HHS Public Access

Author manuscript

Behav Ther. Author manuscript; available in PMC 2020 May 18.

Published in final edited form as:

Behav Ther. 2020 May ; 51(3): 413–423. doi:10.1016/j.beth.2019.07.003.

Exposing Worry's Deceit: Percentage of Untrue Worries in Generalized Anxiety Disorder Treatment

Lucas S. LaFreniere,
The Pennsylvania State University

Michelle G. Newman
The Pennsylvania State University

Abstract

Theories of cognitive therapy have long proposed that those with generalized anxiety disorder (GAD) have inaccurate expectations. By challenging them with objective evidence, symptoms are thought to decrease. To test these premises, this study used ecological momentary assessment (EMA) during the Worry Outcome Journal (WOJ) treatment to determine the percentage of GAD worries that did not come true. We then analyzed the association between participants' untrue worry percentages and GAD symptom change across treatment. Twenty-nine participants with GAD recorded worries when prompted for 10 days, reviewed them online nightly, and tracked their worry outcomes across 30 days. These recordings were then coded by independent raters. Analyses applied bias-correct bootstrapping path analysis on slopes extracted from longitudinal linear mixed models. Primary results revealed that 91.4% of worry predictions did not come true. Higher percentages of untrue worries significantly predicted lower GAD symptoms after treatment, as well as a greater slope of symptom reduction from pre- to post-trial. Participants' average expected likelihoods of worries coming true were much greater than actual observed likelihoods. The most common percentage of untrue worries per person was 100%. Thus, worries in those with GAD were mostly inaccurate. Greater evidence of this inaccuracy predicted greater improvement in treatment. As theorized, disconfirming false expectations may significantly contribute to treatment's effect.

Keywords

generalized anxiety disorder; worry; cognitive therapy; worry outcome monitoring; self-monitoring

Numerous efficacy studies have established the power of cognitive behavioral therapies (CBTs), evincing moderate to large effect sizes (Butler, Chapman, Forman, & Beck, 2006; Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012) and marking many as empirically supported treatments (ESTs; Chambless & Ollendick, 2001). Yet less research has examined how these therapies create change. Many scholars have argued that we should prioritize the study of treatment predictors, processes, and targets, elucidating the factors responsible for

There are no conflicts of interest or financial disclosures related to this work.

therapy outcome (e.g., Arch & Craske, 2008; Kazdin, 2008; Smits, Julian, Rosenfield, & Powers, 2012). Identifying treatment's active ingredients can then foster its augmentation and development. Unfortunately, it is still an empirical question whether many current therapies truly work via their theorized processes. For example, one of the central premises of cognitive therapy (CT) is that by challenging unrealistic thoughts with objective evidence, symptoms can be reduced (Beck, 1979). CT fundamentally assumes that 1) clients' expectations are inconsistent with reality and 2) greater recognition of that inconsistency leads to better outcomes. Yet despite being promoted for decades, the disconfirmation of faulty beliefs has not been examined directly in the treatment of one prevalent syndrome: Generalized anxiety disorder (GAD).

GAD's primary criterion—worry—is a cognition that often contains unrealistic content. Those with GAD are thought to continually expect unreasonable or exaggerated negative future events, regardless of their improbability. Accordingly, CT for GAD teaches clients to examine worries as hypotheses and test them against reality, believing that clients' anxious predictions will largely be exposed as false (Beck & Emery, 1985). Yet well-controlled scientific study has not established that GAD worries are, in fact, irrational. Although excessive worrying and GAD have been associated with higher likelihood estimations for negative outcomes compared to controls (Berenbaum, Thompson, & Bredemeier, 2007; Berenbaum, Thompson, & Pomerantz, 2007; Butler & Mathews, 1983; MacLeod, Williams, & Bekerian, 1991), the only report on worry's accuracy comes from a secondary description of an unpublished study. In a theoretical article, Borkovec, Hazlett-Stevens, and Diaz (1999) briefly mention the results of two questionnaire-based worry outcome monitoring assessments. In the first, 67.9% of the worry predictions of 29 undergraduate participants with GAD were rated as *not* having come true over a two-week period. In the second, 17 clients tracked worries and their outcomes during treatment. For this sample, 85% of worries were rated as having turned out "better than expected." Unfortunately, the strict accuracy of their predictions was not reported. When worries did *not* turn out better than expected, clients still coped better than expected 79% of the time. Although these reports are promising, they were not peer-reviewed, few methodological details were reported, and no mention was made of coders or their reliability. Yet not only is the accuracy of GAD worry predictions unclear, but their association with change during CT has gone unstudied as well. A true test of the theory underlying CT for GAD would require both establishing that worries are commonly false and that recognizing this fact predicts change.

One cognitive treatment for GAD is specifically designed to target inaccurate predictions: The Worry Outcome Journal (WOJ). The WOJ is an ecological momentary intervention in which clients record their worrisome predictions in daily life, rate the in-the-moment costs of each worry (such as distress and interference), and then track their actual outcomes as objectively as possible (LaFreniere & Newman, 2016). It is an augmented form of worry outcome monitoring, a common technique in CT for GAD (Borkovec & Newman, 1998). By monitoring worry outcomes and costs with the WOJ, those with GAD are meant to recognize that their worries are both highly unlikely and costly. Thus, their worry is not worthwhile (the costs of worry outweigh any benefits). In classic CT fashion, the WOJ is meant to provide objective, personalized evidence that challenges the unrealistic predictions associated with worry. In a randomized controlled trial (RCT), the WOJ outperformed an

active control in reducing worry over 10 days and at 30-day follow-up (LaFreniere & Newman, 2016). Yet the primary therapeutic process of the WOJ—revealing worry's inaccuracy—has not been directly tested.

Fortunately, the entries made in the WOJ offer useful ecological momentary assessment (EMA) data that can address this question. Primarily, clients recorded their worrisome predictions in real time, then reported on their outcomes. Thus, WOJ entries can elucidate clients' perceived accuracy of worry after attending to the reality of events—a core process of CT. Moreover, clients also recorded the distress associated with each worry, its duration of cognitive interference, and two subjective probabilities that worries would actualize. These latter ratings included an "emotional," anxiety-driven probability and a "logical," rationally-formed probability. All of this data can inform in-the-moment GAD processes, including the perceived costs of worries, clients' recognition of the gap between emotionally-informed and rationally-informed expectations, and—most importantly—the degree of worry's inaccuracy.

Thus, the study at hand aimed to use WOJ data to address several questions. First, we aimed to more accurately assess how frequently GAD worries did not come true using EMA. We also sought to report other in-journal data from the WOJ (number of unique worries, distress, interference, etc.) at a time very early in treatment, when few to no worries had been tested. Second, we aimed to empirically determine if disconfirmation of poor expectations truly did predict outcome and change in GAD symptoms during treatment. Is training a scientific, evidence-based mind actually associated with change in CT? Understanding the processes of this common therapeutic approach can both further CT's empirical support and allow for its improvement and extension in clinical practice.

To this end, the current study identified worry outcome likelihoods and used bootstrapping path analysis (Preacher & Hayes, 2008) to test the theories undergirding the WOJ in data from its sole RCT (LaFreniere & Newman, 2016). First, we hypothesized that the percentage of worries that did not come true would be relatively high for GAD participants (greater than 75%). Second, we expected that a higher percentage of untrue worries would predict better treatment outcome—lower GAD symptoms post-trial and greater slopes of change in GAD symptoms across the trial.

Method

The current study draws on data from LaFreniere and Newman's (2016) RCT comparison of the WOJ to an active control. It was approved by an Institutional Review Board. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Participants

The condition examined by this study included 29 undergraduate student participants, 26 women and 3 men. Participants met full DSM-IV criteria for GAD (which are identical to that of DSM-5;American Psychiatric Association, 2013) as assessed by the Generalized Anxiety Disorder Questionnaire for DSM-IV (GAD-Q-IV; Newman et al., 2002) at both a

mass screening and during baseline assessment. Inclusion criteria required being over 18 years of age and speaking English. Participants were required to complete 80% or more WOJ entries after two days as assessed by a compliance check, which all passed. Where ethnicity is concerned, the sample was 75.9% White, 10.3% Black, 6.9% Hispanic, and 6.9% Asian. One participant dropped out due to discomfort with recording worries. Although she reported no harm, she was informed of options for treatment.

Self-Report Measures

The Generalized Anxiety Disorder Questionnaire for DSM-IV (GAD-Q-IV; Newman et al., 2002). The GAD-Q-IV was used both to select participants for the study and to measure general GAD symptoms at outcome. The GAD-Q-IV is a 9-item self-report measure designed to assess the degree of GAD symptoms as defined by the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; APA, 1994). To its credit, the GAD-Q-IV is the only self-report measure that evaluates all DSM GAD symptoms (Rodebaugh, Holaway, & Heimberg, 2008). In addition to items evaluating the presence, amount, and topic matter of excessive and uncontrollable worry, the GAD-Q-IV addresses six somatic symptoms (e.g., tension, restlessness, etc.), as well as distress and interference. The current study profits from the measure's ability to be scored both based on diagnostic criteria and dimensionally (Moore, Anderson, Barnes, Haigh, & Fresco, 2014; Newman et al., 2002). The GAD sample was selected using the criterion-based scoring system. Treatment effects were assessed with its continuous score. Research has shown that both functions of the measure are valid and reliable. Its internal consistency is robust (Cronbach's alpha = .94; Newman et al., 2002). Furthermore, it has shown sufficient retest reliability in clinical classification over a two week period (92% of the sample; Newman et al., 2002). It has also demonstrated convergent validity with measures of worry and discriminant validity against measures of social anxiety and panic disorder. Rates of kappa agreement between the GAD-Q-IV and the Anxiety Disorders Interview Schedule (ADIS) are good as well (κ = .67). The measure also demonstrates sensitivity to change as a result of treatment (Andersson et al., 2012; Chen, Liu, Rapee, & Pillay, 2013; Paxling et al., 2011)

Procedure

An online subject pool screening identified university students who met criteria for GAD based on a criterion scoring of the GAD-Q-IV. Qualifying students were invited to participate in the study through a mass email. Those wishing to enroll were randomly assigned to either the WOJ or an active control condition. The current study only addresses WOJ participants, because only these participants monitored whether or not their worries came true.

Participants first presented at the research facility and completed the GAD-Q-IV. They then underwent a 30-minute training session wherein study personnel read from a script as participants followed along. Participants in both conditions were given a treatment rationale to replicate actual cognitive-behavioral clinical practice and to control for treatment expectancy. The rationale for the WOJ read: "Many people who regularly worry believe their worry is useful to them. As you pay attention to how upsetting, disrupting, and costly your worries are, and as you see clear evidence in your life that the things you worry about

actually do not happen, you will recognize the uselessness of worrying and begin to engage in it less. Without these anxious thoughts in your life, your anxiety should also lessen." The active control was also given a rationale. For the WOJ group, the training then defined worry for participants, describing it as "a repetitive anxious thought that an event in the future will turn out badly . . . apprehensive expectation . . . repetitive and hard to control in your mind." They were given several examples. Participants were also instructed that "a worry is *not* just a fleeting concern about the future," but rather must "forecast a negative outcome." They were then provided examples of thoughts that were *not* worries. They were asked to only record specific, concrete, testable worries that they could be sure either came true or did not come true within the 30 days of the study; they were given examples of both ideal and improper entries. WOJ users were also taught how to track outcomes. Next, they were taught how to complete journal ratings and how to use the online system. Lastly, they were given the chance to ask questions, as well as contact information for any later inquiries. The entire session script was included in the initial pages of each WOJ for participants' reference.

The next morning participants began making prompted entries using a commercial mass texting service. For ten days they recorded worries or thoughts in their paper journals when prompted randomly (four times/day within four timeframes: Morning 8:00 AM-12:00 PM, early afternoon 12:00–3:00 PM, late afternoon 3:00–6:00 PM, and evening 6:00–9:00 PM) by text message on their personal mobile phone. They were permitted to make additional entries at any time, but were required to record at least one at the prompted time. Recorded worries or thoughts must have occurred within the two hours prior to the prompt. For each recorded entry, they also noted the prompt time and date.

For WOJ-recorded worries, participants noted their specific, testable predictions about the future. Next, they rated the degree of distress caused by the worry (worry-related distress) on a scale of 1 (no distress) to 7 (severe distress) and their best estimate answering the question, "How much time did that worry take up since I first had the worry?" (recent cognitive interference duration). After completing present-moment cost ratings, they gave two probability estimates with respect to the worry outcome. First, they rated how likely they *felt* that the feared outcome would occur, according to their "gut feelings or intuition" (emotional likelihood of worry outcome). Next, they rated how likely "a person would *logically* conclude the worried outcome would occur if they were thinking as realistically as possible" (logical likelihood of worry outcome). With this rating they were instructed to ask, "If the most rational person in the world were to give a probability as to how likely this event would come true, what would it be?" The purpose of these two ratings was to train participants in realistic prediction formation, draw attention to the discrepancy between their subjective forecasting and an objective prediction attempt, and to teach distance from worry by taking the perspective of "the most rational person." All in-journal ratings were made first in participants' paper journals. They were strongly encouraged to record a worry immediately when prompted, but if they had not experienced a worry since the last prompt time, they were allowed to write "No worry."

Each night at 10:00 PM participants copied all content from their day's paper recordings into an online survey via PsychData. They also indicated the percentage of thinking time spent on each specific worry across the entire day. Next, they were asked to review all

recorded worries to date, to note if any of these worries had led to a new or unrecorded outcome, and if so, if their feared outcome came true. They also rated if the outcome was "as bad as, worse than, or better than expected." They could note these outcomes at any time.

After two days of recording and online entry, participants received a phone call to check compliance rates. They were asked the total number of paper and online entries made and were assessed for any study-related harm. They were required to complete at least 80% of their entries by this time or they were eliminated from the study. No participant had a compliance rate below 80%. Consequently, no one was removed from the study due to low compliance.

The day after the ten-day intervention period they completed the GAD-Q-IV again via PsychData. Participants then returned their paper records to the lab. Twenty days after the final day of worry recording they were provided the entire content of their journals electronically. They were again asked to review each entry and provide the information described above, including if any worries came true in the past 20 days and whether the worries turned out as bad as, worse than, or better than expected. They then completed the outcome measure again. Lastly, participants were compensated with course-specific research credit.

Worry Outcome Coding and Journal Entry Data

Two raters independently coded worry outcome incidence for the WOJ group. Each rater counted the number of worries recorded, the number of worries that came true, the number of outcomes that were perceived as turning out better than expected, as expected, and worse than expected, and the number of worrisome predictions that were untestable within 30 days (e.g., "I will never get married"; "I won't become an engineer"; etc.). Specific rule-based criteria for determining what counted as a worry were established a priori. These rules were: 1) A statement must make a prediction to be counted as a worry; 2) the same worry cannot be counted twice if it is repeated in subsequent entries; 3) if a prediction has a separate test of whether it came true or not, then it is considered a separate worry; 4) if a new entry has the same test of outcome as a past entry, then it is a repeated worry; 5) a statement or entry is considered "double-barreled" if it has multiple independent predictions within it and corresponding tests of those predictions. These multiple, testable predictions must be counted as multiple worries. For example, an entry that reads "I will fail my Monday and Thursday Spanish quizzes this week" counted as two worries, since it made two predictions and had two tests of outcome. Raters were extensively trained on what counted as a worry or certain outcome type and what did not. Prior to final coding, the primary investigator (serving as a master coder) and the two raters each coded the data of two randomly chosen participants. Raters' results were compared to those of the master coder and discrepancies were discussed and resolved. Afterward, the two raters independently coded all journal data. Although participants were instructed to only record worries that could be tested within the 30 days of the trial, 16 worries across the entire sample did not have the chance to be tested in the 30-day timeframe. Thus, 16 was subtracted from the total number of worries counted. The overall percentage of testable worries that did not come true was calculated by taking the total number of worries that did not come true divided by the total number of worries

counted after subtracting the 16 untestable worries. Interrater reliability was high for each result (total worries, ICC(2) = 0.920; number of worries that did not come true, ICC(2) = 0.927; worries that turned out better than expected, ICC(2) = 0.980; worries that turned out as badly as expected, ICC(2) = 0.984; worries that turned out worse than expected, ICC(2) = 0.934). Percentages of untrue worries were also calculated for each individual participant within their person-specific data with the same method as the overall percentage. These values served as the "per person" percentages used in inferential analyses. Each reported finding is the average of the two raters' calculations.

In our descriptive results, we present both worry outcome data as well as averaged data from participant ratings in their early WOJ entries. The early rating means, standard deviations, and ranges presented in Table 1 are based on the average of each participant's ratings across the first two days of the trial (the mean of eight ratings). Ratings from the first two days were averaged in order to capture these worry-related experiences very early in treatment—a time when the WOJ was likely to have the least influence, with few to no worries having actually been tested. These two-day averages may approximate normative GAD worry experiences assessed in participants' natural environments, but treatment influence is possible.

2.6 Planned Statistical Analyses

Statistical analyses utilized both *R* and *Mplus 7* software. 6.9% of values were missing overall. We first conducted paired samples t-tests comparing average worry ratings on day one to average worry ratings on day two. These tests were meant to verify a lack of treatment-related change in these ratings, supporting the notion that these ratings may approximate those of normative worries. For all inferential analyses, we employed bias-corrected bootstrapping regression-based path analysis (Preacher & Hayes, 2008) with full information maximum likelihood (FIML) estimation to account for missingness. All variables in all models were observed. All reported confidence intervals are 95% bootstrap confidence intervals. Acceptable power was considered to be .80 or greater. Almost all participants had high percentages of untrue worries near or at 100%. Consequently, we corrected for a negatively skewed distribution with an arcsine transformation in all analyses including untrue worry percentages, as is recommended for percentage data (Cohen & Cohen, 1983). GAD-Q-IV scores and slopes of GAD-Q-IV change were both normally distributed. To confirm that percentages of untrue worries were not merely a proxy for GAD severity, we ran a bootstrapping path analysis where pre-trial GAD-Q-IV scores predicted participants' percentages of untrue worries. For our primary analyses, we did two separate bootstrapping path analyses: We first regressed post-trial GAD-Q-IV scores on participants' percentage of untrue worries during treatment with 10,000 bootstraps. We then regressed individual slopes of change in GAD-Q-IV scores from pre- to post-trial on percentage of untrue worries with 10,000 bootstraps. Power analysis with *G*Power* software demonstrated acceptable power for both analyses. We used multilevel modeling in the *R* statistical program's *lme4* and *lmerTest* packages to extract each individual's slope of change in GAD-Q-IV scores from pre- to post-trial. To do so, we specified a linear mixed effects model where GAD-Q-IV scores were predicted by a fixed effect of intercept, a random effect of intercept, and a random effect of time trend from pre- to post-trial. The fixed effect of time

trend was not specified so that its common influence would not be removed from individuals' slopes of change. To acquire participants' slopes of change, we then extracted the random effects of time trend on GAD-Q-IV scores for each individual. These slopes were then used as the outcome of our predictor model. Recall that analyses presented in the original RCT already demonstrated significant change in GAD-Q-IV scores from pre- to post-trial (LaFreniere & Newman, 2016). We report unstandardized coefficients for path betas. Cohen's *d* effect sizes were calculated by the formula $d = B/(\sqrt{(n)} * SE)$.

Results

Descriptive Analyses

Averaged across coders, 91.39% of all worries did not come true (91.08% for the first coder and 91.70% for the second). In other words, 8.61% of worries came true. In regard to the percentages calculated per person, the average percentage of untrue worries was 89.60% per person ($SD = 13.12\%$; 10.4% true worries). Per person untrue worry percentages ranged from 53% untrue worries to 100%. The mode of the percentage of untrue worries per person was 100%: For 7 participants, none of their worries came true. The median was 93.50% untrue worries. Of all worries that did come true, 30.10% were rated as turning out better than expected. The average number of unique, distinct worries per person (that had a chance to be tested) was 34.30 ($SD = 18.00$), ranging from 12.50 to 100.50 (the total number of unique worries did not influence any study variable of interest[1]). Other worry outcomes and WOJ rating descriptive results are presented in Table 1. Average worry distress on day one did not differ from average worry distress on day two ($t(28) = 0.261$, $p = 0.796$, $d = .055$). Similarly, day one ratings did not significantly differ from day two ratings for recent cognitive interference duration ($t(28) = -0.396$, $p = 0.695$, $d = -0.083$), cognitive interference duration across the entire day ($t(28) = 0.1.06$, $p = 0.917$, $d = 0.020$), emotional likelihood of worry outcome ($t(28) = 0.265$, $p = 0.793$, $d = 0.047$), or logical likelihood of worry outcome ($t(28) = 0.070$, $p = 0.944$, $d = 0.010$). Refer to Table 2 for zero order correlations between study variables. In contrast to the primary study analyses, these correlations use pairwise deletion rather than accounting for missingness with FIML. They also do not include bias-corrected bootstrapping. Note that pre-trial GAD symptoms were not significantly correlated with percentage of untrue worries. This result suggests untrue worry percentage was not merely a proxy for GAD symptom or worry severity[2].

Predictor Analyses

A bias-corrected bootstrap regression demonstrated that higher percentages of worries that did not come true significantly predicted lower GAD symptoms at post-trial ($B = -69.729$, bootstrap CI = [−165.511, −4.957], $d = -0.372$). Note that percentage of untrue worries also

[1]According to a bias-corrected bootstrapping path analysis with FIML estimation, the percentage of untrue worries was not related to the number of unique worries per person ($B = .000$, $p = .293$, 95% bootstrap CI = [.000, .001]. Cohen's d could not be accurately calculated due to an unstandardized beta coefficient of 0, yet the *R*-squared value was only 0.050. The simple Pearson correlation was 0.224 and it was not significant ($p = .293$). In addition, number of unique worries per person did not predict response to treatment (i.e., pre-post slope of GADQ; $B = 0.012$, CI = [−.001, .043], $d = .018$), nor was it associated with pre-trial GADQ scores ($B = 0.026$, CI = [−.022, .087), $d = .031$).

[2]Note that pre-trial Penn State Worry Questionnaire scores were not significantly correlated with untrue worry percentage either ($r = -.196$, $p = .359$).

predicted post-trial symptoms when controlling for baseline symptoms ($B = -53.509$, bootstrap CI = [−129.395, −3.220], $d = -0.307$). A bias-corrected bootstrap regression also demonstrated that higher percentages of untrue worries predicted greater negative slopes in GAD symptoms from pre-trial to post-trial ($B = -19.679$, bootstrap CI = [−43.458, −0.854], $d = -0.336$). As hypothesized, participants who experienced higher percentages of their worries not coming true had lower GAD symptoms after treatment and greater reductions in symptoms across treatment.

Discussion

The theories supporting cognitive therapy (CT) propose that those with GAD have inaccurate expectations and beliefs. CT is thought to reduce symptoms by correcting these expectations with evidence. One CT intervention, the Worry Outcome Journal (WOJ), proposes that if those with GAD track the actual outcomes of their worrisome predictions and find their worries improbable, worry and anxiety will decrease. Yet controlled research has never examined whether GAD expectations truly are unrealistic, or whether correcting them predicts improvement. Drawing from an RCT of the WOJ (LaFreniere & Newman, 2016), the current study examined EMA data of WOJ users with GAD to determine the percentage of worries that did not come true. This percentage of untrue worries was then used to predict GAD symptoms post-trial and the change in GAD symptoms across the trial. Results revealed that, on average, 91.4% of GAD worries did not come true. As expected, higher percentages of untrue worries significantly predicted both lower GAD symptoms at outcome and greater reduction in GAD symptoms across the trial. According to these findings, the worries of those with GAD do appear to be unlikely. Moreover, greater disconfirmation of these worries with empirical evidence predicted better treatment outcomes. We will discuss each of these results in turn.

To begin, this study's descriptive findings inform several important GAD phenomena. Notably, it appears quite common that worries did not come true for WOJ users with GAD (91.4%). In fact, the most common incidence of false worries among participants was 100% with a median of 93.5%. Clinicians may find these figures valuable in therapy during psychoeducation and cognitive restructuring of worry's validity. Note also that it was participants themselves who determined whether their worries came true or not. Positive beliefs and the theorized functions of worry in GAD (Behar, DiMarco, Hekler, Mohlman, & Staples, 2009) suggest people with GAD view worry as a valued means of coping. Since those with GAD may be motivated to maintain their worry by supporting its accuracy, the true incidence of worry non-outcomes may actually be higher than shown in participant journals. For those worries that did come true, 30% were perceived as having turned out better than expected. These findings partially replicate the informal report of Borkovec et al. (1999), who found a similar nonoutcome figure of 85% when reviewing homework for 17 clients in treatment. Even so, our 91% figure is higher than their other, non-treatment undergraduate GAD sample. In this sample, 67.9% of worries were rated as not having come true in a questionnaire. Yet Borkovec et al.'s studies were not peer-reviewed, they did not report the specifics of their method, they did not prompt for worry recording as in the WOJ, and they did not use coders or a coding system. Their report does not explain how percentages were calculated, how outcomes were reviewed and tracked, whether participants

were truly recording testable worrisome predictions by a standardized definition, and how repeated worries, repeated outcomes, and untestable worries were handled. It is also unclear whether their participants were reviewing all their past worries regularly to note outcomes (as in the WOJ) or only did so for each day's worries or at two weeks' time. It is likely that without prompting and regular review, those with GAD would be selectively biased to recall and attend more to worries that turned out poorly. Regardless, as elaborated by Borkovec et al., the high non-occurrence of feared outcomes in their client sample and our sample suggests frequent opportunity for negative reinforcement to promote worry. Interestingly, that which treats worry when recognized may maintain worry when unattended. Such a process substantiates the underlying premises of self-monitoring.

In addition to the percentage of untrue worries, other descriptive data from WOJ entries were of interest. The average number of unique worries per person was rather high: We found an average of 34.30 distinct worries across 10 days not counting repeated worries, with the highest person's count being over 100. We also report various EMA ratings of worry experiences from very early in treatment (the first two days). The average distress rating per worry very early in treatment was 4.51 (moderate to high). This finding is consistent with the Contrast Avoidance Model of worry, which proposes individual worries are associated with elevated distress (Newman & Llera, 2011). Worries also appeared to take up significant amounts of thinking time for participants (cognitive interference), with individual worries consuming 43.12% of the past two hours and 25.88% of the entire day on average. Lastly, even without practice or instruction, it appears those with GAD are able to recognize the difference between rationally-formed probabilities of future negative outcomes and their automatic, emotional predictions. Whereas emotional probabilities for worry were 62.09% on average, participants' attempts to make logically-formed probabilities were 41.67% on average, even in the first two days of monitoring. Compared to the actual average likelihood of worry outcome across worries (8.6%), it is clear GAD participants have distorted expectations both when driven by emotion and when attempting rational forecasting. We suggest these ratings likely approximate worry experience prior to notable treatment effects due to 1) a lack of change in ratings between day one and day two, as demonstrated by our descriptive analyses, and 2) the low chance of worry predictions being tested (and disconfirmed) within this short time frame.

This study was the first to test whether recognizing greater amounts of these false expectations actually predicted better outcome for a CT intervention. Higher percentages of untrue worries were associated with both outcome level and slope of change in GAD symptoms across WOJ treatment. We recognize that this study does not provide direct evidence that the accuracy of participants' worry changed over time, even though greater worry disconfirmation percentage did predict a change in overall worry severity and frequency scores. However, we can reasonably infer that participants' prediction formation changed from the finding that worry symptoms were reduced: Worry itself is the formation of "apprehensive expectations." Our EMA design allowed for the examination of individual participant expectations (worrisome predictions) and whether they were perceived as actualized or not in real life (event outcomes). By including disconfirmed worries, analyses were able to support CT's theorized belief-correction process for at least one intervention. Although beliefs considered unrealistic by clinicians have often been assessed by

questionnaire in treatment studies (e.g., Epstein & Eidelson, 1981) and the gap between expectations for therapy and actual therapy experience has been researched (e.g., Westra, Aviram, Barnes, & Angus, 2010), no study has empirically derived the incidence of false expectations in daily life for a disordered sample. The assumptions and techniques of behavioral experiments and cognitive restructuring have been extensively applied. Yet if we want our treatments to be empirically supported *throughout,* our studies of outcome must be complemented by tests of such core processes (Kazdin, 2008). Our results contribute to this end by supporting the disconfirmation strategies often used in CBT. It is recommended that CT's disconfirmation processes be examined with further analyses, especially those meeting full criteria for establishing a treatment mechanism.

Unfortunately, there is limited support for *any* proposed mechanism of GAD treatment. Although there are a few exceptions testing GAD treatment mediators generally (e.g., Arch, Wolitzky-Taylor, Eifert, & Craske, 2012; Newman & Fisher, 2013; Newman & Fisher, 2010), none tested the disconfirmation premise of CT. One systematic review by Smits et al. (2012) did examine threat reappraisal as a mechanism of CBT treatment for anxiety, but they noted there were no mediation analyses of the construct for GAD. Although they found reappraisal to be a common treatment mechanism for anxiety across studies, these studies were almost exclusively focused on social anxiety disorder and panic disorder. Our own review of the literature did not reveal any such studies in GAD since the time of the Smits et al. (2012) review. Yet the few studies that *have* examined mechanisms may inform the processes of the WOJ. Aligned with the Smits et al. (2012) review, recognition of more worry non-outcomes may lead to beneficial reappraisal. The WOJ likely demonstrates to participants that their worries are less threatening than previously thought, because those worries seldom come true. This reappraisal of threat may reduce symptoms the way it has for several other anxiety disorders in CBT trials. In addition, a study by Arch et al. (2012) found that reductions in cognitive fusion—or "buying into" maladaptive thoughts—mediated both CBT and acceptance and commitment therapy (ACT) for GAD. Disconfirmation of worries may reduce the "believability" or cognitive fusion of worries, leading to symptom improvement. This would allow for a process that is more generalized across all worrying, rather than tediously countering singular predictions one-by-one ("I'm surprised these few worries were so inaccurate. Maybe I should be questioning all my worries"). Whatever the WOJ's sub-processes may be, more extensive research on predictors, mediators, and treatment mechanisms is certainly advisable for all GAD therapies.

Further research on the WOJ should attend to several other questions as well. We recommend other process predictors of the WOJ be addressed, as well as possible mediators. First, the rationale for the WOJ suggests that recognition of in-the-moment costs of worry promote change alongside recognition of worry likelihood. Accordingly, studies should examine how attention toward worry costs like distress, cognitive interference, and relationship strain influence outcomes. Second, the fact that greater untrue worry percentages predicted greater improvement may be considered a "double-edged sword": If certain participants find that more of their worries actually *do* come true, treatment may be less powerful or even harmful. If such worry confirmations were to occur, other CT techniques may need to be applied. Possible options are reframing outcomes or directing

attention toward client coping ability, nonoutcomes, or the ultimate harmlessness of the situation's consequences. Research should then inquire into the effect of augmenting the WOJ with other CBT techniques, which can be applied in the unlikely event that a worry comes true. Third, our participants were not given therapists with whom they could process their WOJ data, outcomes, and maladaptive beliefs. Studying the WOJ in the context of in-person therapy is equally important to studying it as a stand-alone EMI. Other studies may also benefit from addressing WOJ use in other populations and disorders, other WOJ delivery methods, addition of the WOJ EMI to ongoing face-to-face therapies, and comparisons to other intervention processes, perhaps with moderated mediation analyses.

In designing future extensions of this study, it is important to recognize its limitations. The most salient is our relatively small sample size of 29, although analyses were sufficiently powered. Our population had restricted generalizability as well, being largely White, female, young adult undergraduate students. Participants were also diagnosed by the criterion-scoring of a self-report questionnaire, rather than a clinical interview. Furthermore, we were not able to examine change in percentage of untrue worries. Participants did not mark the time or day that they determined a worry was true or untrue, our coding system did not partition worries into their time of first occurrence or day of the study, and the occurrence of repeated worries across time made this impossible. Yet change in percentage or number of untrue worries may have functioned as a superior predictor of outcome. As mentioned in the original RCT, we were also unable to examine worry outcomes for worries that needed longer periods to be tested, and did not track worries past a month's time. Only tracking proximal worry events may have influenced certain findings, such as the average number of unique worries per person. Yet it is important to recognize that even monitoring outcomes of only proximal worries was able to reduce general worrying and GAD symptoms. In general, it would be beneficial to test the WOJ for a longer duration and examine the journal data therein.

Despite its limitations, the results of this study suggest several clinical implications for GAD treatment. Primarily, therapists should use techniques to draw client attention to evidence that their worries are unrealistic, unlikely, and unhelpful. If clients can track the actual results of their expectations, they may learn to form more adaptive, evidence-based beliefs and predictions. Evidence that worries are unlikely may also increase client expectancy for CBT success. As clients realize their long-held beliefs may be flawed, expectancy for a belief-challenging therapy should increase. Therapists should capitalize on this concordance between the verification of therapy's predictions and the refutation of anxiety's predictions. Increasing client faith in the therapeutic model may then increase treatment's effect. In showing support for disconfirmation, our study suggests CT should continue to aim for correcting maladaptive cognitions with renewed confidence. In cases where a client's worries are often shown to be (or perceived to be) true, alternatives to CT may be chosen. Relaxation training, detachment from one's thoughts (i.e., cognitive diffusion in ACT), or—similarly—present moment awareness in mindfulness are all viable options. Since lower non-outcome percentages did predict less symptom change, worry-consistent events should be addressed—at least if they seduce clients' focus. Therapists may target the rare occurrences that worry did come true, engaging in reframing or identification of errors in outcome interpretations. When worries are supposedly confirmed, therapists may discuss

with clients whether their interpretations of outcome were accurate and whether the consequences were truly so dire. For example, is an 85% exam score *really* confirming the worry, "I will do poorly on the exam"? And are its repercussions *really* so bad? The results of this study support the use of such commonly-used but understudied techniques. In the spirit of establishing an empirical basis for informed, optimal healing, we hope such evaluations of both old and new means of help continue long into the future.

References

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, D.C.: American Psychiatric Association.

American Psychiatric Association, A. P. A. (2013). Diagnostic and Statistical Manual of Mental Disorders (DSM-5; 5th ed.). Arlington, Virginia: American Psychiatric Association. doi: 10.1176/appi.books.9780890425596

Andersson G, Paxling B, Roch-Norlund P, Ostman G, Norgren A, Almlov J, Georen L, Breitholtz E, Dahlin M, Cuijpers P, Carlbring P, & Silverberg F (2012). Internet-based psychodynamic versus cognitive behavioral guided self-help for generalized anxiety disorder: A randomized controlled trial. Psychotherapy and Psychosomatics, 81, 344–355. doi:10.1159/000339371 [PubMed: 22964540]

Arch JJ, Wolitzky-Taylor KB, Eifert GH, & Craske MG (2012). Longitudinal treatment mediation of traditional cognitive behavioral therapy and acceptance and commitment therapy for anxiety disorders. Behavior Research and Therapy, 50, 469–478. doi:10.1016/j.brat.2012.04.007

Arch JJ, & Craske MG (2008). Acceptance and commitment therapy and cognitive behavioral therapy for anxiety disorders: Different treatments, similar mechanisms? Clinical Psychology: Science and Practice, 15, 263–279. doi: 10.1111/j.1468-2850.2008.00137.x

Beck AT (1979). Cognitive therapy and the emotional disorders. New York: Penguin.

Beck AT, & Emery G (1985). Anxiety disorders and phobias: A cognitive perspective. New York, NY: Basic Books.

Behar E, DiMarco ID, Hekler EB, Mohlman J, & Staples AM (2009). Current theoretical models of generalized anxiety disorder (GAD): Conceptual review and treatment implications. Journal of Anxiety Disorders, 23, 1011–1023. doi:10.1016/j.janxdis.2009.07.006 [PubMed: 19700258]

Berenbaum H, Thompson RJ, & Bredemeier K (2007). Perceived threat: Exploring its association with worry and its hypothesized antecedents. Behaviour Research and Therapy, 45, 2473–2482. doi:10.1016/j.brat.2007.03.015 [PubMed: 17512494]

Berenbaum H, Thompson RJ, & Pomerantz EM (2007). The relation between worrying and concerns: The importance of perceived probability and cost. Behaviour Research and Therapy, 45, 301–311. doi:10.1016/j.brat.2006.03.009 [PubMed: 16677599]

Borkovec TD, Hazlett-Stevens H, & Diaz ML (1999). The role of positive beliefs about worry in generalized anxiety disorder and its treatment. Clinical Psychology and Psychotherapy, 6, 126–138.

Borkovec TD, & Newman MG (1998). Worry and generalized anxiety disorder In Salkovskis P (Ed.), Adults: Clinical formulation and treatment (Vol. 6, pp. 439–459). Oxford, England: Pergamon Press. doi: 10.1016/B0080-4270(73)00201-7

Butler AC, Chapman JE, Forman EM, & Beck AT (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. Clinical Psychology Review, 26, 17–31. doi:10.1016/j.cpr.2005.07.003 [PubMed: 16199119]

Butler G, & Mathews A (1983). Cognitive processes in anxiety. Advances in Behaviour Research and Therapy, 5, 51–62. doi:10.1016/0146-6402(83)90015-2

Chambless DL, & Ollendick TH (2001). Empirically supported psychological, interventions: Controversies and evidence. Annual Review of Psychology, 52, 685–716. doi:10.1146/Annurev.Psych.52.1.685

Chen J, Liu X, Rapee RM, & Pillay P (2013). Behavioural activation: A pilot trial of transdiagnostic treatment for excessive worry. Behaviour Research and Therapy, 51, 533–539. doi:10.1016/j.brat.2013.05.010 [PubMed: 23827657]

Cohen J, & Cohen P (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Epstein N, & Eidelson RJ (1981). Unrealistic beliefs of clinical couples: Their relationship to expectations, goals and satisfaction. The American Journal of Family Therapy, 9, 13–22. doi:10.1080/01926188108250420

Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, & Fang A (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. Cognitive Therapy and Research, 36, 427–440. doi:10.1007/s10608-012-9476-1 [PubMed: 23459093]

Kazdin AE (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. American Psychologist, 63, 146–159. doi:10.1037/0003-066X.63.3.146 [PubMed: 18377105]

LaFreniere LS, & Newman MG (2016). A brief ecological momentary intervention for generalized anxiety disorder: A randomized controlled trial of the Worry Outcome Journal. Depression & Anxiety, 33, 829–839. doi:10.1002/da.22507 [PubMed: 27062682]

MacLeod AK, Williams JM, & Bekerian DA (1991). Worry is reasonable: The role of explanations in pessimism about future personal events. Journal of Abnormal Psychology, 100, 478–486. doi:10.1037/0021-843x.100.4.478 [PubMed: 1757661]

Moore MT, Anderson NL, Barnes JM, Haigh EA, & Fresco DM (2014). Using the GAD-Q-IV to identify generalized anxiety disorder in psychiatric treatment seeking and primary care medical samples. Journal of Anxiety Disorders, 28, 25–30. doi:10.1016/j.janxdis.2013.10.009 [PubMed: 24334213]

Newman MG, & Fisher AJ (2013). Mediated moderation in combined cognitive behavioral therapy versus component treatments for generalized anxiety disorder. Journal of Consulting and Clinical Psychology, 81, 405–414. doi:10.1037/a0031690 [PubMed: 23398493]

Newman MG, Zuellig AR, Kachin KE, Constantino MJ, Przeworski A, Erickson T, & Cashman-McGrath L (2002). Preliminary reliability and validity of the Generalized Anxiety Disorder Questionnaire-IV: A revised self-report diagnostic measure of generalized anxiety disorder. Behavior Therapy, 33, 215–233. doi:10.1016/S0005-7894(02)80026-0

Newman MG, & Llera SJ (2011). A novel theory of experiential avoidance in generalized anxiety disorder: A review and synthesis of research supporting a contrast avoidance model of worry. Clinical Psychology Review, 31, 371–382. doi:10.1016/j.cpr.2011.01.008 [PubMed: 21334285]

Newman MG, & Fisher AJ (2010). Expectancy/credibility change as a mediator of cognitive behavioral therapy for generalized anxiety disorder: Mechanism of action or proxy for symptom change? International Journal of Cognitive Therapy, 3, 245–261. doi :10.1521/ijct.2010.3.3.245 [PubMed: 21132075]

Paxling B, Almlöv J, Dahlin M, Carlbring P, Breitholtz E, Eriksson T, & Andersson G (2011). Guided Internet-delivered cognitive behavior therapy for generalized anxiety disorder: A randomized controlled trial. Cognitive Behaviour Therapy, 40, 159–173. doi:10.1080/16506073.2011.576699 [PubMed: 21770848]

Preacher KJ, & Hayes AF (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior Research Methods, 40, 879–891. doi:10.3758/BRM.40.3.879 [PubMed: 18697684]

Rodebaugh TL, Holaway RM, & Heimberg RG (2008). The factor structure and dimensional scoring of the Generalized Anxiety Disorder Questionnaire for DSM-IV. Assessment, 15, 343–350. doi:10.1177/1073191107312547 [PubMed: 18202302]

Smits JA, Julian K, Rosenfield D, & Powers MB (2012). Threat reappraisal as a mediator of symptom change in cognitive-behavioral treatment of anxiety disorders: A systematic review. Journal of Consulting and Clinical Psychology, 80, 624–635. doi:10.1037/a0028957 [PubMed: 22686124]

Westra HA, Aviram A, Barnes M, & Angus L (2010). Therapy was not what I expected: A preliminary qualitative analysis of concordance between client expectations and experience of cognitive-

behavioural therapy. Psychotherapy Research, 20, 436–446. doi:10.1080/10503301003657395 [PubMed: 20552534]

Table 1

Worry outcome incidence and WOJ ratings across entries of the first two days of monitoring

	M (SD)	Min. – Max.
Overall percent of worries that did not come true	91.30%	--
Percent of true worries[a] with outcome better than expected	30.10%	--
Percent of true worries[a] with outcome as bad as expected	44.15%	--
Percent of true worries[a] with outcome worse than expected	25.80%	--
Average number of unique worries per person	34.30 (18.00)	12.50 - 100.50
Early worry-related distress[b]	4.51 (1.49)	3.13 – 6.75
Early recent cognitive interference duration[c]	43.12% (26.08)	18.75% - 75.14%
Early daily cognitive interference duration[d]	25.88% (22.57)	1.25% - 67.86%
Early emotional likelihood of worry outcome	62.09% (26.08)	25.00% - 85.71%
Early logical likelihood of worry outcome	41.67% (22.97)	7.50% - 82.86%

[a]Percentage of worries that were perceived as having come true by participants during monitoring.

[b]Average degree of self-reported distress associated with individual worries on a scale from 1 (*no distress*) to 7 (*severe distress*).

[c]Percentage of time spent on thought in past two hours.

[d]Percentage of time spent on thought across entire day.

Table 2

Zero-order correlations between model variables (using pairwise deletion).

	1	2	3	4
1. % Untrue Worries				
2. GAD-Q Pre-to-Post Slope	−.353†			
3. GAD-Q Pre-Trial	−.270	.150		
4. GAD-Q Post-Trial	−.400†	.950**	.452*	

Note: GAD-Q = Generalized Anxiety Disorder Questionnaire for DSM-IV-TR;

† = $p < .10$.

* = $p < .05$.

** = $p < .001$.