the pool of patients who may be tested for sudden gains. Information external to symptom measures also could be added to *a priori* sudden-gains tests. Researchers and clinicians could sharpen their search for sudden gains by asking the question, “Is there a clear reason to believe that this patient is about to have a sudden gain?” If changes in social-interpersonal functioning, depressive cognitive content, or biological markers suggest that a sudden gain is likely, sudden-gains tests may be justified over the next symptom assessment or two. Our latter suggestion is in some ways the reverse of searching for cognitive changes once sudden gains have been identified [5].

Researchers also could consider alternative strategies to model patients’ symptom change trajectories. Recently, Vittengl *et al.* [46] identified linear, log-linear, and one-step change trajectories in individual patients’ symptom scores during CT for depression using regression analyses with statistical significance thresholds set via Monte Carlo simulation to account for multiple statistical tests and autocorrelation. Simulation Modeling Analysis offers a conceptually similar process for detecting overall changes in individual patients’ symptoms due to treatment in quasi-experimental designs [54], and possibly could be modified for use in detecting specific change patterns (e.g., gradual vs. sudden gains). Moreover, curvilinear regression can identify rapid early response and symptom spikes in psychotherapy [55]; growth mixture modeling can be used to identify groups of patients with different symptom-change trajectories [37]; and growth-curve analysis shows potential as a means to investigate both gradual and discontinuous changes in symptom level and trajectory [56].

We speculate that observed sudden gains, consistent with classical test theory [57], represent a combination true (i.e., valid, systematic) and error (i.e., random, unsystematic) variance. Further, we propose that the contributions of true and error variance to the sudden-gain

symptom score pattern vary from case to case. For example, when real patients do not improve in treatment, and when simulated patient data are drawn from populations with constant symptom means (e.g., mean of 30 on the BDI throughout repeated assessments), then observed sudden gains are entirely random and invalid (i.e., contain zero true sudden-gain variance). Consequently, observed sudden gains are less common in this scenario because random variance alone forms the sudden-gain pattern. In a second scenario in which real patients improve gradually (e.g., decrease the equivalent of 1 or 2 BDI points per assessment), and when simulated patients are drawn from populations with gradually decreasing symptom means, then observed sudden gains are partly valid (patients have real gains) and partly error (patients have gradual, not sudden, gains). Observed sudden gains are more frequent because the amount of random variance necessary to produce the sudden-gain pattern is reduced by the real gradual gain. Finally in a third scenario where patients improve abruptly (e.g., one big step down in symptoms, from scores of 30 to 14 on the BDI), and when simulated patients are drawn from populations with consistently high symptom means followed by consistently low symptom means, then observed sudden gains are fully valid (i.e., the patients improved, and the improvement was sudden, not gradual) and not produced by random variance. Observed sudden gains are most frequent in the final scenario because random variance can only obscure the true sudden gain.

In this theoretical framework, there are two measurement tasks—identification of gains and identification of suddenness. We propose that the sudden-gains literature sometimes conflates the two tasks. For example, it is entirely expected (and logically circular) to find that patients who had sudden gains experienced better outcomes than patients without sudden gains--the patients with sudden gains all had gains, and so by definition improved, whereas the patients without sudden gains likely represent a mix of patients with gradual gains plus patients with no gains, and so improved less on average. A more interesting and important question is whether patients with sudden gains have better outcomes than patients who had gradual gains *of equivalent magnitude pre- to post-treatment* (e.g., decrease of 20 BDI points from pre- to post-treatment). For example, Vittengl *et al.* [46] differentiated patients with linear, log-linear, one-step (a form of sudden gains), and undefined change patterns during acute phase CT. The first three groups had largely equivalent outcomes and all were superior to the undefined change group. Because measures of depressive cognitive content and social-interpersonal functioning correlate substantively with depressive symptom measures [58], patients with sudden gains also show relatively better functioning in these domains [16]. Similarly, cognitive changes before sudden gains [5] might be compared to cognitive changes before gradual reductions in symptoms of equivalent magnitude to understand whether there is anything unique about possible cognitive mechanisms in the suddenness of gains.

Monte Carlo simulations are not equivalent to study of real patient data, which is both a strength and weakness. Simulations contain exactly (and only) the information programmed into them, allowing researchers to investigate properties of statistical tests, such as those for sudden gains. We programmed population means, variances, and covariances for repeated symptom scores in our simulations. Some of the simulations had constant means across time (no gains), some had linearly and log-linearly decreasing means (gradual gains), and some

had abruptly and stably decreasing means (sudden gains). Our analyses did not attempt to model the unknown reasons that patients display such patterns clinically (e.g., due to insights, or biochemical or social-interpersonal events). Consequently, our analyses inform measurement theory for sudden gains, not clinical theory or therapeutic technique directly. Nonetheless, as scientists we assert that appropriate measurement of symptom changes (e.g., differentiating sudden from gradual and no gains) is prerequisite to productive clinical study, and thus we offer empirical guidance for improving study of sudden gains.

We did not attempt to model all measurement issues that may influence detection of sudden gains, suggesting areas for future research. For example, missing data are common in clinical settings (e.g., due to patient fatigue) and complicate detection of sudden gains. Missing data may obscure the magnitude and suddenness of gains, as well as muddy judgments about the stability of gains. Missing data likely affect the Type I error rate and power of each of the criterion sets tested and could be explored in future analyses. In addition, depression symptom measures are limited not only in their reliability (modeled in the current analyses as retest stability) but also in their validity (not modeled here; e.g., convergence between clinician and patient reports [16]). Multivariate procedures including structural equation modeling and factor analyses have the potential to account for measurement error and focus sudden gains tests on valid symptom variance.

We aim to encourage the field to better differentiate sudden gains from other patterns of change and from no change in depressed patients' symptoms. We do not dispute that real sudden gains can occur during treatment of MDD, as have been reported in the literature. However, we submit that existing sudden-gains criteria may have been inadequately sensitive (missing some sudden gains) and specific (misidentifying many other symptom changes as sudden gains). Moreover, sudden gains may not add information beyond the previously identified, arguably simpler concept of early rapid response [30]. We did not test all possible implementations of sudden-gains criteria, but we did simulate a relatively large, arguably realistic range of conditions, and found inadequate psychometric performance for three popular criterion sets in each scenario. Consequently, our results set the stage for careful work in defining sudden gains empirically before theorizing about their possible origins in disease or treatment processes. The number of sudden gains identified with more rigorous criteria is likely to be considerably smaller than reported in the literature to date. On the other hand, sudden gains identified with rigorous criteria also are more likely to be "real" and to have important implications for long-term outcomes. Focusing on real, not false-positive, sudden gains may lead more quickly to improved knowledge of the causes and consequences of this interesting and potentially important phenomenon.

Acknowledgments

This report was supported by Grants Number K24 MH001571, R01 MH58397, R01 MH69619 (to Robin B. Jarrett, Ph.D.) and R01 MH58356 and R01 MH69618 (to Michael E. Thase, M.D.) from the National Institute of Mental Health (NIMH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIMH or the National Institutes of Health. We also appreciate the careful review by members of the trial's Data Safety and Monitoring Board. We are indebted to our research teams and our colleagues at The University of Texas Southwestern Medical Center at Dallas, the University of Pittsburgh (where Dr. Thase was located during patient accrual), and the University of Pennsylvania (Dr. Thase's current affiliation). We appreciate the participation of colleagues, previously named, and study participants without whom such research could not have been completed.

All authors contributed substantially to this work and approved its publication. Dr. Vittengl reviewed the literature, conducted the analyses, and wrote the manuscript. Drs. Clark, Thase, and Jarrett interpreted findings from the literature review and analyses, and edited and revised the manuscript.

References

1. Beck, AT.; Rush, AJ.; Shaw, BF.; Emery, G. Cognitive therapy of depression. New York: Guilford; 1979.
2. Craighead, WE.; Sheets, ES.; Brosse, AL.; Ilardi, SS. Psychosocial treatments for major depressive disorder. In: Nathan, PE.; Gorman, JM., editors. A guide to treatments that work. 3. New York: Oxford; 2007. p. 289-307.
3. Vittengl, JR.; Jarrett, RB. Major depressive disorder. In: Hofman, SG., editor. The Wiley handbook of cognitive behavioral therapy Part Two. New York: Wiley; 2013. p. 1-29.
4. Kazdin AE. Mediators and mechanisms of change in psychotherapy research. *Annu Rev Clin Psychol.* 2007; 3:1-27. [PubMed: 17716046]
5. Tang TZ, DeRubeis RJ. Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *J Consult Clin Psychol.* 1999; 67:894-904. [PubMed: 10596511]
6. Lorenz TA, Pulverman CS, Meston CM. Sudden gains during patient-directed expressive writing treatment predicts depression reduction in women with history of childhood sexual abuse: Results from a randomized clinical trial. *Cognitive Therapy and Research.* 2012; 37:690-6. [PubMed: 25484475]
7. Hofmann SG, Schulz SM, Meuret AE, Moscovitch DA, Suvak M. Sudden gains during therapy of social phobia. *J Consult Clin Psychol.* 2006; 74:687-97. [PubMed: 16881776]
8. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry.* 1961; 4:561-71. [PubMed: 13688369]
9. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960; 23:56-62. [PubMed: 14399272]
10. Driessen E, Hollon SD. Cognitive behavioral therapy for mood disorders: efficacy, moderators and mediators. *Psychiatr Clin North Am.* 2010; 33:537-55. [PubMed: 20599132]
11. Elliott R. Psychotherapy change process research: realizing the promise. *Psychother Res.* 2010; 20:123-35. [PubMed: 20099202]
12. Messer, SB.; McWilliams, N. Insight in psychodynamic therapy: Theory and assessment. In: Castonguay, LG.; Hill, C., editors. *Insight in psychotherapy.* Washington, DC: American Psychological Association; 2007. p. 9-29.
13. Grosse Holtforth, M.; Castonguay, LG.; Boswell, JF.; Wilson, LA.; Kakouros, AA.; Borkovec, TD. Insight in cognitive-behavioral therapy. In: Castonguay, LG.; Hill, C., editors. *Insight in psychotherapy.* Washington, DC: American Psychological Association; 2007. p. 57-80.
14. Tang TZ, DeRubeis RJ, Beberman R, Pham T. Cognitive changes, critical sessions, and sudden gains in cognitive-behavioral therapy for depression. *J Consult Clin Psychol.* 2005; 73:168-72. [PubMed: 15709844]
15. Tang TZ, Derubeis RJ, Hollon SD, Amsterdam J, Shelton R. Sudden gains in cognitive therapy of depression and depression relapse/recurrence. *J Consult Clin Psychol.* 2007; 75:404-8. [PubMed: 17563157]
16. Vittengl JR, Clark LA, Jarrett RB. Validity of sudden gains in acute phase treatment of depression. *J Consult Clin Psychol.* 2005; 73:173-82. [PubMed: 15709845]
17. Hardy GE, Cahill J, Stiles WB, Ispan C, Macaskill N, Barkham M. Sudden gains in cognitive therapy for depression: a replication and extension. *J Consult Clin Psychol.* 2005; 73:59-67. [PubMed: 15709832]
18. Barlow, DH.; Nock, MK.; Hersen, M. Single case experimental designs. 3. New York: Pearson; 2009.
19. Gaynor ST, Weersing VR, Kolko DJ, Birmaher B, Heo J, Brent DA. The prevalence and impact of large sudden improvements during adolescent therapy for depression: a comparison across cognitive-behavioral, family, and supportive therapy. *J Consult Clin Psychol.* 2003; 71:386-93. [PubMed: 12699032]

20. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol.* 1991; 59:12–9. [PubMed: 2002127]
21. Stiles WB, Leach C, Barkham M, Lucock M, Iveson S, Shapiro DA, Iveson M, Hardy GE. Early sudden gains in psychotherapy under routine clinic conditions: practice-based evidence. *J Consult Clin Psychol.* 2003; 71:14–21. [PubMed: 12602421]
22. Drymalski WM, Washburn JJ. Sudden gains in the treatment of depression in a partial hospitalization program. *J Consult Clin Psychol.* 2011; 79:364–8. [PubMed: 21381809]
23. Kelly MA, Roberts JE, Ciesla JA. Sudden gains in cognitive behavioral treatment for depression: when do they occur and do they matter? *Behav Res Ther.* 2005; 43:703–14. [PubMed: 15890164]
24. Busch AM, Kanter JW, Landes SJ, Kohlenberg RJ. Sudden gains and outcome: a broader temporal analysis of cognitive therapy for depression. *Behav Ther.* 2006; 37:61–8. [PubMed: 16942961]
25. Hopko DR, Robertson SM, Carvalho JP. Sudden gains in depressed cancer patients treated with behavioral activation therapy. *Behav Ther.* 2009; 40:346–56. [PubMed: 19892080]
26. Hunnicutt-Ferguson K, Hoxha D, Gollan J. Exploring sudden gains in behavioral activation therapy for Major Depressive Disorder. *Behav Res Ther.* 2012; 50:223–30. [PubMed: 22336434]
27. Kelly MA, Roberts JE, Bottonari KA. Non-treatment-related sudden gains in depression: the role of self-evaluation. *Behav Res Ther.* 2007; 45:737–47. [PubMed: 16905116]
28. Kelly MA, Cyranowski JM, Frank E. Sudden gains in interpersonal psychotherapy for depression. *Behav Res Ther.* 2007; 45:2563–72. [PubMed: 17765867]
29. Andrusyna T, Luborsky L, Pham T, Tang T. The mechanisms of sudden gains in supportive-expressive therapy for depression. *Psychotherapy Research.* 2006; 16:526–36.
30. Thomas C, Persons JB. Sudden gains can occur in psychotherapy even when the pattern of change is gradual. *Clin Psychol: Sci Pract.* 2013; 20:127–142.
31. Lipsey, MW.; Wilson, DB. Practical meta-analysis. Thousand Oaks, CA: Sage; 2001.
32. Aderka IM, Nickerson A, Bøe HJ, Hofmann SG. Sudden gains during psychological treatments of anxiety and depression: a meta-analysis. *J Consult Clin Psychol.* 2012; 80:93–101. [PubMed: 22122290]
33. Everitt, BS.; Wessely, S. Clinical trials in psychiatry. 2. West Sussex, England: Wiley; 2008.
34. Kenny DA, Judd CM. Consequences of violating the independence assumption in analysis of variance. *Psychol Bull.* 1986; 99:422–31.
35. Greenfield MF, Gunthert KC, Haaga DA. Sudden gains versus gradual gains in a psychotherapy training clinic. *J Clin Psychol.* 2011; 67:17–30. [PubMed: 21058390]
36. Barkham M, Stiles WB, Shapiro DA. The shape of change in psychotherapy: longitudinal assessment of personal problems. *J Consult Clin Psychol.* 1993; 61:667–77. [PubMed: 8370863]
37. Uher R, Muthén B, Souery D, Mors O, Jaracz J, Placentino A, Petrovic A, Zobel A, Henigsberg N, Rietschel M, Aitchison KJ, Farmer A, McGuffin P. Trajectories of change in depression severity during treatment with antidepressants. *Psychol Med.* 2010; 40:1367–77. [PubMed: 19863842]
38. Dimidjian S, Hollon SD, Dobson KS, Schmaling KB, Kohlenberg RJ, Addis ME, Gallop R, McGlinchey JB, Markley DK, Gollan JK, Atkins DC, Dunner DL, Jacobson NS. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol.* 2006; 74:658–70. [PubMed: 16881773]
39. Segal ZV, Kennedy S, Gemar M, Hood K, Pedersen R, Buis T. Cognitive reactivity to sad mood provocation and the prediction of depressive relapse. *Arch Gen Psychiatry.* 2006; 63:749–55. [PubMed: 16818864]
40. Beck AT, Steer RA, Garbin MG. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clin Psychol Rev.* 1988; 8:77–100.
41. Howard KI, Kopta MS, Krause MS, Orlinsky DE. The dose-effect relationship in psychotherapy. *Am Psychol.* 1986; 41:159–64. [PubMed: 3516036]
42. Lutz W, Martinovich Z, Howard KI, Leon SC. Outcomes management, expected treatment response, and severity-adjusted provider profiling in outpatient psychotherapy. *J Clin Psychol.* 2002; 58:1291–304. [PubMed: 12357444]

43. Jarrett RB, Kraft D, Doyle J, Foster BM, Eaves GG, Silver PC. Preventing recurrent depression using cognitive therapy with and without a continuation phase: a randomized clinical trial. *Arch Gen Psychiatry*. 2001; 58:381–8. [PubMed: 11296099]
44. Jarrett RB, Minhajuddin A, Gershenfeld H, Friedman ES, Thase ME. Preventing depressive relapse and recurrence in higher-risk cognitive therapy responders: a randomized trial of continuation phase cognitive therapy, fluoxetine, or matched pill placebo. *JAMA Psychiatry*. 2013; 70:1152–60. [PubMed: 24005123]
45. Jarrett RB, Thase ME. Comparative efficacy and durability of continuation phase cognitive therapy for preventing recurrent depression: design of a double-blinded, fluoxetine- and pill placebo-controlled, randomized trial with 2-year follow-up. *Contemp Clin Trials*. 2010; 31:355–77. [PubMed: 20451668]
46. Vittengl JR, Clark LA, Thase ME, Jarrett RB. Nomothetic and idiographic symptom change trajectories in acute-phase cognitive therapy for recurrent depression. *J Consult Clin Psychol*. 2013; 81:615–26. [PubMed: 23627652]
47. Watson D, Tellegen A. Toward a consensual structure of mood. *Psychol Bull*. 1985; 98:219–35. [PubMed: 3901060]
48. Fan, X.; Felsovalyi, A.; Sivo, SA.; Keenan, SC. SAS for Monte Carlo studies: A guide for quantitative researchers. Cary, NC: SAS Institute; 2002.
49. Harding KJ, Rush AJ, Arbuckle M, Trivedi MH, Pincus HA. Measurement-based care in psychiatric practice: a policy framework for implementation. *J Clin Psychiatry*. 2011; 72:1136–43. [PubMed: 21295000]
50. Hayes AM, Feldman GC, Beevers CG, Laurenceau JP, Cardaciotto L, Lewis-Smith J. Discontinuities and cognitive changes in an exposure-based cognitive therapy for depression. *J Consult Clin Psychol*. 2007; 75:409–21. [PubMed: 17563158]
51. Cohen J. A power primer. *Psychol Bull*. 1992; 112:155–9. [PubMed: 19565683]
52. West SG, Hepworth JT. Statistical issues in the study of temporal data: daily experiences. *J Pers*. 1991; 59:609–62. [PubMed: 1960644]
53. Kazdin, AE. Single-case research designs: Methods for clinical and applied settings. New York: Oxford; 2011.
54. Borckardt JJ, Nash MR, Murphy MD, Moore M, Shaw D, O'Neil P. Clinical practice as natural laboratory for psychotherapy research: a guide to case-based time-series analysis. *Am Psychol*. 2008; 63:77–95. [PubMed: 18284277]
55. Hayes AM, Laurenceau JP, Feldman G, Strauss JL, Cardaciotto L. Change is not always linear: the study of nonlinear and discontinuous patterns of change in psychotherapy. *Clin Psychol Rev*. 2007; 27:715–23. [PubMed: 17316941]
56. Singer, JD.; Willett, JB. Applied longitudinal data analysis: Modeling change and event occurrence. New York: Oxford; 2003.
57. Crocker, L.; Algina, J. Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston; 1986.
58. Clark LA, Vittengl JR, Kraft D, Jarrett RB. Shared, not unique, components of personality and psychosocial functioning predict depression severity after acute-phase cognitive therapy. *J Pers Disord*. 2003; 17:406–30. [PubMed: 14632375]

Figure 1.

Figure 2.

Table 1

Characteristics of Sudden Gains in Treatment of Unipolar Depression

Study	Treatment	N	Sudden Gains		Patients With versus Without Sudden Gains		Follow-up <i>d</i>
			Patients with	<i>M</i> BDI Decrease	End of Treatment <i>d</i>		
[24]	CT/FECT ¹	38	42%	9.3	BDI = -0.004 Recovery = -0.28		Not available.
[22]	PHP ²	664	41%	1.5 ^a	BASIS = 0.66 [*]		Not available.
[19]	CT ¹ SBFT ² NST ²	32 27 28	50% 26% 39%	10.8 10.7 11.8	Clinically significant improvement on the BDI = 0.29 (CT) 0.64 (SBFT) 0.90 [*] (NST)		Not available for sudden gains separate from pre-treatment gains.
[17]	CT ¹	76	41%	13.3	BDI = 0.79 [*] Recovery = 0.60 [*]		4-month BDI = 1.00 [*]
[25]	BA ²	26	50%	11.8	HRSD = 1.26 [*] BDI = 2.33 [*] Response = 1.34 [*]		3-month Response = 1.24 [*]
[26]	BA ²	42	36%	Not available.	QIDS-R = 0.48 Response = 0.74 [*] Remission = 0.27 ^b		Not available.
[23]	CT ¹	31	42%	Not available.	BDI = 0.42 Response = 0.42		Not available.
[27]	None, ³	60	60%	12.5	No treatment.		No treatment to follow up.
[28]	IP ¹ ²	185	34%	13.0	BDI = +/- 0.06 ^b HRSD = +/- 0.03 ^b Remission = -0.18		23-month Relapse = -0.22
[5]	CT ¹	61	39%	11.2	BDI = 1.00 [*] Recovery = 0.81 [*]		6-month BDI = 1.10 [*] 12-month BDI = 0.12 18-month BDI = 0.93 [*] 18-month relapse = +/- 0.13 ^b 6-months BDI = 0.48
[29]	SE ²	35	43%	10.5	BDI = 0.80 [*] Recovery = 0.72 [*]		
[14]	CT ¹ AT ²	46 37	43% 46%	10.2 11.5	BDI = 0.45 [*] ^c		Not available.
[15]	CT ¹	60	40%	11.0	BDI = 0.80 [*] HRSD = 0.80 [*] Response = 0.73 [*]		24-month Relapse/recr = 0.80 [*]

Study	Treatment	N	Sudden Gains Patients with	MBDI Decrease	End of Treatment <i>d</i>	Patients With versus Without Sudden Gains	Follow-up <i>d</i>
[30]	CT ¹	22	45%	10.9	Not available.		Not available.
[16] dataset 1	CT ¹	33	33%	9.7	BDI = 0.52* <i>d</i>		Not available.
	MED ²	32	47%	10.1	HRSD = 0.47* <i>d</i>		
	PBO ³	24	25%	11.2	IDSC = 0.55* <i>d</i>		
					IDSR = 0.57* <i>d</i>		
					Response = 0.49*		
[16] dataset 2	CT ¹	138	46%	10.2	Included above, <i>d</i>		24-month BDI = -0.49* 24-month Relapse/recr = -0.07

Note. BDI = Beck Depression Inventory. HRSD = Hamilton Rating Scale for Depression. IDS = Inventory of Depressive Symptomatology (IDSR = self-report, IDSC = clinician report). CT = cognitive therapy. AT = automatic thought treatment (a component of CT). BA = behavioral activation (a component of CT). SE = supportive-expressive therapy. NST = nondirective supportive therapy. FEET = functional-analysis enhanced CT (pooled with CT patients by Busch *et al.* [24]). SBFT = systemic behavioral family therapy. PHP = partial hospitalization program including CT, psychoeducation, and pharmacotherapy. MED = pharmacotherapy. PBO = pill placebo. We omitted studies of patients with primary diagnoses other than depression and patients with depression pooled with other diagnoses. When not provided directly in articles, effect size *d* was estimated from summary statistics using standard formulas [31]. Positive *ds* indicate lower symptoms / less relapse among patients with versus without sudden gains, whereas negative *ds* indicate higher symptoms / more relapse among patients with versus without sudden gains. We use the terms response, remission, recovery, relapse, and recurrence matching the cited studies' terminology, although definitions vary.

* $p < .05$.

¹Included in summary of CT datasets (see text).

²Included in summary of other active treatment datasets (see text).

³Included in summary of non-active control datasets (see text).

^aDepression score from the Behavior and Symptom Identification Scale-24 (BASIS).

^bThe direction of this non-significant effect is unclear in the published report.

^cTang *et al.* [14] controlled treatment condition and pre-treatment BDI in an omnibus test; *d* estimated from full model R^2 .

^dVittengl *et al.* [16] pooled patients because sudden gain rates and sizes did not differ significantly across treatments and datasets

Table 2

Proportion of False Positives Sudden Gains for Three Criterion Sets

Repeated BDI measures	Retest $r = .90$			Retest $r = .80$			Retest $r = .70$		
	T99	G03	K05	T99	G03	K05	T99	G03	K05
Population with no change in BDI scores									
10	.046	.055	.106	.092	.111	.348	.117	.143	.520
15	.079	.087	.140	.153	.170	.442	.193	.217	.637
20	.110	.119	.166	.215	.231	.498	.267	.290	.710
Population with gradual linear decreases in BDI scores									
10	.182	.199	.182	.289	.322	.463	.325	.368	.625
15	.332	.347	.198	.491	.514	.478	.529	.561	.642
20	.470	.482	.177	.642	.662	.418	.683	.707	.564
Population with gradual log-linear decreases in BDI scores									
10	.296	.411	.238	.434	.564	.484	.457	.587	.613
15	.389	.494	.257	.554	.662	.513	.584	.688	.638
20	.459	.558	.267	.638	.733	.512	.673	.759	.639

Note. $N = 10,000$ per table cell. BDI = Beck Depression Inventory. T99, G03, and K05 are sudden-gains criteria presented by Tang and DeRubeis [5], Gaynor *et al.* [19], and Kelly *et al.* [23], respectively. Sudden gains are drops in BDI scores between adjacent measurements that are (a) at least 7 points, (b) at least 25% of the pre-gain score, and (c) relatively stable. As detailed in the Method, criterion (c) is variously defined as a significant t -test comparing pre- and post-gain score means (Tang *et al.*); maintenance of at least 50% of the gain across the next two scores (Gaynor *et al.*); or a drop in scores at least 1.5 times the standard deviation of the patient's entire series of scores (Kelly *et al.*). Expected values in each cell are .025 based on nominal alphas of .05, two-tailed, for hypothesis tests.

Table 3

Power to Detect Sudden Gains for Three Criterion Sets

Repeated BDI measures	Retest $r = .90$			Retest $r = .80$			Retest $r = .70$		
	8pt	12pt	16pt	8pt	12pt	16pt	8pt	12pt	16pt
Tang and DeRubeis [5] Criteria									
10	.411	.717	.848	.412	.645	.807	.404	.597	.762
15	.466	.777	.906	.484	.717	.869	.481	.679	.828
20	.499	.807	.929	.535	.763	.902	.549	.737	.868
Gaynor <i>et al.</i> [19] Criteria									
10	.426	.733	.863	.428	.669	.827	.428	.625	.784
15	.477	.787	.912	.498	.737	.880	.506	.698	.840
20	.508	.810	.933	.550	.774	.906	.579	.772	.903
Kelly <i>et al.</i> [23] Criteria									
10	.440	.663	.781	.569	.677	.756	.678	.721	.748
15	.436	.667	.788	.615	.695	.767	.722	.744	.765
20	.432	.654	.771	.631	.693	.747	.749	.748	.760

Note. $N = 10,000$ per table cell. 8pt, 12pt, 16pt refer to 8, 12, and 16 point drops in Beck Depression Inventory (BDI) scores between measurements (small, moderate, and large sudden gains, respectively). There were no changes in population means other than those due to sudden gains. Values indicating with traditionally acceptable levels of power (i.e., $> .80$) are in **bold**. Sudden gains defined as drops in BDI scores between adjacent measurements that are (a) at least 7 points, (b) at least 25% of the pre-gain score, and (c) relatively stable. As detailed in the Method, criterion (c) is variously defined as a significant t -test comparing pre- and post-gain score means (Tang *et al.*); maintenance of at least 50% of the gain across the next two scores (Gaynor *et al.*); or a drop in scores at least 1.5 times the standard deviation of the patient's entire series of scores (Kelly *et al.*).