Efficacy and Mediators of a Group Cognitive–Behavioral Therapy for Hoarding Disorder: A Randomized Trial

David F. Tolin
Institute of Living, Hartford, Connecticut, and Yale University School of Medicine

Bethany M. Wootton, Hannah C. Levy, Lauren S. Hallion, and Blaise L. Worden
Institute of Living, Hartford, Connecticut

Gretchen J. Diefenbach
Institute of Living, Hartford, Connecticut, and Yale University School of Medicine

James Jaccard
New York University

Michael C. Stevens
Institute of Living, Hartford, Connecticut, and Yale University School of Medicine

Objective: Hoarding disorder (HD) is a common and potentially debilitating psychiatric disorder. Thus far, psychological treatments have yielded modest effects and/or were time-consuming and costly to deliver. The aim of the present study was to test the efficacy of a brief group cognitive–behavioral therapy (CBT) for adults with HD and to test hypothesized mediators of treatment outcome. Method: Eighty-seven adults with a primary diagnosis of HD were randomized to either immediate CBT or wait list. CBT consisted of 16 weekly, 90-min group sessions that emphasized in-session practice of discarding and refraining from acquiring, decision-making and problem-solving training, emotional distress tolerance, motivational interviewing strategies, and contingency management. Participants were assessed at pretreatment, midtreatment, and posttreatment by an independent evaluator unaware of treatment condition. Results: CBT was efficacious for the symptoms of HD compared with wait list. Saving-related cognitions, but not subjective cognitive impairment, partially mediated treatment outcomes. Conclusion: Brief Group CBT is an efficacious and feasible treatment for adults with HD, and is partially mediated by reductions in maladaptive beliefs about possessions. Superiority trials comparing CBT to active treatments, and additional research into mechanisms of treatment outcome, are warranted.

What is the public health significance of this article?
Hoarding disorder (HD) is a common and often severe psychiatric illness. As a new diagnosis, it has received less research attention than many other disorders. Treatment efficacy for HD also lags behind that of many other psychiatric disorders. This study suggests that a relatively brief (16-session) Group CBT is efficacious for adults with HD, with 83% of patients classified as improved. These results support the consideration of Group CBT as a first-line treatment for HD.

Keywords: group therapy, obsessive–compulsive disorder, contingency management, behavior therapy

Supplemental materials: http://dx.doi.org/10.1037/ccp0000405.supp

This article was published Online First April 22, 2019.
David F. Tolin, Institute of Living, Hartford, Connecticut, and Department of Psychiatry, Yale University School of Medicine; Bethany M. Wootton, Hannah C. Levy, Lauren S. Hallion, and Blaise L. Worden, Institute of Living; Gretchen J. Diefenbach, Institute of Living, and Department of Psychiatry, Yale University School of Medicine; James Jaccard, Silver School of Social Work, New York University; Michael C. Stevens, Institute of Living, and Department of Psychiatry, Yale University School of Medicine.

Bethany M. Wootton is now at the Department of Psychology, University of Technology Sydney. Lauren S. Hallion is now at the Department of Psychology, University of Pittsburgh.

This work was funded by National Institute of Mental Health grant R01 MH101163. Clinicaltrials.gov identifier NCT01956344. The authors thank Amber Billingsley, Akanksha Das, Marla Genova, Benjamin Katz, and James Ransom for their assistance with data collection. The authors also thank Elizabeth Davis, Christina Gilliam, Scott Hannan, and Kristen Springer for serving as independent evaluators.

Correspondence concerning this article should be addressed to David F. Tolin, Institute of Living, 200 Retreat Avenue, Hartford, CT 06106. E-mail: david.tolin@hhchealth.org
Hoarding disorder (HD) was established as an independent diagnosis in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM–5)* (American Psychiatric Association, 2013). The hallmark characteristic of HD is difficulty discarding possessions because of a perceived need to save them or significant distress upon discarding. The excessive saving behavior results in the accumulation of clutter that precludes the use of living spaces for their intended purpose (e.g., an individual with HD might no longer be able to cook in the kitchen or sleep in the bedroom because of clutter). Excessive acquiring behavior (either through buying or the accumulation of free objects) is listed as a specifier in *DSM–5*, though research suggests that acquiring behavior is present in most cases of HD (Frost, Rosenfield, Steketee, & Tolin, 2013), and current models of the disorder posit a critical role of acquiring behavior in illness maintenance (e.g., Kyrios, 2014).

Estimates of the prevalence of HD suggest that it is a common mental disorder, with prevalence rates in epidemiological studies ranging from 2%–4% (Mueller, Mitchell, Crosby, Glaesmer, & de Zwaan, 2009; Nordsletten et al., 2013). HD is associated with substantial functional impairment and poor health-related quality of life (Drury, Ajmi, Fernandez de la Cruz, Nordsletten, & Mataix-Cols, 2014), as well as significant public health cost (Tolin, Frost, Steketee, Gray, & Fitch, 2008). Current cognitive–behavioral models posit that individuals with HD save items because of impaired decision-making processes, maladaptive beliefs about possessions, and behavioral avoidance. Decision-making impairments may relate, in part, to impaired cognitive function as evidenced by neuropsychological testing (see Woody, Kellman-McFarlane, & Welsted, 2014) and standardized self-report (Samuels et al., 2018). Clinical observation suggests that patients with HD frequently seem to have difficulty focusing on topics and tasks and often become easily overwhelmed under task demands. Although evidence for executive function impairment is mixed, some evidence points to problems of categorization (e.g., Grisham, Norberg, Williams, Certoma, & Kadib, 2010) that comport well with behavioral observations of difficulty making rapid decisions, prioritizing, and distinguishing important from unimportant information. Maladaptive beliefs about possessions include a heightened sense of responsibility for possessions, excessive sentimental attachment to objects, lack of memory confidence, and exaggerated need for control (Steketee, Frost, & Kyrios, 2003). As a result of these processes, the act of decision-making is perceived as aversive, leading people with HD to avoid these activities, which ultimately leads to increased clutter.

Cognitive–behavioral therapy (CBT) targets these purported maintenance factors via efforts to improve decision-making, challenge maladaptive beliefs, and decrease behavioral avoidance (Steketee & Frost, 2014). The efficacy of individual CBT was determined in a randomized controlled trial (RCT), in which CBT proved superior to a wait-list control condition, with a high rate of clinical response (Steketee, Frost, Tolin, Rasmussen, & Brown, 2010). Across studies, the pretreatment to posttreatment effect size of CBT for HD is large (g = 0.82; Tolin, Frost, Steketee, & Muroff, 2015). However, it is also noted that the rate of clinically significant change (CSC), defined as scores closer to the normative range than to the HD range (Jacobson & Truax, 1991), is only 35%, indicating that most patients with HD remain significantly symptomatic after treatment (Tolin et al., 2015).

There is significant room for improvement of CBT for HD, from both a feasibility and an efficacy perspective. The protocol used in the initial trial of individual CBT is lengthy (twenty-six 120-min sessions) and includes home visits that are not feasible for most community clinicians. In the current study, the protocol was shortened to sixteen 90-min sessions, and home visits were not included to maximize disseminability (i.e., acceptability and feasibility for practicing clinicians). Treatment was also conducted in a group, rather than individual format, with the aim of improving feasibility and decreasing sense of stigma (Chasson, Guy, Bates, & Corrigan, 2018). Group CBT for HD appears promising (Bodyrzlova, Audet, Bergeron, & O’Connor, 2018; Gilliam et al., 2011; Muroff, Steketee, Bratiotis, & Ross, 2012; Muroff et al., 2009), and pre- to posttreatment effects of individual versus Group CBT are comparable (Tolin et al., 2015). In the present study, several modifications to the CBT protocol were made in hopes of increasing treatment efficacy. First, the protocol (Table 1) includes regular in-session discarding practice (Steketee & Frost, 2014), in which patients bring possessions to the clinic and practice sorting and discarding. Second, problem-solving training is employed to improve cognitive efficiency. Third, emotional distress tolerance is emphasized as a means of emotion regulation, given findings that poor distress tolerance is associated with saving behavior (e.g., Shaw & Timpano, 2016). Fourth, strategies from motivational interviewing are included midway through treatment to improve compliance, which can be quite variable and is associated with treatment outcomes (Tolin, Frost, & Steketee, 2007). Finally, some elements of contingency management were also included to increase compliance; similar contingencies have shown promise when added to Group CBT in an open trial for HD (Worden, Bowe, & Tolin, 2017). An initial open trial (N = 20) of the present treatment protocol, combined with an Internet support program, showed large reductions in HD severity (d = 1.57) that were well above the average pre- to posttreatment effect in prior research (Ivanov et al., 2018).

In addition to examining the efficacy of Group CBT for HD, it is also important to understand mediators of clinical improvement, which are a necessary (though not sufficient) criterion for identifying mechanisms (see Tryon, 2018). Here, we examined two strong candidates: maladaptive cognitions and subjective cognitive impairment. Maladaptive cognitions such as a heightened sense of responsibility, excessive sentimental attachment, poor memory confidence, and need for control have been proposed to play a central mechanistic role in HD (Steketee et al., 2003). A previous RCT that utilized a somewhat more cognitively based individual treatment for HD (Steketee et al., 2010) found that reductions in these maladaptive beliefs mediated HD symptom reduction among treatment completers (Levy et al., 2017). It remains to be seen whether cognitive change would also mediate outcomes in the present group-based treatment, which is more skill-based than cognitively focused, or whether such an intervention would operate via an alternate mechanism. However, previous research in related disorders such as depression suggests that changes in maladaptive cognitions may be a key mechanism by which CBT exerts its effects, even when direct cognitive interventions are de-emphasized (Lorenzo-Luaces, German, & DeRubeis, 2015).

The skills-based nature of the present protocol suggests the possibility that improved cognitive function might be a critical mechanism of symptom improvement. Impaired cognitive function has been documented in patients with HD via objective neuropsychological testing, albeit with mixed results (see Woody et al., 2014). Here, we focus on subjective cognitive impairment, which may predict functioning and quality of life beyond the
contribution of objective neuropsychological test scores (e.g., Hill et al., 2017). HD is associated with a wide range of self-reported cognitive impairments (Samuels et al., 2018), including memory, distractibility, and inattention; these impairments in cognitive function are thought to maintain the avoidance of decision-making characteristic of HD. In particular, distractibility is a promising potential mediator. Impaired capacity to sustain attention is perhaps the most robust neuropsychological finding in HD (Woody et al., 2014), and subjective distractibility shows a larger difference between HD and healthy control participants than do other subjective cognitive functions (Tolin, Hallion, et al., 2018). Distractibility may contribute to difficulty discarding by impeding task persistence; improved focus, conversely, might mediate improved discarding and decreased clutter over the course of treatment. Given the emphasis on improving cognitive efficiency in the present CBT, it would be informative to determine the degree to which improved subjective cognitive function (particularly distractibility, but perhaps other cognitive failures as well) mediates outcomes.

The primary aim of the current study was to determine the efficacy of our redesigned 16-session Group CBT intervention for adults with HD. It was predicted that patients receiving CBT would show a greater reduction on measures of HD severity at posttreatment than would patients assigned to a wait list control condition. The secondary aim was to examine potential mediators of treatment outcome for patients receiving CBT. It was predicted that reductions in both maladaptive cognitions and in subjective cognitive impairment would be significant mediators of HD symptom reduction.

**Method**

**Participants**

Eighty-seven adult patients with a primary DSM-5 diagnosis of HD were enrolled in this study involving the neural mechanisms of CBT response for HD.1 Prospective participants were primarily recruited from newspaper, radio, and billboard advertisements; online advertising; and community lectures. Participant flow is outlined in Figure 1. Participants were enrolled between October 22, 2013, and

---

1 Functional magnetic resonance imaging (fMRI) data for this study will be presented in a separate article.
November 27, 2017. To be included in the study, clinical participants were required to (a) have a primary diagnosis of HD on the Diagnostic Interview for Anxiety, Mood, and Obsessive-Compulsive and Related Neuropsychiatric Disorders (Tolin, Gilliam, Wootton, et al., 2018) that was of at least moderate severity according to the Clinician’s Global Impression-HD (Tolin, Gilliam, Davis, et al., 2018); (b) be aged 20–65; (c) be taking no psychiatric medications or on a stable dose of psychiatric medications (antidepressants, benzodiazepines, and stimulants were allowed) for at least 8 weeks; (d) be willing and able to abstain from the use of stimulant or benzodiazepine medications on the day of testing (though not necessarily during the treatment itself); (e) be free of serious mental illness such as schizophrenia, bipolar disorder, or active substance abuse; (f) be right-handed; and (g) be free of nonremovable metal in the body, claustrophobia, or other factors that would preclude functional magnetic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet inclusion criteria; the most common reasons for exclusion were diagnostic resonance imaging (fMRI). Of the 135 prospective participants that were interviewed, 48 were excluded because of failing to meet

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) diagram.
“much improved”) and nonresponders (3 = “minimally improved” or higher). The CGI-HD shows good interrater and test–retest reliability (Tolin, Gilliam, Davis, et al., 2018).

Potential mediators. The Saving Cognitions Inventory (SCI; Steketee et al., 2003) is a 24-item self-report measure that yields four subscales: emotional attachment (e.g., “Losing this possession is like losing a friend”), concerns about memory (e.g., “Saving this means I don’t have to rely on my memory”), control over possessions (e.g., “I like to maintain sole control over my things”), and responsibility toward possessions (e.g., “I am responsible for the well-being of this possession”). Respondents rate from 1 to 7 the extent to which they experienced each thought when attempting to discard an object within the past week. The SCI has demonstrated good internal consistency and convergent and discriminant validity in previous samples (Steketee et al., 2003). Internal consistency estimates were acceptable in the present sample (attachment α = .94, memory α = .78, control α = .64, responsibility α = .73).

The Cognitive Failures Questionnaire (CFQ; Broadbent, Cooper, FitzGerald, & Parkes, 1982) is a 25-item self-report measure assessing the patient’s perception of his or her overall cognitive status. Items are rated on a 5-point scale and the CFQ is comprised of four reliable subscales (Wallace, Kass, & Stanny, 2002): Memory (α = .85 in the present sample), Distractibility (α = .83 in the present sample), Blunders (α = .82 in the present sample), and Memory for Names (α = .57 in the present sample).

Other measures. DSM–5 diagnoses were assessed using the Diagnostic Interview for Anxiety, Mood, and Obsessive-Compulsive and Related Neuropsychiatric Disorders (DIAMOND; Tolin, Gilliam, Wootton, et al., 2018), a semistructured clinical interview. The DIAMOND HD diagnosis shows excellent interrater reliability (κ = 0.86), very good test–retest reliability (κ = 0.64), and strong convergence with the Saving Inventory—Revised (Tolin, Gilliam, Wootton, et al., 2018). Primacy of diagnosis was determined using the DIAMOND severity ratings, in which each diagnosis is rated on a scale from 1 (normal or in remission) to 7 (extreme) based on the participant’s level of distress and functional impairment. The Depression Anxiety Stress Scales (DASS; Lovibond & Lovibond, 1995) is a 42-item self-report measure assessing three domains of negative emotion: depression, anxiety, and stress/tension. Each item is rated on a 4-point scale assessing symptom frequency over the past week. Subscales of the DASS show high internal consistency (α = .89–.96) and good discriminant and divergent validity (Lovibond & Lovibond, 1995). In the present study, the depression (DASS-D, α = .94), anxiety (DASS-A, α = .90) and stress (DASS-S, α = .92) subscales showed excellent internal consistency.

Treatment

The CBT followed a published treatment protocol that includes a therapist manual (Tolin, Wootton, Worden, & Gilliam, 2017) and client workbook (Tolin et al., 2017). The group followed a didactic and experiential format for 16 weekly, 90-min sessions. Table 1 outlines the major themes of the treatment and the major content of each session. A class-like group structure was intended to help focus attention on the topic at hand, as well as reduce self-stigma about hoarding behaviors. Two group leaders (a licensed psychologist, who had significant experience working with HD patients, as well as a postdoctoral fellow in clinical psychology) and 4–11 participants per group (M = 7.8) were present for each session. In-session exercises were used for skill acquisition and rehearsal. To practice refraining from acquiring, we created a “virtual store” of 10–20 free items that prior group members had identified as desirable; patients practiced looking at, holding, talking about, and then putting back the items. To practice sorting, decision-making, and discarding, participants were instructed to bring a box or bag of items from home to each session. The group room had boxes labeled “recycle/donate” and “trash” in the middle of the room. Psychoeducation modules involved using the “virtual store” and sorting boxes (e.g., during psychoeducation about emotions, participants were asked to discard an item, or to refrain from acquiring an item, and to identify their emotions). Early in the group, participants were educated about four hypothesized mechanisms of HD: (a) poor decision-making and problem-solving, (b) intense emotions, (c) maladaptive thought processes, and (d) low motivation. Two sessions were dedicated to address each of these mechanisms, with all including psychoeducation and in-session practice of skills. Sessions 11–14 employed all of these themes as patients continued in-session sorting and discarding practice. The final two sessions were dedicated to relapse prevention.

Homework followed each group session and included the completion of assignments and worksheets which helped the patient to review concepts and skills, track possessions going in and out of the home, and monitor between session (at-home) sorting and discarding exercises. Each session began with a homework review and ended with the setting of new discarding goals for the week. Participants were also instructed to take digital pictures of their living spaces at the beginning and end of treatment, which were reviewed in the group on a projector screen.

Several steps were taken to maximize motivation and treatment compliance. First, midway through treatment (where, in our earlier clinical observations, we found motivation often begins to wane), group time was spent using strategies from motivational interviewing (Miller & Rollnick, 2013), including weighing the pros and cons of behavioral change, and reviewing progress on long-term goals and values and comparing these to current behavior. Second, contingency management strategies, as employed in prior HD research (Worden et al., 2017), were used to target treatment attendance and homework completion. For example, participants kept logs of their goals each week (e.g., discard 10 newspapers) and a point was given when each homework assignment was completed and weekly goal met. At 6-week intervals, if 75% of the possible points had been awarded, a group reward was given (e.g., a free movie ticket or group party, but not an object that could add to clutter). The group rewards were earned in all but one group. At an individual level, participants were instructed to brainstorm meaningful, nonacquiring self-rewards (e.g., watch a favorite TV show, go out for ice cream) and contracted to self-administer these rewards on a weekly basis when that week’s goal was met.

Finally, we had a written procedure for managing noncompliance with treatment. An episode of noncompliance was defined as a missed session, or not completing more than 30% of the homework assignments for that session, which typically included the In/Out Log, the weekly discarding goal set in collaboration with the group leaders, treatment worksheets, and compliance with instructions to bring in items for in-session discarding decision-making exercises. At the first instance of noncompliance, the group leader reminded the patient of the importance of regular
Whether each of the elements of the session manual were present in the session. The trial was registered with clinicaltrials.gov (NCT01956344).

Data Analytic Strategy

Data were sought for all individuals at all time points whether they completed treatment or not for purposes of complete-case (intent-to-treat) analysis. Of the 14 individuals who dropped out of the study during treatment, posttreatment data were obtained for 12 of them. The use of complete case analysis often underestimates treatment efficacy but has the advantage of preserving randomization, which is not necessarily the case for the analysis of just treatment completers, also known as per protocol analysis (Mitka, 2012; National Research Council, 2010). Complete case analyses are reported here to preserve randomization but per protocol analyses that better estimate efficacy presuming dropouts do not disrupt randomization are provided in the online supplemental material. Sensitivity analyses are therefore necessary to provide estimates of robustness of findings, though these are not presented here. Treatment completers, also known as per protocol analyses, are reported here to preserve randomization but per protocol analyses that better estimate efficacy presuming dropouts do not disrupt randomization are provided in the online supplemental material.

HOARDING DISORDER
relationships (see online supplemental material). Significance tests of mediation used the joint significance test (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002), which tends to perform as well as more complex bootstrap methods for mediation analysis (Hayes & Scharkow, 2013).

Finally, to supplement the SEM analysis of mediation, a fixed effects regression analysis for panel data (Allison, 2009) was undertaken using all individuals, including the crossovers from the wait list control after they completed treatment; that is, there was no control group and all individuals had been exposed to treatment. The analysis used three time points, the baseline, a midtreatment assessment (at Week 8), and the posttest. Time was dummy-coded in the analysis. Fixed effects regression analysis removes from the data all sources of between-subjects variability, so any between-subjects confounds (e.g., gender, depression) are moot. The method essentially regresses the outcome variable onto a mediator on a within-person basis, documenting how much the outcome changes given a one unit change in the mediator within an individual, on average.

Results
Sample Description and Participant Flow

As shown in Figure 1, of the 41 patients randomized to wait list, 5 (12%) withdrew from that condition. Of the 46 patients randomized to CBT, 15 (33%) withdrew from that condition: six dropped out on their own accord (three prior to treatment and three during treatment), while nine were removed from treatment because of noncompliance. Of the 36 patients who completed the wait list condition and progressed to CBT, nine subsequently withdrew from that condition: five dropped out on their own accord (one prior to treatment and four during treatment) and four were removed because of noncompliance. Overall, of the 78 patients who started CBT, 7 (9%) dropped out on their own accord and 13 (17%) were removed from the group for noncompliance.

Descriptive statistics for the sample are shown in Table 2. No meaningful pretreatment differences were found on any measure between the CBT and wait list groups, suggesting adequacy of randomization. Scores on the SI-R and HRS-I were well within the clinical range, and the average patient was rated as “markedly ill” on the CGI-HD. Depression scores on the DASS were in the mild range on average, whereas the anxiety and stress scores were not clinically elevated. Scores on all SCI subscales were more than 1 SD above the normative means reported in the original SCI publication (Steketee et al., 2003), indicating significant endorsement of maladaptive saving beliefs. The CFQ mean total score was 45.27 (SD = 15.80), which is nearly 2 SD higher than the healthy control comparison sample used for the fMRI data (results to be published in a separate paper). Similarly, scores on the Memory, Distractibility, and Blunders subscales were more than 1 SD higher in the hoarding patients than the healthy controls. The Memory for Names subscale score was within 1 SD of the healthy control group.

Treatment Fidelity and Independent Evaluator Awareness Check

Scores on the fidelity checklist were calculated as a percentage of items checked. Ratings across sessions ranged from 66.7% to 100% (M = 94.8, SD = 11.74). Thus, fidelity to the treatment protocol appeared excellent. Out of 167 assessments, raters guessed participants’ condition correctly 116 (69.5%) times, with 30 (18.0%) instances of reported unblinding by participants (the remainder of guesses were because of symptom change). Mean confidence rating (0–100%) for guesses was 65.27 (SD = 21.02).

Efficacy of Group CBT for Hoarding Disorder

Primary outcome. Table 3 presents means and clinically significant change (CSC) percentages for the treatment and control conditions for the SI-R total and subscale scores. We determined CSC by calculating the statistic C, representing the cutoff between clinical and normative groups. C is calculated as $C = ((M_{HD})(\sigma) + (\mu)(SD_{HD})/(SD_{HD} + \sigma))$, where $M_{HD}$ and $SD_{HD}$ are the established mean and SD for patients with HD, and $\mu$ and $\sigma$ are the normative mean and SD (Tolin et al., 2015). We used the HD and normative means provided by Muroff, Underwood, and Steketee (2014), which resulted in the following cutoff values; SI-R total, 43.22; SI-R saving, 14.50; SI-R clutter, 17.89; and SI-R acquiring, 9.92. The differences were statistically significant ($p < .01$) in favor of the CBT (vs. wait list) condition for all major outcomes and all subscales. The standardized mean differences (based on Cohen’s $d$) generally were large (1.27 to 1.39) and the percent differences generally were substantial ($-4.21$ to $-16.24$). On the
Outcome means at posttreatment contrasts for study outcomes

Mediators of change in group CBT

Joint significance test. Table 4 presents group differences in means and percentages for the treatment and control conditions for the proposed mediators as well as their subscales. The group differences for CBT versus control were significant and moderate to large (\(d = 0.52\) to 0.86) for all mediators except for the CFQ Names subscale. The CFQ Memory subscale was statistically significant for the unadjusted contrast (critical ratio \(\text{CR} = 2.31, p < .05\)) but not for the cluster adjusted contrast \(\text{CR} = 1.87, p < .07\). The Cohen’s \(d\) coefficients for statistically significant effects generally were near 0.70.

The primary mediational analyses focused on the total scores of the SCI (maladaptive beliefs about saving) and the CFQ (subjective cognitive impairment). The joint significance test for the relevant structural model (see the online supplemental material) predicts that in addition to significant treatment-control group differences on the mediators reported in Table 4, the path/regression coefficients from SCI and CFQ to the primary outcome variables also should be statistically significant. Huber-White robust regression was performed on each outcome separately using as predictors the SCI, CFQ, the treatment condition (dummy coded), and the baseline measure of the outcome (to increase statistical power in an analysis of covariance sense; see Rausch, Maxwell, & Kelley, 2003). Table 5 presents the relevant regression coefficients, standard error of estimates, and squared multiple

### Table 3

**Contrasts for Study Outcomes**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>CBT group</th>
<th>Waitlist group</th>
<th>CBT − waitlist difference</th>
<th>Critical ratio</th>
<th>Cohen’s (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRS-I Total</td>
<td>17.28</td>
<td>25.84</td>
<td>−8.56 ± 2.06</td>
<td>8.27</td>
<td>0.58</td>
</tr>
<tr>
<td>HRS-I Clutter</td>
<td>41.7</td>
<td>8.1</td>
<td>33.6 ± 18.8</td>
<td>3.58</td>
<td>—</td>
</tr>
<tr>
<td>HRS-I Saving</td>
<td>41.7</td>
<td>5.4</td>
<td>36.3 ± 18.0</td>
<td>4.02</td>
<td>—</td>
</tr>
<tr>
<td>HRS-I Acquiring</td>
<td>52.8</td>
<td>13.5</td>
<td>39.3 ± 20.0</td>
<td>3.91</td>
<td>—</td>
</tr>
<tr>
<td>Mediator means CBT group</td>
<td>22.2</td>
<td>2.33</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

### Table 4

**Contrasts for Mediators**

<table>
<thead>
<tr>
<th>Mediator means</th>
<th>CBT group</th>
<th>Waitlist group</th>
<th>CBT − waitlist difference</th>
<th>Critical ratio</th>
<th>Cohen’s (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCI-Total</td>
<td>79.23</td>
<td>98.32</td>
<td>−19.09 ± 10.70</td>
<td>3.57</td>
<td>.86 ± .48</td>
</tr>
<tr>
<td>SCI-Emotion</td>
<td>29.50</td>
<td>37.36</td>
<td>−7.86 ± 5.08</td>
<td>3.03</td>
<td>.71 ± .49</td>
</tr>
<tr>
<td>SCI-Memory</td>
<td>16.34</td>
<td>20.65</td>
<td>−4.31 ± 2.55</td>
<td>3.39</td>
<td>.82 ± .49</td>
</tr>
<tr>
<td>SCI-Responsibility</td>
<td>19.35</td>
<td>23.77</td>
<td>−4.43 ± 3.06</td>
<td>2.83</td>
<td>.66 ± .49</td>
</tr>
<tr>
<td>SCI-Control</td>
<td>13.87</td>
<td>16.14</td>
<td>−2.28 ± 1.69</td>
<td>2.64</td>
<td>.63 ± .50</td>
</tr>
<tr>
<td>CFQ-Total</td>
<td>41.26</td>
<td>47.22</td>
<td>−5.96 ± 4.18</td>
<td>2.85</td>
<td>.67 ± .46</td>
</tr>
<tr>
<td>CFQ-Memory</td>
<td>10.69</td>
<td>12.55</td>
<td>−1.86 ± 1.60</td>
<td>2.31</td>
<td>.55 ± .50</td>
</tr>
<tr>
<td>CFQ-Distraction</td>
<td>16.75</td>
<td>19.61</td>
<td>−2.86 ± 1.81</td>
<td>3.17</td>
<td>.75 ± .46</td>
</tr>
<tr>
<td>CFQ-Blunder</td>
<td>10.36</td>
<td>11.85</td>
<td>−1.50 ± 1.30</td>
<td>2.25</td>
<td>.52 ± .45</td>
</tr>
<tr>
<td>CFQ-Names</td>
<td>4.56</td>
<td>4.76</td>
<td>−.19 ± .57</td>
<td>(.67 (ns))</td>
<td>.16 ± .48</td>
</tr>
</tbody>
</table>

Note. CBT = cognitive–behavioral therapy; SCI = Saving Cognitions Inventory; CFQ = Cognitive Failures Questionnaire. Critical ratio (CR) is the parameter estimate divided by its estimated \(SE\). CR > 1.96 is \(p < .05\) and CR > 2.56 is \(p < .01\). CBT \(n = 46\), Waitlist \(n = 41\); entries after ± are half widths of 95% confidence intervals, analogous to a margin of error.

SI-R. 42% of treated patients met CSC criteria, compared with 8% of waitlist patients. The number needed to treat (NNT) index (Pinson & Gray, 2003) was 2.98.

Secondary outcomes. As shown in Table 3, HRS-I scores significantly favored CBT over wait list, with a large standardized mean difference (1.95). It is noted, however, that only 17% of treated patients (vs. 0% of waitlist patients) met criteria for CSC on the HRS-I. We used the same formula described above for determining CSC, and used the HD and normative means from the original HRS-I validation study (Tolin et al., 2010). This yielded a cutoff score of 13.09 for CSC on the HRS-I. For the CGI-HD scale, 83.3% of the treated individuals were rated as much improved or very much improved, compared with 11.1% in the waitlist control group, a rate that is 7.5 times higher.

Mediators of Change in Group CBT
correlations. The CFQ did not independently predict SI-R outcomes after the other variables were controlled, which is inconsistent with mediation. The SCI significantly and independently predicted change in SI-R. Treatment condition remained a statistically significant regression coefficient for both hoarding outcomes, suggestive of partial mediation by change in SCI (i.e., SCI accounted for some but not all of the effect of CBT on SI-R). An estimate of how much of the overall treatment effect on SI-R was because of the SCI mediator was 16%, with about 84% of the program effect being because of mechanisms/mediators other than SCI. Tests of treatment by mediator interactions on the outcome using product terms did not yield statistically significant coefficients, suggesting no meaningful treatment by mediator interaction.

Fixed effects panel regression. The fixed effects panel regression analyses tested if changes in the two mediators (entered into the regression equation simultaneously) were associated with changes in each hoarding outcome from baseline to midtreatment to posttreatment on a purely within-individual basis. For the SI-R outcome, there was statistically significant reduction in total SI-R scores between baseline and midtreatment (mean difference = −6.52 ± 2.79, CR = 4.58, p < .05) as well as a statistically significant reduction from midtreatment to the immediate posttreatment (mean difference = −6.40 ± 2.65, CR = 4.72, p < .05). The coefficient for CBT was 0.09 ± 0.24 (CR = 0.73, ns), suggesting that changes in CFQ did not mediate changes in SI-R.

Additional analyses (see online supplemental material) demonstrated that the study was adequately powered, that aggregation bias did not appear to affect results, and that the conclusions of the mediation analysis were reasonable in the presence of modest violations of sequential ignorability (i.e., correlated disturbances between the mediators and outcomes; see Imai, Keele, & Yamamoto, 2010).

Discussion

The first aim of this study was to investigate the efficacy of a 16-session Group CBT intervention for adults with HD in a randomized controlled trial comparing an immediate treatment group to a wait list control condition. It was hypothesized that the Group CBT intervention would outperform the wait list control condition at posttreatment, with large between group effect sizes. The present results provide preliminary support for the efficacy of Group CBT for HD. Between-groups effects were large (d = 1.59 for the SI-R Total and 1.95 for the HRS-I). This effect compares favorably to prior controlled trials for HD. For example, Steketee et al. (2010) conducted a trial of individual CBT for HD and after 12 weeks the effect size between CBT and wait list groups was 1.07 on the SI-R and 0.71 on the HRS-I. In another study of a 20-week Group CBT with multiple home visits, Muroff et al. (2012) obtained a difference of d = 1.03 on the SI-R but only 0.21 on the HRS-I compared to bibliotherapy. Given the larger effects seen in the present study it is possible that the protocol used results in enhanced outcomes over previously used protocols for both individual and Group CBT for HD. However, this requires further evaluation in future controlled trials.

In the present study 42% of patients met criteria for CSC on the SI-R, which compares favorably to the 35% rate of CSC seen across CBT studies (Tolin et al., 2015), though the majority (58%) of patients were still in the clinical range (lack of CSC) at the conclusion of treatment. Rates of CSC were significantly worse on the HRS-I, with only 17% meeting criteria (and thus 83% still remaining in the clinical range). The HRS-I is likely a more complete (and conservative) measure of outcome, as evidenced by its smaller pre- to posttreatment effects in this and prior research Muroff et al. (2012). This could be because the revised HRS-I used in this study has detailed probe and clarification questions (Tolin, Gilliam, Davis, et al., 2018) that help the interviewer obtain a clearer impression of the severity of symptoms; it is noted that the correlation between the SI-R total score and HRS-I was modest. It is also noted that the HRS-I measures not only the domains tapped by the SI-R (clutter, difficulty discarding, and acquiring) but also hoarding-related distress and impairment; it is possible that these factors attenuated the outcomes as changes in functional impairment frequently lag behind those in the behavioral features of hoarding (Tolin et al., 2015). It is also quite possible that the HRS-I is less sensitive to treatment effects; additional research into this topic is warranted.

There are a number of factors which may enhance the efficacy of CBT for HD. First, as noted previously (Tolin et al., 2015), across studies longer treatment is associated with stronger effects, particularly in the domains of clutter and functional impairment, and it is possible that a longer course of treatment, would be of benefit. Second, enhancing between-session compliance during treatment may also be beneficial given the strong relationship between compliance and outcome (Ayers, Wetherell, Golshan, & Saxena, 2011; Tolin et al., 2007). Given the high rates of attentional deficits (e.g., Grisham, Brown, Savage, Steketee, & Barlow, 2007) and memory confidence concerns (Hartl et al., 2004) seen in this population, automated prompts and reminders may be a pragmatic low-cost way to enhance between-session compliance. While the results from this study provide preliminary support for Group CBT for HD, future research should focus on how to improve this effect further with dismantling studies.

The second aim of the present study was to examine potential mediators of treatment outcome for patients receiving CBT. It was predicted that reductions in both maladaptive cognitions and in

Table 5
Saving Inventory–Revised (SI-R) Based Mediational Regressions

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Unstandardized coefficient</th>
<th>Critical ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>16.45 ± 22.04</td>
<td>—</td>
</tr>
<tr>
<td>Baseline measure</td>
<td>−5.8 ± .25</td>
<td>4.57</td>
</tr>
<tr>
<td>Treatment group</td>
<td>−14.36 ± 5.78</td>
<td>6.31</td>
</tr>
<tr>
<td>SCI</td>
<td>.14 ± .09</td>
<td>2.92</td>
</tr>
<tr>
<td>CFQ</td>
<td>−.14 ± .22</td>
<td>1.21</td>
</tr>
<tr>
<td>Multiple R²</td>
<td>.55</td>
<td>—</td>
</tr>
<tr>
<td>Standard error of estimate</td>
<td>9.68</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. SCI = Saving Cognitions Inventory; CFQ = Cognitive Failures Questionnaire. Critical ratio (CR) is the parameter estimate divided by its estimated standard error. CR > 1.96 is p < .05 and CR > 2.56 is p < .01. Entries after ± are half widths of 95% confidence intervals, analogous to a margin of error. Total N = 87.
subjective cognitive impairment would be significant mediators of HD symptom reduction. Consistent with prior research using a more cognitively based, manualized, individual CBT (Levy et al., 2017), the findings further suggest a mechanistic role for changes in hoarding-related beliefs in the beneficial effects of CBT on HD, accounting for 16% of treatment effects. This finding is consistent with emerging evidence that cognitive change mediates outcomes even for more behavioral treatments in related disorders (e.g., depression; Lorenzo-Luaces et al., 2015). Thus, it is important for future research to investigate how these hoarding-related beliefs can be more effectively addressed in treatment. The failure of subjective cognitive impairment to mediate treatment outcomes is also notable given the greater emphasis on cognitive skill-building and lesser emphasis on cognitive restructuring, compared with previous CBT protocols. Patients with HD’s subjective reports of cognitive functioning do not reliably align with objective measures (Moshier et al., 2016); it therefore remains plausible that objective improvements in cognitive functioning may have contributed to treatment gains despite being undetected by patients. Future research should examine both psychological and biological variables that might further account for symptom improvement. Prior research has begun to link the discarding decision-making process to specific abnormalities in neural activity (e.g., Tolin et al., 2012). Neuroimaging was included in the present study at pre-, mid-, and posttreatment; future analyses should examine whether abnormal patterns of neural activity mediate treatment response as well as hypothesized mediators (e.g., distractibility).

Several important limitations of the present study should be noted. First, the use of a wait list control rather than placebo or *bona fide* treatment precludes strong conclusions about the specificity of the effects to CBT per se (i.e., vs. nonspecific factors). Wait list was deemed to be the most appropriate control condition for the present study given the dearth of randomized controlled trials of CBT for HD in the literature (e.g., only one small controlled trial of Group CBT in which treatment was compared with a bibliotherapy condition; Muroff et al., 2012). However, we note that wait list can serve as a “nocebo” condition, potentially inflating estimates of efficacy (Furukawa et al., 2014). Future studies should compare Group CBT to active control conditions.

Second, there was no follow-up assessment after treatment; therefore, it is unclear whether treatment gains were maintained after treatment discontinuation. This concern is attenuated somewhat by the findings of Muroff, Steketee, Frost, and Tolin (2014), who found generally strong maintenance of treatment gains 1 year after individual CBT for HD. Nevertheless, it is important to examine the long-term efficacy of Group CBT for HD. Such an investigation is particularly important for the current treatment protocol given it differs significantly from treatment delivered in previous trials of CBT for HD, which focused more heavily on traditional cognitive interventions.

Third, though treatment fidelity of prescribed interventions was high there were no systematic ratings made for the inclusion of nonmanual interventions. While the fidelity rater informally noted that these occurred rarely, if at all, in sessions, the lack of a formal quantitative assessment is a limitation of the current study. Fourth, because these data were collected as part of an ongoing clinical trial using fMRI, our inclusion criteria (e.g., medication restrictions, at least moderate HD severity, right-handedness) may not be representative of all treatment-seeking HD patients. The age restriction (65 years or younger, resulting in a mean age of 54 years), used to avoid age-related neuroimaging confounds, is particularly noteworthy: HD is common among older adults (Samuels et al., 2008), and some evidence suggests attenuated treatment response in older HD patients (Ayers et al., 2011). Aging-related decline in executive functioning may introduce unique complications (e.g., exacerbation of impaired decision-making). It will be important to examine whether HD patients over the age of 65 respond to this treatment as well as do those under 65, or whether additional age-related interventions are needed (e.g., Ayers et al., 2014).

Fourth, the unblinding of 18% of IE ratings is another limitation that attenuates the strength of conclusions that can be drawn. Ideally, these raters would have been replaced. In terms of the statistical modeling of mediation, parameter estimates can be biased by endogeneity, specification error, omitted variables, and measurement error, thus requiring caution in conclusions about mediation.

Fifth, we did not administer clutter-specific measures (e.g., the Clutter Image Rating; Frost, Steketee, Tolin, & Renaud, 2008), although we assessed self-reported clutter severity using the clutter subscale of the SI-R. Clutter-specific measures may have provided additional insight into HD symptom change and potential mechanisms of treatment response. We also did not explicitly assess HD-related impairment using an impairment-specific measure. However, we note that impairment is one of the five HRS-I items completed by the IEs.

Sixth, the 26% noncompletion rate is a limitation of the present study. Of the 78 patients who started CBT, 9% dropped out on their own accord and 17% were removed for noncompliance. Other studies of CBT have not included the stepped consequences for noncompliance included in the present protocol, which we thought to be important for maximizing treatment compliance (particularly in a group setting, in which a grossly noncompliant patient could easily affect the compliance of other group members). Therefore, it is problematic to compare the noncontinuation rate with the patient-initiated dropout rates in other studies of HD. Steketee et al.’s (2010) patient-initiated dropout rate of 10% is comparable to the 9% seen here. Mataix-Cols, Marks, Greist, Kobak, and Baer (2002), in a large study of OCD patients, found that 27% of those with significant hoarding symptoms dropped out of behavior therapy, a figure that is quite similar to the noncompletion rate seen here. The pros and cons of removing 17% of patients for noncompliance in the present study is debatable, though we note that this was done only after five instances of noncompliance with repeated warnings.

Finally, in many clinical trials, the sample was restricted. Our sample was largely White and female, which likely does not reflect the entire HD population in the United States (Samuels et al., 2008). For this reason, it will be important to replicate these findings in more diverse samples of individuals with HD in the future.

The preliminary efficacy of Group CBT demonstrated in the present study suggests a need for both dismantling studies and comparisons with other treatments. Dismantling studies would be valuable as a means to further clarify the active ingredients that are most strongly associated with improvement, which would in turn promote a better understanding of underlying mechanisms. Comparisons with pharmacotherapy, and examinations of the efficacy of combined therapy versus monotherapy, would be useful. The efficacy of pharmacotherapy for HD has not been established in
controlled trials, although some medications, particularly venlafaxine (Saxena & Sumner, 2014) and atomoxetine (Grassi et al., 2016) appear promising in pilot open trials. No non-CBT psychosocial treatments have been tested to date for HD, rendering comparisons to other psychosocial treatment (other than active placebo interventions) premature. Finally, the disseminability of this Group CBT approach to community mental health settings should be examined. Unlike some of the prior work on CBT, the present protocol was designed to be feasible for this application, with no home visits and only 16 weekly sessions. Dissemination efforts have already suggested that many of the principles of CBT can be used successfully by lay group leaders (Frost, Pekareva-Kochergina, & Maxner, 2011; Frost, Ruby, & Shuer, 2012), and a broader effectiveness trial is warranted.

References


