Published in final edited form as:

J Mood Anxiety Disord. 2024 September ; 7: . doi:10.1016/j.xjmad.2024.100060.

Assessing in-session rumination during CBT for depression: Replication and further evaluation of an observational measure

J.C. Kennedy[a,*], **B.W. Dunlop**[a], **S.E. Baldwin**[a], **C.J. Liu**[b], **H.S. Mayberg**[c,1], **W.E. Craighead**[a,d,1]

[a]Department of Psychiatry and Behavioral Sciences, School of Medicine, Emory University School of Medicine, Atlanta, GA, United States

[b]Department of Psychology, American University, Washington, D.C., United States

[c]The Nash Family Center for Advanced Circuit Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[d]Department of Psychology, Emory University, Atlanta, GA, United States

Abstract

This study sought to further evaluate an observational measure of rumination that occurs during psychotherapy (i.e., in-session rumination). Specifically, the study aimed to replicate the reliability of an observational rating procedure in new therapy sessions and a new sample, clarify the relationship between in-session rumination and depressive symptoms, and evaluate for the first time the relationship between in-session rumination and self-reported rumination. A team of trained research assistants produced observational ratings of in-session rumination occurring during video-taped sessions of CBT from two separate treatment studies. Thirty-five patients with major depressive disorder (MDD) from one study had their final session rated, and 17 patients with MDD from another study had their first session rated. Results showed that the observational ratings were reliable, and that in-session rumination generally correlated with depressive symptom severity as expected, with higher in-session rumination predicting higher depressive symptom levels both cross-sectionally and longitudinally. Preliminary evidence also found that in-session rumination correlated with self-reported rumination, and exploratory analyses provided preliminary evidence supporting the incremental validity of in-session rumination for predicting depression severity after treatment. The results indicate that in-session rumination can be reliably identified during CBT sessions and consistently predicts higher depressive severity, both of which support efforts to develop treatments that specifically target rumination.

Keywords

In-session rumination; Rumination during CBT; Rumination; Psychotherapy observational ratings

[*]Correspondence to: 12 Executive Park, Suite 247, Emory University School of Medicine, Atlanta, GA 30329, United States. jamie.kennedy@emory.edu (J.C. Kennedy).

[1]Indicates equal senior authorship.

1. Introduction

Rumination is a construct grounded in basic science that carries important implications for the treatment of depression. Within Nolen-Hoeksema's Response Styles Theory, which emphasizes the particularly depressogenic form of rumination (often called depressive rumination), rumination refers to a maladaptive way of responding to one's distress, including "repetitively focusing on the fact that one is depressed; on one's symptoms of depression; and on the causes, meanings, and consequences of depressive symptoms." [1] (p. 569) Rumination begins as an attempt to explain one's current mood or distress (e.g., "why is this happening to me?" or "what does this mean?") and frequently expands to a repetitive, somewhat uncontrollable chain of thoughts about other problems, self-blame, criticism, or emotions [2,3]. Although it shares characteristics, like negative valence, with other depressogenic cognitive variables (e.g., negative automatic thoughts [4]), rumination is characterized by its typical content (e.g., causes and consequences of depression), and by its repetitive, passive, abstract, evaluative, and overgeneralized characteristics [5,6].

Considerable evidence implicates rumination in the etiology and maintenance of depression. Several meta-analyses show that self-reported rumination and depressive symptoms are consistently associated in both clinical and non-clinical samples [7,8]. Moreover, longitudinal studies show that, after controlling for baseline depressive symptoms, self-reported rumination predicts the onset and recurrence of depressive episodes [9–12]. Experimental research has also found that rumination produces numerous detrimental effects, such as exacer-bating negative mood [13], and clinical neuroscience research links rumination to abnormalities associated with the pathophysiology of depression [14].

Rumination can also interfere with evidence-based psychotherapy, such as cognitive behavior therapy (CBT). For example, one study found that higher pre-treatment self-reported rumination predicted lower frequency of remission to CBT [15], and another study found that greater baseline self-reported rumination predicted higher levels of depressive symptoms at the end of treatment [16]. Given these findings, some researchers concluded that some depressed patients may need interventions that specifically target repetitive thought processes like rumination [17]. Patients who habitually ruminate, for instance, may do so automatically without awareness, and might benefit from targeted, behaviorally oriented interventions aimed at creating habit change and emphasizing the functional antecedents and consequences of ruminative thinking rather than its accuracy [18,19].

In our prior work [20], we sought to add to the rumination literature by developing, implementing, and evaluating a measure of rumination that occurs during psychotherapy for depression. We labeled this clinical process as "in-session rumination." We chose not to use the term "rumination" alone because it has primarily been conceptualized and measured as an *intrapersonal*, cognitive process, whereas our measure assessed in-session behavior, an *interpersonal* process. Having an observational measure of ruminative behavior supplements the existing literature that typically relies on self-reported rumination. Although self-report has many benefits, including efficiently assessing patients' subjective experiences, it can also be affected by the cognitive biases associated with psychopathology [21,22]. Observational measures, on the other hand, are shielded from these biases by

assessing overt behavior. Further, observational measures permit novel research questions, such as evaluating how patients' in-session behaviors relate to treatment processes and outcomes. Having a reliable, validated measure of in-session rumination may also ultimately help practicing clinicians identify this important patient behavior.

Our initial study provided a preliminary evaluation of a measure of in-session rumination occurring during psychotherapy for depression [20]. Specifically, we rated depressed patients' video-recorded first and eighth CBT sessions and found that both sessions could be rated reliably (*ICC*s > 0.69). Moreover, the study found that in-session rumination correlated with depressive symptoms as expected, with higher in-session rumination associated with higher depressive symptoms during treatment. However, there were some limitations to the study. First, the study was limited to one sample and two therapy sessions, so we were unable to determine how our methodology might generalize to other therapy sessions and to other research samples. Additionally, there were some inconsistencies in how our measure correlated with depressive symptoms, with some effects only detected with clinician-rated depressive symptoms and other effects only detected with self-reported depression symptoms. Finally, the study did not include a measure of self-reported rumination, which prevented us from evaluating how in-session rumination relates to self-reported rumination.

1.1. The present study

The present study aimed to address the limitations of our prior work. Specifically, the study sought to further investigate the generalizability of an observational measure of in-session rumination by replicating its reliability in both new therapy sessions and in a new clinical-trial research sample. The study also sought to clarify the relationship of in-session rumination to depressive symptoms by evaluating how it relates to clinician-rated and self-reported depressive symptoms both cross-sectionally and longitudinally. Finally, the study allowed for the first evaluation of the relationship between in-session rumination and self-reported rumination. Understanding how in-session rumination relates to self-reported rumination allows investigators to better contextualize our prior findings and the results of the current study within the broader rumination literature.

2. Materials and methods

The current investigation reports on a secondary project of the Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) and the InSuLa Assessed Needs for Depression (ISLAND) studies. Prior publications described the protocols and outcomes of PReDICT [23–26] and ISLAND [27]. PReDICT participants in the current study were patients participating in 12-week CBT monotherapy. ISLAND participants included in the current study were participants who did not remit to their initial 12-week open-label treatment with a selective serotonin reuptake inhibitor (SSRI) and were beginning 12 weeks of CBT while maintaining the SSRI at the same dose they were on at week 12. All participants provided written informed consent. The Emory Institutional Review Board approved both studies, and they were conducted in accordance with the 1975 Helsinki Declaration and its amendments. PReDICT data were gathered from 2007 to 2013, ISLAND

data were gathered from 2014 to 2019, and the observational ratings data for the current study were produced in 2022–2023.

2.1. Participants

PReDICT participants eligible for the present study were patients participating in 12 weeks of CBT monotherapy who had a video-recording of their initial CBT session rated in a prior study [20] and also completed a neuroimaging scan during the study; these scans are being analyzed in another project. Of the 42 eligible PReDICT patients, 35 (83.3 %) had a video-recording of their final CBT session to be rated. ISLAND participants eligible for the present study were patients who completed 12 weeks of SSRI monotherapy treatment and did not remit to treatment (n = 24). Of the eligible patients, 17 (70.8 %) agreed to participate in 12 weeks of CBT and had a session video-recording available for rating. As part of the ISLAND and PReDICT studies, all participants completed the Structured Clinical Interview for DSM-IV [28], were rated on the 17-item Hamilton Depression Rating Scale (HAM-D) [29], and completed an independent interview by a board-certified psychiatrist. Participants had at least moderately severe depression, defined as a HAM-D score ≥ 18 at screening and ≥ 15 at their study baseline visit. The individual study publications provide additional details on inclusion and exclusion criteria [23,27].

2.2. Materials

Patients were rated on their in-session rumination *intensity,* which is based on a four-point scale (0 = *no in-session rumination* to 3 = *severe in-session rumination*). More intense ratings were signified by the patient repeatedly engaging in the following behaviors: returning to a ruminative topic, having trouble disengaging from one, providing excessive detail about one, and/or responding in ways that seem increasingly abstract and overgeneralized. Ratings were produced for individual 10-minute segments of the session, and the maximum rating across all the segments was used as the datum for that session. For each of the observed therapy segments, raters also evaluated the *duration* of in-session rumination using a five-point scale (0 = *not at all* to 4 = *very much*). Anchors were provided for each rating point (1, ruminating up to 25 % of the time; 2, 25–50 %; 3, 50 – 75 %; and 4, 75 – 100 %). Ratings were made for each individual segment of therapy, and the average rating across the four segments was used as the datum for that session.

As part of the ISLAND study, patients also completed the Ruminative Response Scale (RRS), which is a subscale of the Response Styles Questionnaire [30]. The RRS assesses one's self-reported tendency to engage in ruminative behaviors during periods of distress (e.g., "think about how sad you feel" and "think—why do I always react this way?"). Both the ISLAND and PReDICT studies collected clinician-rated and self-reported measures of depressive symptoms. The 17-item Hamilton Depression Rating Scale (HAM-D) [29,31] is a clinician-rated measure of depressive symptom severity, and the Beck Depression Inventory (BDI) [32] is a widely-used, self-report measure of depressive symptoms. Total HAM-D and BDI scores were used to represent depression severity in the current study.

2.3. Procedure

The observational ratings produced for the current study used a methodology developed to measure in-session rumination in video-recordings of patients' first and eighth CBT sessions in the PReDICT study [20]. In short, the rating scheme operationalized in-session rumination as negative, repetitive, and passive talking about typical ruminative topics, namely: *dwelling* on one's symptoms or feelings, *rehashing* the details of a problem or situation, or *speculating* about the causes and consequences of a problem, situation, or symptom [1,33]. All observational raters were postbaccalaureate research assistants who were unaware of patient outcomes or the hypotheses of the current study. Training included receiving didactic instructions and making practice ratings on video-recordings of CBT sessions from a different treatment study before producing ratings for the current study. Weekly rating meetings were conducted to maintain alignment on the procedure and to discuss ambiguous or difficult-to-rate sessions; final scores for each rating were locked before these discussions. All sessions were double rated by a team of research assistants, and the order in which the sessions were rated was randomized, meaning that participants were not rated in the order in which they entered the study.

CBT was provided in a manner consistent with the Beck and colleagues protocol [4]. Participants met with their therapist twice per week for the first four weeks and then weekly for the remaining eight weeks. We first rated PReDICT patients' final CBT session, as we sought to further evaluate the reliability of our rating scheme in a new session context and evaluate how the ratings correlated with end of treatment depressive symptoms. We then rated ISLAND patients' first CBT session, which were structured around a relatively open-ended collection of patient history involving discussing past events that might trigger rumination, and thus, create variability in observed in-session rumination. Moreover, we sought to replicate our prior findings that evaluated if early-treatment in-session rumination predicted symptoms later in treatment [20].

2.4. Data analysis

Preliminary analyses yielded the demographic and clinical characteristics of the study sample. Chi-Square tests and Analysis of Variance (ANOVA) were used to compare eligible participants who did not have a rated recorded session to those who did have one. Descriptive statistics were also calculated to summarize the incidence and distribution of in-session rumination. To assess the reliability of the observational measures, intraclass correlation coefficients (*ICC*) [34] with absolute agreement produced an estimate of the ratio of true score variance to total variance. These correlations provided a reliability estimate of the mean scores of the rating teams considered as a whole, allowing for generalizability of the results to other samples. Pearson correlations were used to evaluate the relationships among the measures of in-session rumination, self-reported rumination, and depressive symptoms, and multiple regressions were used to further evaluate the relationships among the variables. All analyses were conducted at a statistical significance of $p < .05$ (2-tailed). SPSS 28.0 was used for the analyses.

3. Results

3.1. Sample characteristics

Table 1 describes the demographic and clinical characteristics of the participants. Both ISLAND and PReDICT participants included in the current study did not significantly differ on any of the clinical or demographic variables from the eligible patients who did not have a video recording to be rated.

3.2. Description of observational ratings and inter-rater reliability

Thirty-five final CBT sessions from the PReDICT study were rated for in-session rumination. The ratings included all levels of in-session rumination *intensity* (31.4 %, None; 54.3 %, Mild; 8.6 %, Moderate; 5.7 %, Severe), and the *duration* ratings ranged from no in-session rumination (0) to over half of the session time (2.13); (M = 0.63; SD = 0.59). *ICCs* were estimated for the averaged ratings using two-way random effects models with absolute agreement (model (2,2)) [34], and they were evaluated relative to established guidelines [35]. Reliability for both in-session rumination *intensity* and *duration* were in the "excellent" range (ICCs >.75); specifically, the ICCs were: .841 for *intensity* and .872 for *duration*. These data extend our prior findings regarding the reliability of in-session rumination for the PReDICT study data by demonstrating reliable ratings for an additional session at the end of CBT treatment.

Seventeen CBT sessions from the ISLAND study were rated for in-session rumination. Most of the *intensity* ratings were in the "Mild" range, though some of the ratings were distributed across the other intensities except for "Severe" (17.6 %, None; 64.7 %, Mild; 17.6 %, Moderate; 0 %, Severe). The *duration* ratings ranged from no in-session rumination (0) to over a quarter of the session time (1.63); (M =0.61; SD = 0.43). Reliability for in-session rumination *intensity* was in the "good" range, .697, and reliability for *duration* was in the "excellent" range, .918. The reliability for *intensity* was lower than we had hoped, and it may reflect limited variability in the *intensity* ratings of the sample (i.e., the *intensity* ratings in this sample did not include a patient with a "severe" rating). Since these patients participated it SSRI treatment prior to starting CBT, we speculate that this prior treatment may have reduced their symptoms or trained them to participate in a clinical setting in such a way that limited them from displaying "severe" in-session rumination *intensity*. However, the reliability was still in the "good" range and together with the "excellent" ratings for *duration* indicate that our rating procedure can reliably be applied to a new sample.

3.3. Evaluating the relationship of in-session rumination and self-reported rumination to depressive symptom severity

The ratings from the PReDICT sample provided consistent evidence of a positive association between measures of in-session rumination and depressive symptom severity. In-session rumination *intensity* observed during patients' final CBT session had a positive relationship with clinician-rated HAM-D total scores at the end of treatment, $r(35) = .642$, $p < .001$, as did the *duration* ratings, $r(35) = .611$, $p < .001$. Similarly, the final CBT session ratings also had a positive association with patient self-reported depressive symptom severity on the BDI at the end of treatment (*intensity*, $r(35) = .515$, $p = .002$; *duration*, $r(35) = .508$, $p = .002$).

Table 2 displays the zero-order correlations in the ISLAND sample among the measures, including in-session rumination and depressive symptom severity on the HAM-D and BDI. The evaluated correlations also provided consistent evidence for the expected association between in-session rumination and depressive symptom severity. When evaluated cross-sectionally at the beginning of CBT, both measures of in-session rumination had a strong, positive relationship with both clinician-rated HAM-D symptom severity (rs > .532) and self-reported depressive symptom severity on the BDI (rs > .568). Similarly, self-reported rumination on the RRS had a medium to strong, positive relationship with both clinician-rated HAM-D symptom severity, $r(17) = .489$, $p = .046$, and self-reported BDI symptom severity, $r(17) = .706$, $p = .002$, at the beginning of CBT.

The results also supported the value of in-session rumination at the beginning of CBT predicting depressive symptom severity at the end of the 12 weeks of CBT in the ISLAND sample. Specifically, both measures of in-session rumination at the first CBT session had a strong, positive association with clinician-rated HAM-D symptom severity at the end of CBT (rs > .498). In-session rumination *intensity* had a positive relationship with self-reported BDI symptom severity at the end of CBT treatment that fell short of statistical significance, $r(17) = .457$, $p = .065$, and the *duration* ratings had a statistically significant, positive, and strong association with BDI symptom severity at the end of CBT, $r(17) = .579$, $p = .015$. In contrast, self-reported rumination on the RRS did not predict end of CBT HAM-D symptom severity, $r(17) = .185$, $p = .478$, yet did predict end of CBT BDI scores, $r(17) = .556$, $p = .020$.

3.4. Evaluating the relationship of in-session rumination to self-reported rumination

An important step in evaluating a new clinical measure is to investigate its relationship with more established measures of similar constructs or processes (i.e., convergent validity). The ISLAND sample provided the first evaluation of the relationship between in-session rumination and self-reported rumination on the RRS, and the results generally revealed a medium to large positive association between the two measures (Table 2). In-session rumination *intensity* was moderately correlated with self-reported rumination at the beginning of CBT, though, this association did not reach statistical significance, $r(17) = .395$, $p = .117$. This effect may be attenuated due to the lower variability and reliability of the *intensity* ratings. In-session rumination *duration* was strongly correlated with the RRS, $r(17) = .523$, $p = .031$. Moreover, both measures of in-session rumination had a strong, positive relationships with self-reported rumination scores at the end of CBT (rs > .591).

3.5. Exploring the incremental validity of in-session rumination for predicting end of treatment depressive severity

Another important step in evaluating a new clinical measure is to investigate its predictive validity above and beyond existing, well-established predictors [36]. Thus, we conducted analyses to evaluate the incremental validity of in-session rumination in predicting subsequent depressive symptoms after controlling for baseline depressive symptoms, one well-established predictor of depressive symptoms [24]. Specifically, we reanalyzed data from the PReDICT sample [20], which found that in-session rumination observed during patients' first CBT session predicted higher clinician-rated symptom severity later in

treatment. Forty-two patients in the sample had in-session rumination ratings from their first CBT session, as well as both baseline and end of treatment HAM-D scores. Hierarchical regressions were used to evaluate incremental validity, with baseline HAM-D symptom severity entered in step one, and in-session rumination, measured during CBT session one, entered in step two. The initial models found that baseline depressive symptom severity predicted depressive symptom severity at the end of treatment (HAM-D, β = .316, p = .037). When in-session rumination *intensity* was added to the initial model, neither baseline HAM-D scores (β = .272, p = .066) nor in-session rumination *intensity* (β = .273, p = .065) were statistically significant predictors individually, yet the overall model including both predictors was statistically significant and explained an additional 7.3 % variance in end of treatment symptom severity relative to the initial model. Remarkably, the effect sizes for baseline HAM-D scores and in-session rumination *intensity* during the first CBT session were nearly identical, suggesting that they equally contributed to predicting end of treatment symptoms. In-session rumination *duration* did not emerge as a statistically significant predictor of subsequent depressive symptoms were also entered in the model (β = .173, p = .244). These analyses provide some evidence that in-session rumination has incremental validity and utility in predicting subsequent depressive symptom severity when baseline depressive symptoms were also entered in the model (β = .173, p = .244). These analyses provide some evidence that in-session rumination has incremental validity and utility in predicting subsequent depressive symptom severity above and beyond what was predicted by baseline depressive symptom severity.

To further evaluate the incremental validity of in-session rumination, we compared how self-reported rumination and in-session rumination predicted subsequent depressive symptom severity. More specifically, we conducted a series of simultaneous regressions in which self-reported rumination on the RRS and measures of in-session rumination concurrently predicted end of treatment depressive symptom severity among 17 ISLAND study participants. The results are summarized in Table 3. Measures of in-session rumination reliably predicted subsequent clinician-rated HAM-D scores (βs >.503, ps, <.06), while self-reported rumination on the RRS did not (βs < 0, ps, >.555). When predicting subsequent self-reported depressive symptom scores on the BDI, the results were mixed. Self-reported rumination on the RRS appeared to be a slightly stronger predictor of subsequent BDI symptom severity than in-session rumination *intensity* (RRS, β = .445, p = .073; in-session rumination *intensity,* β = .281, p = .242). Self-reported rumination and in-session rumination *duration* appeared to have a similar effect size, though, the results were not statistically significant (RRS, β = .349, p = .165; in-session rumination *intensity,* β = .396, p = .119). These results suggest that in-session rumination generally predicts subsequent depressive symptoms and provides unique information not captured be self-reported rumination.

4. Discussion

The current study evaluated a recently developed observational measure of in-session rumination that occurs during CBT for depression. In-session rumination refers to patients negatively, repetitively, and passively talking about common ruminative topics, like *dwelling* on their symptoms or feelings, *rehashing* the details of their problems, or *speculating* about the causes and consequences of their problems [1,33]. A team of two trained

postbaccalaureate research assistants produced ratings of both the *intensity* and *duration* of in-session rumination in patients' final CBT session in the PReDICT sample and their first CBT session in the ISLAND sample.

Results supported the reliability of the observational ratings. All of the ratings were in the "good" to "excellent" range of reliability according to published standards [35]. These finding are consistent with our prior work that reported reliable ratings of patients' first and eighth CBT sessions [20]. The data are also consistent with other work that reliably observed rumination in other contexts [37,38]. Importantly, the results of the current study show that our in-session, behaviorally-based rating procedure can generalize to new therapy sessions (i.e., patients' final session) and to a new sample. Having a reliable measure of in-session rumination provides a complementary research methodology to self-report measures and allows for evaluation of how in-session behaviors relate to clinical processes and outcomes.

The current study also found consistent evidence of a strong, positive association between in-session rumination and depressive symptom severity. In the PReDICT sample, in-session rumination observed during patients' final CBT session predicted both clinician-rated and self-reported depressive symptom severity at the end of treatment. In the ISLAND sample, in-session rumination occurring in patients' first CBT session correlated with higher clinician-rated and self-reported depressive symptom severity at the beginning of CBT and predicted higher levels of depression 12 weeks later at the end of CBT. These results help clarify the relationship between in-session rumination and depressive symptoms. Our prior work found inconsistencies in how in-session rumination related to clinician-rated and self-reported depressive symptom severity, with some effects only detected for one measure of depression and other effects only detected for the other [20]. We speculated on potential explanations for the inconsistencies, including differences in the symptom measures (e.g., HAM-D emphasizes observable aspects of depression), shared method variance (i.e., HAM-D and in-session rumination are both observer-rated), and the use of clinician-rated symptoms to select participants in the study. The results of the current study, however, suggest that in-session rumination reliably correlates with both clinician-rated and self-reported symptom severity. These results are consistent with meta-analytic evidence of the relationship between rumination and depression severity [7,8], and suggest that the relationship between in-session rumination and depression symptom severity is consistent across depression assessment methodologies.

The results showing that in-session rumination predicted higher end of treatment depressive symptom severity in the ISLAND study add a unique contribution to the literature. Our prior research [20] and the PReDICT analyses in the current study found that in-session rumination predicted higher subsequent depressive symptoms in patients participating in CBT monotherapy. The analyses of the ISLAND sample in the current study add to our prior research by indicating that in-session rumination predicts poorer treatment outcomes among patients who did not remit to medication monotherapy and were participating in combined medication and psychotherapy treatment. These results are consistent with prior studies, which found that self-reported rumination predicts poorer treatment response among patients participating in combined CBT and medication treatment [39]. Together with our

prior research [20], the results of the current study indicate that in-session rumination is a clinically important process for both patients who begin treatment with CBT and patients who add CBT after failing to remit to SSRI treatment. As suggested by other researcher investigators [17], some ruminative patients may benefit from specialized psychotherapies that explicitly target repetitive thought processes like rumination.

The current study also evaluated the relationship between in-session rumination and self-reported rumination, and the results generally revealed a medium to large positive association between the measures. These results provide important evidence that in-session rumination is indeed measuring a ruminative process. Our measure of in-session rumination was developed through a thorough review of the relevant literature and extensive professional consultation, and we believe it processes good face validity in capturing some of the observable behavioral components of rumination. However, prior to the current study, it was unknown how in-session rumination related to self-reported rumination. Indeed, while other researchers have produced reliable observational ratings of ruminative processes [33,37,38], this is the first study to evaluate how an observational measure of ruminative behavior correlates with self-reported rumination. Having preliminary evidence that in-session rumination converges with self-reported rumination permits research investigators and clinicians to better contextualize our findings within the broader rumination literature.

The current study also explored the incremental validity of in-session rumination for predicting end-of-treatment depressive severity after controlling for baseline depression severity. More specifically, in the PReDICT sample, in-session rumination *intensity* observed at the beginning of CBT explained additional variance in clinician-rated depression severity 12 weeks later at the end of treatment, even after controlling for baseline depression severity. Such an effect was not detected for in-session rumination *duration.* These results provide preliminary evidence that in-session rumination demonstrates incremental validity in predicting subsequent depressive symptom severity and that its predictive utility cannot be fully attributed to its shared variance with depression [36]. Moreover, as an observable behavior, in-session rumination may ultimately prove to have practical utility, as clinicians might learn to observe and intervene on in-session rumination, something that is less possible with a multidimensional construct like depression, which comprises both overt and covert phenomena.

Finally, the study also compared the utility of self-reported rumination and in-session rumination in predicting subsequent depressive symptom severity in the ISLAND sample. The results indicated that in-session rumination reliably outperformed self-reported rumination in predicting subsequent clinician-rated depression severity. The results were mixed when predicting subsequent self-reported depressive symptom severity, with self-reported rumination outperforming in-session rumination *intensity* yet performing similarly to in-session rumination *duration.* One possible explanation for these results is shared method variance, meaning that the observational measures tended to correlate with each other, and that the self-reported measures tended to correlate with each other [40]. However, the finding that in-session rumination *duration* had a similar effect size to self-reported rumination in predicting subsequent self-reported depressive symptoms suggests that shared method variance cannot fully explain the results. Specifically, it suggests that in-session

rumination provides unique information relevant to predicting self-reported depressive symptom severity, in addition to reliably predicting clinician-rated severity. These results indicate that observational ratings of in-session rumination provide unique information relevant to clinical outcomes.

4.1. Limitations and future directions

There are several limitations to the current study. First, an important limitation of any treatment study is the risk that treatment drop-out may bias results and limit the generalizability of the findings. Indeed, our ratings of the final CBT session in the PReDICT sample did not include patients who dropped out earlier in treatment. However, our aim was to replicate our rating methodology in a new CBT session, and we also note that the results of the current study PReDICT analyses converged with our prior work [20], which rated in-session rumination in PReDICT patients' first CBT session and thus, were less vulnerable to the bias of treatment drop out. Similarly, the ratings of the ISLAND sample in the current study were made for patients' first CBT session and did not depend on their completing CBT treatment.

Another important limitation of the current study is the small sample sizes. Small sample sizes limit statistical power and reduce the precision of point estimates of statistical relationships. Thus, the effect estimates reported in the current study should be viewed as preliminary estimates, needing further evaluation in larger samples. Moreover, the sample size might have contributed to type II errors, in which we failed to detect effects that do exist. For example, our analyses of the ISLAND study sample found that self-reported rumination did not predict subsequent clinician-rated depressive symptom severity (r = .185), an effect that would likely be detected in a larger sample and that has been detected in other treatment studies [15]. We do feel confident, however, about the effects that the current study did detect, as they were replicated across two independent samples.

Another limitation of the small sample sizes is that it prohibited us from conducting more complex analyses to further evaluate the validity of in-session rumination. For example, the data indicated that all the measured variables, in-session rumination, self-reported rumination, and depressive symptom severity, were all moderately to strongly intercorrelated, making it difficult to clarify the relationships among the variables. Future work using methods suited for larger samples, like factor analysis or principal component analysis, could be used to evaluate the extent to which ruminative measures and depressive measures converge and diverge as expected. Additionally, future work could further evaluate the relative predictive power of self-reported rumination and in-session rumination to determine the extent to which they share variance or contribute uniquely to clinical outcomes.

4.2. Conclusion

In summary, this study further evaluated and validated a recently developed observational measure of in-session rumination, which refers to patients talking about common ruminative topics (i.e., *dwelling*, *speculating*, and *rehashing*) in a repetitive, negative, and passive way. Results from two samples indicated that in-session rumination could be reliably

rated and that the ratings correlated with depressive symptom severity as expected. Specifically, higher in-session rumination predicted higher depressive symptom severity both cross-sectionally and longitudinally. The study also found preliminary evidence that in-session rumination correlated with self-reported rumination, providing important evidence supporting the validity of the observational measure. Exploratory analyses also provided preliminary evidence supporting the incremental validity and utility of in-session rumination in predicting depressive symptom severity. Having a reliable observational measure of in-session rumination provides a complementary research approach to self-report measures and permits the investigation of important research questions, such as how in-session behavior relates to clinical outcomes.

Acknowledgements

The PReDICT study was supported by NIH grants P50 MH077083; RO1 MH080880; UL1RR025008; M01 RR0039; and K23 MH086690, and the ISLAND study was supported by the NIH grant R01 MH073719. Gifts from the Fuqua Family Foundations also supported the research, and a generous gift from an anonymous donor supported the in-session rumination project. All funding sources were not involved in any aspect of the design or implementation of the research. We would also like to thank the extended PReDICT Team: Treating Psychotherapists: Carla Alvarez, Ph.D., Julie Etzel, Ph.D., Rosario Falero, M.D., Ph.D., Maryrose Gerardi, Ph.D., Mary Heekin, LCSW, Meredith Jones, M.A., Noriel Lim, Ph.D., Vivianna Mahoney, Ph.D., Cynthia Ramirez, Ph.D., Sheethal Reddy, Ph.D., Lorie Ritschel, Ph.D., Jill Rosenberg, LCSW, Diana Simeonova, Ph.D., Patrick Sylvers, M.A., and Alexandra Zagoloff, Ph.D. Therapy supervisors: WEC and Linda Wilcoxon Craighead, Ph.D. Treating Physicians: BWD, Nicole Almeida, M.D., Corey Beck, M.D., Steve Garlow, M.D., Ph.D., Ebrahim Haroon, M.D., Maryann Jacob, M. D., Jeffrey Rakofsky, M.D., Dylan Wint, M.D. Clinical Coordinators: Yara Betancourt, Beatriz Blastos, Ronald Chismar, R.N., Melanie Galanti, Rachelle Gibson, R.N., Lauren Marx, Melissa McKenzie, and Tanja Mletzko Crowe. The ISLAND study also owes thanks to Tanja Mletzko Crowe, Kathleen Helms, Amber Lechner, and Valerie Cruz for their assistance as study coordinators; CBT therapists Allison Lopilato, Noriel Lim, and Devon Loparo; JungHo Cha for his assistance in imaging data preprocessing. Finally, we would like to thank all the participants who donated their time to the study.

Declaration of Competing Interest

Forest Laboratories and Elli Lilly donated the PReDICT study medications, escitalopram and duloxetine, respectively, and they were otherwise uninvolved in the study design, data collection, data analysis, or interpretation of findings. The authors report the following declaration of interests and financial disclosures: JCK has no financial disclosures. BWD: Dr. Dunlop received research support from Boehringer Ingelheim, Compass Pathways, NIMH, Otsuka, Sage, Usona Institute, and Takeda and has served as a consultant for Biohaven, Cerebral Therapeutics, Greenwich Biosciences, Myriad Neuroscience, NRx Pharmaceuticals, Otsuka, Sage, and Sophren Therapeutics. SEB: has no financial discolosres. CL: has no financial disclosures. HSM: Dr. Mayberg receives IP licensing fees from Abbott Labs and consults Abbott Labs, BlackRock Neuro, NextSense and Cogwear. WEC: Dr. Craighead receives research support from the NIH; is a board member of Hugarheill ehf, an Icelandic company dedicated to the prevention of depression; receives book royalties from John Wiley; and is supported by the Mary and John Brock Foundation, the Pitts Foundation, and the Fuqua family foundations. He is a consultant to the George West Mental Health Foundation and a member of the Scientific Advisory Boards of AIM for Mental Health and the ADAA.

References

[1]. Nolen-Hoeksema S Responses to depression and their effects on the duration of depressive episodes. J Abnorm Psychol 1991;100(4):569–82. 10.1037/0021-843X.100.4.569. [PubMed: 1757671]

[2]. Lyubomirsky S, Tucker KL, Caldwell ND, Berg K. Why ruminators are poor problem solvers: clues from the phenomenology of dysphoric rumination. J Pers Soc Psychol 1999;77(5):1041–60. 10.1037/0022-3514.77.5.1041. [PubMed: 10573879]

[3]. Morrow J, Nolen-Hoeksema S. Effects of responses to depression on the remediation of depressive affect. J Pers Soc Psychol 1990;58(3):519–27. 10.1037/0022-3514.58.3.519. [PubMed: 2324941]

[4]. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive Therapy of Depression. New York, NY: Guilford; 1979.

[5]. Nolen-Hoeksema S, Wisco BE, Lyubomirsky S. Rethinking rumination. Perspect Psychol Sci 2008;3(5):400–24. 10.1111/j.1745-6924.2008.00088.x. [PubMed: 26158958]

[6]. Watkins ER. Constructive and unconstructive repetitive thought. Psychol Bull 2008;134(2):163–206. 10.1037/0033-2909.134.2.163. [PubMed: 18298268]

[7]. Aldao A, Nolen-Hoeksema S, Schweizer S. Emotion-regulation strategies across psychopathology: a meta-analytic review. Clin Psychol Rev 2010;30(2):217–37. 10.1016/j.cpr.2009.11.004. [PubMed: 20015584]

[8]. Olatunji BO, Naragon-Gainey K, Wolitzky-Taylor KB. Specificity of rumination in anxiety and depression: a multimodal meta-analysis. Clin Psychol Sci Pract 2013; 20(3):225–57. 10.1111/cpsp.12037.

[9]. Abela JRZ, Hankin BL. Rumination as a vulnerability factor to depression during the transition from early to middle adolescence: a multiwave longitudinal study. J Abnorm Psychol 2011;120(2):259–71. 10.1037/a0022796. [PubMed: 21553940]

[10]. Just N, Alloy LB. The response styles theory of depression: tests and an extension of the theory. J Abnorm Psychol 1997;106(2):221–9. 10.1037/0021-843X.106.2.221. [PubMed: 9131842]

[11]. Nolen-Hoeksema S The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. J Abnorm Psychol 2000;109(3):504–11. 10.1037/0021-843X.109.3.504. [PubMed: 11016119]

[12]. Spinhoven P, Drost J, de Rooij M, van Hemert AM, Penninx BWJH. Is experiential avoidance a mediating, moderating, independent, overlapping, or proxy risk factor in the onset, relapse and maintenance of depressive disorders? Cogn Ther Res 2016;40(2):150–63. 10.1007/s10608-015-9747-8.

[13]. Nolen-Hoeksema S, Morrow J. Effects of rumination and distraction on naturally occurring depressed mood. Cogn Emot 1993;7(6):561–70. 10.1080/02699939308409206.

[14]. Hamilton JP, Farmer M, Fogelman P, Gotlib IH. Depressive rumination, the default-mode network, and the dark matter of clinical neuroscience. Biol Psychiatry 2015; 78(4):224–30. 10.1016/j.biopsych.2015.02.020. [PubMed: 25861700]

[15]. Jones NP, Siegle GJ, Thase ME. Effects of rumination and initial severity on remission to cognitive therapy for depression. Cogn Ther Res 2008;32(4):591–604. 10.1007/s10608-008-9191-0.

[16]. Teismann T, Willutzki U, Michalak J, Schulte D. Bedeutung von Rumination und Ablenkung für den Therapieerfolg depressiver Patienten. Verhaltenstherapie 2008; 18(4):215–22. 10.1159/000165687.

[17]. Watkins ER. Rumination-Focused Cognitive-Behavioral Therapy for Depression. New York, NY, US: Guilford Press; 2016.

[18]. Watkins ER, Nolen-Hoeksema S. A habit-goal framework of depressive rumination. J Abnorm Psychol 2014;123(1):24–34. 10.1037/a0035540. [PubMed: 24661156]

[19]. Watkins ER, Roberts H. Reflecting on rumination: consequences, causes, mechanisms and treatment of rumination. Behav Res Ther 2020;127:103573. 10.1016/j.brat.2020.103573. [PubMed: 32087393]

[20]. Kennedy JC, Dunlop BW, Craighead LW, Nemeroff CB, Mayberg HS, Craighead WE. Assessing in-session rumination and its effects on CBT for depression. Behav Res Ther 2022;159:104209. 10.1016/j.brat.2022.104209. [PubMed: 36283238]

[21]. Gotlib IH, Joormann J. Cognition and depression: Current status and future directions. Annu Rev Clin Psychol 2010;6(1):285–312. 10.1146/annurev.clinpsy.121208.131305. [PubMed: 20192795]

[22]. Hawes DJ, Dadds MR, Pasalich D. Observational coding strategies. In: Comer JS, Kendall PC, editors. The Oxford Handbook of Research Strategies for Clinical Psychology. New York, NY: Oxford University Press; 2013. p. 120–41.

[23]. Dunlop BW, Binder EB, Cubells JF, et al. Predictors of remission in depression to individual and combined treatments (PReDICT): study protocol for a randomized controlled trial. Trials 2012;13:106. 10.1186/1745-6215-13-106. [PubMed: 22776534]

[24]. Dunlop BW, Kelley ME, Aponte-Rivera V, et al. Effects of patient preferences on outcomes in the Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) study. Am J Psychiatry 2017;174(6):546–56. 10.1176/appi.ajp.2016.16050517. [PubMed: 28335624]

[25]. Dunlop BW, Rajendra JK, Craighead WE, et al. Functional connectivity of the subcallosal cingulate cortex and differential outcomes to treatment with cognitive-behavioral therapy or antidepressant medication for major depressive disorder. Am J Psychiatry 2017;174(6):533–45. 10.1176/appi.ajp.2016.16050518. [PubMed: 28335622]

[26]. Kennedy JC, Dunlop BW, Craighead LW, Nemeroff CB, Mayberg HS, Craighead WE. Follow-up of monotherapy remitters in the PReDICT study: Maintenance treatment outcomes and clinical predictors of recurrence. J Consult Clin Psychol 2018;86(2): 189–99. 10.1037/ccp0000279. [PubMed: 29369664]

[27]. Kelley ME, Choi KS, Rajendra JK, et al. Establishing evidence for clinical utility of a neuroimaging biomarker in major depressive disorder: Prospective testing and implementation Challenges. Biol Psychiatry 2021;90(4):236–42. 10.1016/j.biopsych.2021.02.966. [PubMed: 33896622]

[28]. First MB, Spitzer RL, Gibbon M, Williams JBW. Structured Clinical Interview for DSM-IV Axis I Disorders: Patient Edition (SCID-I/P, Version 2.0). New York: Biometrics Research Department, New York State Psychiatric Institute; 1995.

[29]. Hamilton M Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967;6(4):278–96. 10.1111/j.2044-8260.1967.tb00530.x. [PubMed: 6080235]

[30]. Nolen-Hoeksema S, Morrow J. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta Earthquake. J Pers Soc Psychol 1991;61(1):115–21. 10.1037/0022-3514.61.1.115. [PubMed: 1890582]

[31]. Williams JW. A structured interview guide for the Hamilton Depression Rating Scale. Arch Gen Psychiatry 1988;45(8):742–7. 10.1001/archpsyc.1988.01800320058007. [PubMed: 3395203]

[32]. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. Beck depression inventory (BDI). Arch Gen Psychiatry 1961;4(6):561–71. [PubMed: 13688369]

[33]. Rose AJ, Schwartz-Mette RA, Glick GC, Smith RL, Luebbe AM. An observational study of co-rumination in adolescent friendships. Dev Psychol 2014;50(9): 2199–209. 10.1037/a0037465. [PubMed: 25069053]

[34]. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420–8. 10.1037/0033-2909.86.2.420. [PubMed: 18839484]

[35]. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6(4): 284–90. 10.1037/1040-3590.6.4.284.

[36]. Haynes SN, Lench HC. Incremental validity of new clinical assessment measures. Psychol Assess 2003;15(4):456–66. 10.1037/1040-3590.15.4.456. [PubMed: 14692842]

[37]. Brady F, Warnock-Parkes E, Barker C, Ehlers A. Early in-session predictors of response to trauma-focused cognitive therapy for posttraumatic stress disorder. Behav Res Ther 2015;75:40–7. 10.1016/j.brat.2015.10.001. [PubMed: 26523887]

[38]. Nolen-Hoeksema S, McBride A, Larson J. Rumination and psychological distress among bereaved partners. J Pers Soc Psychol 1997;72(4):855–62. 10.1037/0022-3514.72.4.855. [PubMed: 9108698]

[39]. Ciesla JA, Roberts JE. Self-directed thought and response to treatment for depression: A preliminary investigation. J Cogn Psychother 2002;16(4):435–53. 10.1891/jcop.16.4.435.52528.

[40]. Hershenberg R, McDonald WM, Crowell A, et al. Concordance between clinician-rated and patient reported outcome measures of depressive symptoms in treatment resistant depression. J Affect Disord 2020;266:22–9. 10.1016/j.jad.2020.01.108. [PubMed: 32056880]

Table 1

Demographic and clinical characteristics of participants.

	PReDICT Sample (N = 35)		ISLAND Sample (N = 17)	
Characteristic	*M*	*SD*	*M*	*SD*
Age (yrs)	40.2	12.0	40.3	11.4
CBT Baseline[1] HAM-D	18.3	3.5	13.1	5.7
CBT Baseline[1] BDI	19.8	7.0	15.8	10.4
CBT Baseline[1] RRS	–	–	54.9	10.8
	n	%	*n*	%
Sex				
Female	15	42.9	11	64.7
Male	20	57.1	6	35.3
Race				
Black	4	11.4	6	35.3
Other	2	5.7	3	17.6
White	29	82.9	8	47.1
Ethnicity				
Hispanic	1	2.9	2	11.8
Non-Hispanic	34	97.1	15	88.2
Married/Cohabitating				
Yes	17	48.6	8	47.1
No	18	51.4	9	52.9
Employed full-time				
Yes	19	54.3	9	52.9
No	16	45.7	8	47.1
Anxiety disorder at baseline				
Yes	13	37.1	10	58.8
No	22	62.9	7	42.2
Previous episodes				
1	19	54.3	5	29.4
2	5	14.3	3	17.6
≥3	11	31.4	9	52.9
History of suicide attempt	1	2.9	2	11.8
Insurance status				
Yes	22	62.9	10	58.8
No	13	37.1	7	41.2

Note: HAM-D =17-item Hamilton Depression Rating scale; BDI = Beck Depression Inventory; RRS = Ruminative Response Scale;

[1]Baseline for PReDICT participants was the beginning of study participation, and baseline for the ISLAND study refers to the beginning of CBT, which occurred after 12 weeks of SSRI treatment

Table 2

Correlations among measures of in-session rumination, self-reported rumination, and depressive symptom severity in the ISLAND sample.

Variable	1	2	3	4	5	6	7
1. In-Session Rumination *Intensity*	1						
2. In-Session Rumination Duration	.943**	1					
3. CBT Baseline RRS	.395	.523*	1				
4. EOT RRS	.591*	.771**	.746**	1			
5. CBT Baseline HAM-D	.532*	.645**	.489*	.727**	1		
6. EOT HAM-D	.498*	.567*	.185	.322	.414	1	
7. CBT Baseline BDI	.568*	.644**	.706**	.685**	.843**	.566*	1
8. EOT BDI	.457	.579*	.556*	.647*	.500*	.876**	.727**

Note: RRS =Ruminative Response Scale; HAM-D =17-item Hamilton Depression Rating scale; BDI =Beck Depression Inventory; Baseline refers to the beginning of CBT, which occurred after 12 weeks of SSRI treatment; EOT = end of 12 weeks of CBT treatment

** $p < .01$

* $p < .05$

J Mood Anxiety Disord. Author manuscript; available in PMC 2024 November 01.

Table 3

Simultaneous regressions predicting end of CBT depressive symptom severity in the ISLAND sample with measures of baseline self-reported rumination and in-session rumination.

Dependent variable	Predictor Variable	β	*p*	R^2
End of treatment HAM-D				
	Model 1		.136	.248
	Baseline RRS	−.014	.957	
	In-session rumination *intensity*	.503	.066	
	Model 2		.055	.339
	Baseline RRS	−.154	.555	
	In-session rumination *duration*	.648	.023	
End of treatment BDI				
	Model 3		.037	.376
	Baseline RRS	.445	.073	
	In-session rumination *intensity*	.281	.242	
	Model 4		.021	.423
	Baseline RRS	.349	.165	
	In-session rumination *duration*	.396	.119	

Note: *RRS = Ruminative Response Scale; HAM-D =17-item Hamilton Depression Rating scale; BDI = Beck Depression Inventory*; Baseline for the refers to the beginning of CBT, which occurred after 12 weeks of SSRI treatment