
1-1-2018

Psychometric Properties of the Hoarding Rating Scale-Interview

David F. Tolin
Institute of Living

Christina M. Gilliam
Institute of Living

Elizabeth Davis
Institute of Living

Kristen Springer
Institute of Living

Hannah C. Levy
Institute of Living

See next page for additional authors

Follow this and additional works at: https://scholarworks.smith.edu/psy_facpubs



Part of the [Psychiatry and Psychology Commons](#)

Recommended Citation

Tolin, David F.; Gilliam, Christina M.; Davis, Elizabeth; Springer, Kristen; Levy, Hannah C.; Frost, Randy O.; Steketee, Gail; and Stevens, Michael C., "Psychometric Properties of the Hoarding Rating Scale-Interview" (2018). Psychology: Faculty Publications, Smith College, Northampton, MA.
https://scholarworks.smith.edu/psy_facpubs/33

This Article has been accepted for inclusion in Psychology: Faculty Publications by an authorized administrator of Smith ScholarWorks. For more information, please contact scholarworks@smith.edu

Authors

David F. Tolin, Christina M. Gilliam, Elizabeth Davis, Kristen Springer, Hannah C. Levy, Randy O. Frost, Gail Steketee, and Michael C. Stevens



HHS Public Access

Author manuscript

J Obsessive Compuls Relat Disord. Author manuscript; available in PMC 2019 September 20.

Published in final edited form as:

J Obsessive Compuls Relat Disord. 2018 January ; 16: 76–80. doi:10.1016/j.jocrd.2018.01.003.

Psychometric Properties of the Hoarding Rating Scale-Interview

David F. Tolin^{1,2}, Christina M. Gilliam¹, Elizabeth Davis¹, Kristen Springer¹, Hannah C. Levy¹, Randy O. Frost³, Gail Steketee⁴, Michael C. Stevens^{1,2}

¹Institute of Living, Hartford, CT

²Yale University School of Medicine

³Smith College

⁴Boston University School of Social Work

Abstract

The present study tested the psychometric properties of an expanded version of the *Hoarding Rating Scale* (HRS-I), a semistructured interview for hoarding disorder (HD). Eighty-seven adults with HD and 44 healthy control (HC) participants were assessed using the HRS-I and completed a battery of self-report measures of HD severity, negative affect, and functional impairment. All interviews were audio recorded. From the HD participants, 21 were randomly selected for inter-rater reliability (IRR) analysis and 11 for test-retest reliability (TRR) analysis. The HRS-I showed excellent internal consistency ($\alpha = 0.87$). IRR and TRR in the HD sample were good (intra-class coefficients = 0.81 and 0.85, respectively). HRS-I scores correlated strongly with scores on the self-report Saving Inventory-Revised (SI-R); partial correlations indicated that the HRS-I clutter, difficulty discarding, and acquiring items correlated significantly and at least moderately with corresponding SI-R subscales, when controlling for the other SI-R subscales. The HD group scored significantly higher on all items than did the HC group, with large effect sizes ($d = 1.28$ to 6.58). ROC analysis showed excellent sensitivity (1.00) and specificity (1.00) for distinguishing the HD and HC groups with a cutoff score of 11. Results and limitations are discussed in light of prior research.

Key works

Hoarding disorder; assessment; interview; psychometrics

The Hoarding Rating Scale-Interview (HRS-I; Tolin, Frost, & Steketee, 2010) is a 5-item semi-structured interview that was designed to capture the key aspects of hoarding disorder (HD): (1) clutter in the home, (2) difficulty discarding possessions, (3) excessive acquiring of possessions, (4) distress due to hoarding, and (5) functional impairment due to hoarding.

Author for correspondence: David F. Tolin, Ph.D., Institute of Living, 200 Retreat Avenue, Hartford, CT 06106, Phone 860-545-7685, Fax 860-545-7156, david.tolin@hhchealth.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Each item is rated on a 9-point scale from 0–8, and the item scores are summed to create a total score (range = 0–40), with higher scores indicating greater HD severity. The initial validation study (Tolin et al., 2010) was conducted using 73 adults with HD, 19 with obsessive-compulsive disorder (OCD), and 44 healthy control (HC) participants. Reliability was determined by having the same rater complete the HRS-I on two different occasions, first in the clinic, and then in the participants' homes. Correlations among these two administrations, for the HRS-I individual items and for the total score, were very good, ranging from 0.85–0.94. The HRS-I correlated significantly with a self-report measures of HD, the Saving Inventory-Revised (SI-R; Frost, Steketee, & Grisham, 2004), and reliably distinguished participants with HD from those without [area under the curve (AUC) ranged from 0.93–0.99]. A cutoff score of 14 on the HRS-I total score showed optimal sensitivity (0.97) and specificity (0.97).

Subsequent research using the HRS-I (Wootton et al., 2015) demonstrated that this measure correlated strongly with the hoarding subscale of the Obsessive-Compulsive Inventory-Revised (Foa et al., 2002). In youths, the HRS-I showed excellent internal consistency and scores differed significantly between those with and without HD (Park et al., 2016). The HRS-I appears sensitive to the effects of cognitive-behavioral therapy (Steketee, Frost, Tolin, Rasmussen, & Brown, 2010), with scores decreasing significantly after treatment. In a population-based survey, a self-report version of the measure was shown to correlate significantly with measures of buying and acquiring free things, as well as associated features of perfectionism, indecision, and procrastination (Timpano et al., 2011). Thus, the research to date suggests that the HRS-I is both reliable and valid as a measure of HD severity.

The aim of the present study was to address several important methodological limitations of the Tolin et al. (2010) initial validation study. First, as no validated diagnostic measure for HD existed at that time, the HD sample was diagnosed using the HRS-I itself, possibly inflating the estimated known-group validity. To address this concern, in the present study participants were diagnosed based on a validated structured diagnostic interview. Second, the initial validation study did not measure inter-rater reliability; we therefore examined the inter-rater reliability of the HRS-I in the present study. Third, the test-retest reliability assessment in the initial validation study was confounded by context (the rater completed the measure first in the clinic, then in the participant's home) which may have affected the correlation coefficients. Accordingly, in the present study test-retest reliability was assessed in the same context. Finally, the test-retest/cross-context reliability analyses in the original validation study were conducted using the entire sample (HD, OCD, and HC) which had non-overlapping distributions that could have inflated the reliability estimates. We therefore examined inter-rater and test-retest reliability specifically in the HD sample. Using expanded instructions for the HRS-I (see *Method*), we predicted that the measure would show good inter-rater and test-retest reliability, as well as good convergent validity with self-report measures and good known-groups validity as evidenced by strong sensitivity and specificity to differentiate HD from non-HD participants.

Method

Participants

Eighty-seven adult outpatients meeting DSM-5 (American Psychiatric Association, 2013) criteria for HD were sampled as part of a large clinical trial examining the neural mechanisms of CBT response in hoarding disorder. To be included in the study clinical participants were required to (1) have a primary diagnosis of HD of at least moderate severity; (2) be age 18–65; (3) be unmedicated or on a stable dose of psychiatric medications for at least 8 weeks, (4) be willing and able to abstain from the use of stimulant or benzodiazepine medications on the day of testing; (5) be right-handed, and (6) be free of non-removable metal in the body, claustrophobia, or other factors that would preclude functional magnetic resonance imaging (fMRI). Of 135 prospective clinical participants, 48 were excluded due to failing to meet inclusion criteria; the most common reasons for exclusion were HD not being the primary diagnosis ($n = 9$), HD symptoms being too mild ($n = 9$), and presence of a serious mental disorder (e.g., psychosis, bipolar disorder; $n = 9$).

Forty-four healthy control (HC) participants were also recruited. To be eligible for the study the HC participants were required to 1) have no current or past psychiatric diagnosis or treatment; (2) be aged 40–65 (for age matching to the HD sample); (3) be right-handed; and (4) be free of non-removable metal in the body, claustrophobia, or other factors that would preclude fMRI. Of the 60 prospective HC participants, 16 were excluded due to failing to meet inclusion criteria; the most common reasons for exclusion were subclinical HD symptoms ($n = 4$), current psychiatric symptoms ($n = 4$), and abnormal MRI findings ($n = 4$).

Measures

DSM-5 diagnoses were assessed using the *Diagnostic Interview for Anxiety, Mood, and Obsessive-Compulsive and Related Neuropsychiatric Disorders* (DIAMOND; Tolin et al., 2016), a semi-structured clinical interview. The DIAMOND HD diagnosis shows excellent inter-rater reliability ($\kappa = 0.86$), excellent test-retest reliability ($\kappa = 0.64$), and strong convergence with the Saving Inventory-Revised (Tolin et al., 2016). The DIAMOND HD diagnosis consists of yes/no questions, with optional prompt questions, for clutter in the home, difficulty discarding, distress about symptoms, and functional impairment. The diagnosis is assigned according to the symptom criteria listed in the *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (American Psychiatric Association, 2013).

The determination of at least moderate HD severity was made using the *Clinician's Global Impression-Hoarding Disorder* (CGI-HD) scale. Based on the original CGI (Guy, 1976), the CGI-HD, a new scale, is an 8-point rating from 1 (normal, not at all ill) to 7 (extremely ill). Interviewers rate, on this scale, the severity of 6 dimensions: (a) clutter, (b) difficulty discarding, (c) acquiring, (d) health or safety hazard, (e) functional impairment, and (f) distress. The CGI-HD score is calculated as the highest of these ratings (e.g., a “severe” rating for health and safety hazard merits an overall CGI-HD score of “severe,” even if certain other features such as acquiring are not coded as severe). The CGI-HD showed good

inter-rater reliability (ICC = 0.72) and test-retest reliability (ICC = 0.81) in the present sample.

Hoarding symptom severity was assessed with the *Saving Inventory-Revised* (SI-R; Frost et al., 2004), a 23-item self-report measure that yields a total score as well as three subscales: Clutter (α in present sample = .98), difficulty discarding (α in present sample = .96), and acquiring (α in present sample = .94). The SI-R readily discriminates HD from OCD patients and community controls, and correlates significantly with ratings of clutter and impairment (Frost et al., 2004).

Affective symptoms were measured using the *Depression Anxiety Stress Scales* (DASS; Lovibond and Lovibond 1995), a 42-item self-report measure assessing three subscales of negative emotion: depression (DASS-D), anxiety (DASS-A), and stress/tension (DASS-S). Each item is rated on a 4-point scale assessing symptom frequency over the past week. DASS subscales have shown high internal consistency ($\alpha = 0.89\text{--}0.96$) and good discriminant and divergent validity (Brown, Chorpita, Korotitsch, & Barlow, 1997); internal consistency was excellent in the present study (DASS-D, $\alpha = .95$; DASS-A, $\alpha = .91$; DASS-S, $\alpha = .95$).

Functional impairment was assessed using the Emotional Role Functioning subscale of the *36-Item Short Form Health Survey (SF-36)* (Ware, 1993), a common measure of health-related quality of life (HRQoL). Three of the 36 items are summed to estimate role limitations due to emotional problems, which we used as a measure of functional impairment secondary to mental health concerns. Higher scores indicate better HRQoL (less impairment). Internal consistency in the present study was acceptable for this 3-item scale ($\alpha = .68$).

The *Hoarding Rating Scale-Interview* (HRS-I; Tolin et al., 2010) (see *Introduction*) is a 5-item semi-structured interview that assesses clutter, difficulty discarding, acquiring, distress, and impairment. Because initial pilot testing yielded inconsistent inter-rater reliability, two significant modifications (see Appendix) were made. First, scale anchor points were expanded to be more descriptive and to encompass a broader range of possible scenarios (e.g., for difficulty discarding, a rating of 4 is accompanied by the anchor description “Moderate, feels moderately distressed by discarding or avoids discarding some things (e.g., 50%) because of distress.” Second, each key question included supplemental follow-up questions that the interviewer could use as needed. For example, the interviewer could follow the key question “To what extent do you have difficulty discarding (or recycling, selling, giving away) ordinary things that other people would get rid of?” with (a) “How often do you try to discard things?”; (b) “When you try to discard things, how hard is it? How much discomfort do you feel?”; and/or (c) “Do you avoid discarding things? Why is that? What kinds of things do you avoid discarding, and what kinds of things do you not avoid? How hard would it be to discard the things you have been avoiding?”

Procedure

All study procedures were approved by the Hartford Hospital Institutional Review Board, and all participants provided written informed consent prior to any study procedures. The

present study was conducted as part of a clinical trial in which the HD group was seeking treatment. The HD group was recruited from the flow of clinic patients as well as from newspaper advertisements and flyers in the community. The HC group was recruited via newspaper advertisements and flyers in the community. Participants met with a doctoral-level psychologist or supervised postdoctoral fellow, trained in the use of the DIAMOND and HRS-I (training involved co-rating audiorecorded sessions, with a deviation of no more than one point on any given item over 2 consecutive assessments), in the clinic (no interviews were conducted in participants' homes). All interviews were digitally audio recorded. On a subsequent day (mean 53.07 days later, with no treatment provided in the interim), participants completed a battery of self-report measures, including the SI-R and DASS. Among HD participants, 21 were randomly selected for IRR analysis and 11 for TRR analysis using raters ($n = 5$) with qualifications identical to those of the interviewers who listened to audiotapes of the HRS-I and made their own ratings of each HRS-I item. Initial interviewers, but not subsequent raters, were aware of the participant's likely group (HD vs. HC). With 40% relative error, a sample of 10 or more is powered to detect 80% or higher agreement, as expected for TRR; and a sample of 17 or more is powered to detect 60% or higher agreement, as expected for IRR (Gwet, 2008).

Data Analytic Plan

Internal consistency was assessed using Cronbach's α . Inter-rater and test-retest reliability were calculated for the subsample of HD patients using intra-class correlation (ICC) coefficients.¹ Convergent and divergent validity were examined using Pearson's correlation coefficients between HRS-I items and SI-R subscales. Partial correlations (Spearman's ρ) were used to examine the relationship of HRS-I items to SI-R subscales, controlling for the other SI-R subscales, and for HRS-I items and the DASS total score, controlling for the SF-36 Emotional Role Functioning Score, and vice versa. Known-groups validity was examined using independent-samples t tests with effect sizes calculated using Cohen's d . Receiver operating characteristic (ROC) analyses were used to determine the optimal cut points for distinguishing the HD and HC groups (i.e., the scores with the smallest difference between sensitivity and specificity), and to identify the sensitivity and specificity of those scores.

Results

Sample Characteristics

As shown in Table 1, the sample was mainly female and non-Hispanic white, with a mean age of 53.8 years ($SD = 5.9$). The HD and HC groups did not differ significantly in gender, age, ethnicity, or race. As expected, the HD group scored significantly higher than did the HC group on all SI-R and DASS subscales, and significantly lower (more impairment) on the SF-36 Emotional Role Functioning subscale. Consistent with previous research (Frost, Steketee, & Tolin, 2011), over half of HD participants met DSM-5 criteria for a depressive disorder, and over one-third met criteria for an anxiety disorder.

¹As noted by Yen and Lo (2002), ICC is preferred over Pearson's product-moment correlations for reliability analyses because it is not appropriate to apply the correlation to a case of two measures from the same variable, and because Pearson's correlations cannot detect the existence of systematic error.

Internal Consistency and Reliability

Internal consistency of the HRS-I was high in the total sample ($\alpha = 0.87$). Inter-item correlations ranged from 0.31 (between the acquiring and impairment items) to 0.92 (between the clutter and difficulty discarding items).

As shown in Table 2, inter-rater reliability for HD participants was adequate for each item (ICC = 0.71–0.90) and good for the total score (ICC = 0.81). Test-retest reliability was also adequate for each item (ICC = 0.72–0.94) and good for the total score (ICC = 0.85).

Validity

As shown in Table 3 (top), HRS-I and SI-R items and total scores correlated strongly with one another (range = 0.74–0.94), although correlations among like items did not stand out against the background of high correlations among all items. Similarly, the HRS-I distress item correlated moderately highly with the DASS total score, and the HRS-I impairment item showed a moderately high correlation with the SF-36 Emotional Role Functioning Score, and again, all of the items were strongly intercorrelated.

A series of partial correlations (ρ) shown in Table 3 (bottom) aimed to clarify the specific relationship of HRS-I items to other relevant scales. The HRS-I clutter item correlated significantly with the SI-R clutter subscale, when controlling for the SI-R difficulty discarding and acquiring subscales. Similarly, the HRS-I acquiring item correlated significantly with the SI-R acquiring subscale when controlling for the SI-R clutter and difficulty discarding subscales. The HRS-I difficulty discarding item, however, correlated significantly and equally with the SI-R difficulty discarding ($\rho = 0.42$) and clutter ($\rho = 0.48$) subscales. In general, these partial correlations showed that HRS-I clutter, difficulty discarding, and acquiring items correlated most strongly with their counterpart subscales on the SI-R, although this was less pronounced for difficulty discarding. The HRS-I distress item correlated significantly and most strongly with the DASS total score. However, all HRS-I items showed strong partial correlations with SF-36 Emotional Role Functioning when controlling for DASS total, while the HRS-I functioning item did not show the expected stronger relationship.

Table 4 shows scores on each HRS-I item, and the HRS-I total score, for the HD and HC groups. As expected, the HD group scored significantly higher on all items than did the HC group, with large effect sizes ($d = 1.28$ to 6.58). ROC analysis (see Table 5) showed excellent sensitivity and specificity for distinguishing the HD and HC groups; the optimal cutoff total score was 11.

Discussion

In the original psychometric examination of the HRS-I, Tolin et al. (2010) used correlations (r) across time and context together. In a combined HD/HC sample, results suggested high reliability. The present study examined test-retest reliability without the confound of context, using ICC rather than r , and similarly found evidence of strong test-retest reliability. The present study added an analysis of inter-rater reliability to Tolin et al.'s prior report, with

results indicating good consistency across raters. Together, these results suggest that the HRS-I is reliable for clinical and research use.

The establishment of a clear test-retest reliability statistic also facilitates calculation of reliable change indices (Jacobson & Truax, 1991). The reliable change index (a change that is unlikely to be attributable to chance) is calculated as a difference greater than $1.96(SD_1\sqrt{2}\sqrt{1-r})$, where SD_1 is the standard deviation of the measure, and r is the reliability of the measure (Evans, Margison, & Barkham, 1998). Using the test-retest reliability and standard deviation from the present study, reliable change would be present with a decrease of 7.13 points or more for the HRS-I total score. Clinically significant change, commonly considered to reflect clinical remission, can be defined as (A) a post-treatment score that is 2 or more SD from the clinical mean, (B) within 2 SD of the normative mean, or (C) more likely to come from the normative mean than the clinical mean (McGlinchey, Atkins, & Jacobson, 2002). Using the current data, then, criterion A would be met if a participant's post-treatment score was lower than or equal to 14.45. Criterion B would be met if a participant's post-treatment score was lower than or equal to 1.92. Criterion C is calculated as $\frac{(M_{clin} \times SD_{norm}) + (M_{norm} \times SD_{clin})}{SD_{norm} + SD_{clin}}$, where *norm* and *clin* refer to the normative and clinical groups, respectively (Evans et al., 1998); using this formula, Criterion C would be met if a participant's post-treatment score was lower than or equal to 3.24. Clearly, these three criteria differ markedly in terms of stringency, and will need to be selected carefully based on the aims of the study.

Like Tolin et al. (2010), we examined convergent validity by comparing participants' scores on the HRS-I with those on the SI-R. Correlating matching HRS-I items with SI-R subscales for clutter, difficulty discarding, acquiring, and total scores, Tolin et al. found correlations (r) ranging from 0.76–0.92. The present results were comparable, with correlations (ICC) ranging from 0.81–0.94. The present results add to those of Tolin et al. by demonstrating that the matching correlations remain significant even when controlling for the other SI-R subscales. This helps clarify that the relationships between the HRS-I and SI-R scores are domain-specific and cannot be attributed solely to a global hoarding severity dimension. One exception was that the HRS-I difficulty discarding item was also related strongly to the degree of self-reported clutter in the home as measured by the SI-R. The validity of the HRS-I distress item was evident in its significant association with negative affect, even when controlling for impairment on the SF-36. While the HRS-I impairment item was significantly associated with impairment in emotional role functioning even when controlling for negative affect on the DASS, the latter measure was also associated moderately with all other HRS-I items, suggesting that all HD symptoms are associated with emotional role impairment.

We note that, compared to the samples in Tolin et al. (2010), our HD sample had significantly higher HRS-I scores [27.73 (6.64) vs. 24.22 (5.67), $t = 3.52$, $p < 0.001$], and our HC sample showed significantly lower HRS-I scores than did that of Tolin et al. (2010) [0.36 (0.78) vs. 3.34 (4.97), $t = 5.39$, $p < 0.001$], despite comparable inclusion criteria (absence of lifetime psychiatric disorder or psychiatric treatment). This may suggest that the

new anchor points and follow-up questions added to the HRS-I (see Appendix) enhance separation between the groups, including reducing false positive ratings in HCs. The lower scores in the present HC sample are responsible for the lower cutoff score in the present study (11) compared to that of Tolin et al. (14). The lower cutoff score in the present study does not appear attributable to Tolin et al.'s inclusion of obsessive-compulsive disorder (OCD) patients in the ROC analysis, as the OCD group scored even lower on the HRS-I than did the HCs in that study. The cut score of 11 showed excellent sensitivity and specificity, making it a useful threshold for determining the presence of clinically significant change.

A significant limitation of the present study is our exclusive use of treatment-seeking HD patients. Many individuals with HD do not seek treatment voluntarily (Frost, Tolin, & Maltby, 2010), and it is not clear how inclusion of these (potentially more severe) cases would have altered the results. The requirement of at least moderate HD severity in the HD group, and the exclusion of prospective HC participants with a history of psychiatric illness, is both a strength and a limitation. Its strength lies in helping create distinct groups with known characteristics, but its limitation lies in omitting the middle of the HD continuum (Timpano et al., 2013), thus potentially inflating sensitivity and specificity estimates. Our use of audio recordings of the HRS-I interviews ensured that raters were rating the same responses, but eliminated the potentially important variability in participant responses over time. The lack of in-home measurement, as well as the lack of a visual measure of clutter severity such as the Clutter Image Rating (Frost, Steketee, Tolin, & Renaud, 2008) also limits the confidence that can be placed on the accuracy of self-reports of clutter. The long delay between interview and self-report measures is a further limitation, and may have lessened relationships among the measures. Finally, our sample was primarily female and white. Though the prevalence of HD by race and ethnicity is not clear, epidemiologic research suggests that men are at least as likely as women to meet criteria for HD (Samuels et al., 2008; Timpano et al., 2011), and this discrepancy could have affected the results.

The present results confirm that the HRS-I is reliable and valid for use with clinical populations of HD patients, and suggests that it is suitable for both clinical and research applications. The amendments to the HRS-I, which include more descriptive anchor points and the addition of optional follow-up questions, appear to have improved the already-strong sensitivity and specificity of this measure, and may have decreased false positive ratings among non-clinical respondents.

Acknowledgments

Funded by NIMH grant R01 MH101163. Clinicaltrials.gov identifier NCT01956344.

The authors thank Marla Genova, James Ransom, Amber Billingsley, Akanksha Das, Benjamin Katz, Bethany Wootton, and Lauren Hallion for their assistance.

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5. Washington, DC: 2013.

- Brown TA, Chorpita BF, Korotitsch W, Barlow DH. 1997; Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy*. 35:79–89. DOI: 10.1016/S0005-7967(96)00068-X [PubMed: 9009048]
- Evans C, Margison F, Barkham M. 1998; The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence Based Mental Health*. 1(3):70–72.
- Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G, Salkovskis PM. 2002; The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment*. 14(4):485–496. DOI: 10.1037/1040-3590.14.4.485 [PubMed: 12501574]
- Frost RO, Steketee G, Grisham J. 2004; Measurement of compulsive hoarding: Saving Inventory-Revised. *Behaviour Research and Therapy*. 42(10):1163–1182. DOI: 10.1016/j.brat.2003.07.006 [PubMed: 15350856]
- Frost RO, Steketee G, Tolin DF. 2011; Comorbidity in hoarding disorder. *Depression and Anxiety*. 28(10):876–884. DOI: 10.1002/da.20861 [PubMed: 21770000]
- Frost RO, Steketee G, Tolin DF, Renaud S. 2008; Development and validation of the Clutter Image Rating. *Journal of Psychopathology and Behavioral Assessment*. 32:401–417.
- Frost RO, Tolin DF, Maltby N. 2010; Insight-related challenges in the treatment of hoarding. *Cognitive and Behavioral Practice*. 17:404–413.
- Guy, W. *Assessment manual for psychopharmacology*. Washington, DC: U.S. Government Printing Office; 1976.
- Gwet KL. 2008; Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*. 61(Pt 1):29–48. DOI: 10.1348/000711006X126600 [PubMed: 18482474]
- Jacobson NS, Truax P. 1991; Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*. 59(1):12–19. DOI: 10.1037/0022-006X.59.1.12 [PubMed: 2002127]
- Lovibond, S, Lovibond, P. *Manual for the Depression Anxiety Stress Scales*. Sydney: The Psychology Foundation of Australia: Inc; 1995.
- McGlinchey JB, Atkins DC, Jacobson NS. 2002; Clinical significance methods: Which one to use and how useful are they? *Behavior Therapy*. 33:529–550. DOI: 10.1016/S0005-7894(02)80015-6
- Park JM, Samuels JF, Grados MA, Riddle MA, Bienvenu OJ, Goes FS, ... Geller DA. 2016; ADHD and executive functioning deficits in OCD youths who hoard. *Journal of Psychiatric Research*. 82:141–148. DOI: 10.1016/j.jpsychires.2016.07.024 [PubMed: 27501140]
- Samuels JF, Bienvenu OJ, Grados MA, Cullen B, Riddle MA, Liang KY, ... Nestadt G. 2008; Prevalence and correlates of hoarding behavior in a community-based sample. *Behaviour Research and Therapy*. 46(7):836–844. [PubMed: 18495084]
- Steketee G, Frost RO, Tolin DF, Rasmussen J, Brown TA. 2010; Waitlist-controlled trial of cognitive behavior therapy for hoarding disorder. *Depression and Anxiety*. 27(5):476–484. DOI: 10.1002/da.20673 [PubMed: 20336804]
- Timpano KR, Broman-Fulks JJ, Glaesmer H, Exner C, Rief W, Olatunji BO, ... Schmidt NB. 2013; A taxometric exploration of the latent structure of hoarding. *Psychological Assessment*. 25(1):194–203. DOI: 10.1037/a0029966 [PubMed: 22984803]
- Timpano KR, Exner C, Glaesmer H, Rief W, Keshaviah A, Braehler E, Wilhelm S. 2011; The epidemiology of the proposed DSM-5 hoarding disorder: exploration of the acquisition specifier, associated features, and distress. *Journal of Clinical Psychiatry*. 72(6):780–786. DOI: 10.4088/JCP.10m06380 [PubMed: 21733479]
- Tolin DF, Frost RO, Steketee G. 2010; A brief interview for assessing compulsive hoarding: the Hoarding Rating Scale-Interview. *Psychiatry Research*. 178(1):147–152. DOI: 10.1016/j.psychres.2009.05.001 [PubMed: 20452042]
- Tolin DF, Gilliam C, Wootton BM, Bowe W, Bragdon LB, Davis E, ... Hallion LS. 2016; Psychometric Properties of a Structured Diagnostic Interview for DSM-5 Anxiety, Mood, and Obsessive-Compulsive and Related Disorders. *Assessment*. doi: 10.1177/1073191116638410
- Ware, JE. *SF-36 Health Survey Manual and Interpretation Guide*. Boston: The Health Institute, New England Medical Center; 1993.

- Wootton BM, Diefenbach GJ, Bragdon LB, Steketee G, Frost RO, Tolin DF. 2015; A contemporary psychometric evaluation of the Obsessive Compulsive Inventory-Revised (OCI-R). *Psychological Assessment*. 27(3):874–882. DOI: 10.1037/pas0000075 [PubMed: 25664634]
- Yen M, Lo LH. 2002; Examining test-retest reliability: an intra-class correlation approach. *Nursing Research*. 51(1):59–62. [PubMed: 11822570]

Appendix. Hoarding Rating Scale – Interview

Instructions for interviewers

- The HRS-I is designed to be a semi-structured interview. The initial question should be asked, and then additional supplemental questions (including, but not necessarily limited to, those listed below the initial questions) should be asked in order to arrive at a rating.
- The numeric scales and anchors should not be read to the interviewee. The HRS-I ratings should reflect the judgment of the interviewer, based on all of the available information.
- When two different rating descriptions seem to apply, choose the higher of the two.

I'd like to get a sense of how much you have been affected by saving, acquiring, and clutter over the past week.

Over the past week...

1. Because of the clutter or number of possessions, how difficult is it to use the rooms in your home?

Supplemental questions:

- a. Could you imagine that we're walking through your home together? As we go into each room, what would we see?
- b. How much does clutter interfere with your ability to do things like cook and eat in the kitchen, sleep in the bedroom, sit in the living area, or use the bathroom for toileting and bathing?
- c. How much does clutter interfere with your ability to move easily through all of the living spaces?
- d. Are there any safety hazards in the home, for example, risk of fire, risk of falling, blocked exits, cluttered stairways, or conditions that would prevent emergency workers from entering?

0	1	2	3	4	5	6	7	8
No problem		Mild, a few (e.g., 25%) of the living spaces are unusable or		Moderate, some (e.g., 25–50%) of the living spaces are unusable or		Severe, most (e.g., 75%) of the living spaces are unusable or		Extreme, nearly all of the living spaces are unusable or

0	1	2	3	4	5	6	7	8
		unsafe, but most spaces are usable		unsafe for use		unsafe for use		unsafe for use

2. To what extent do you have difficulty discarding (or recycling, selling, giving away) ordinary things that other people would get rid of?

Supplemental questions:

- a. How often do you try to discard things?
- b. When you try to discard things, how hard is it? How much discomfort do you feel?
- c. Do you avoid discarding things? Why is that? What kinds of things do you avoid discarding, and what kinds of things do you not avoid? How hard would it be to discard the things you have been avoiding?

0	1	2	3	4	5	6	7	8
No problem		Mild, feels mildly distressed by discarding or avoids discarding some things (e.g., < 25%) because of distress		Moderate, feels moderately distressed by discarding or avoids discarding some things (e.g., 50%) because of distress		Severe, feels strongly distressed by discarding or avoids discarding most things (e.g., 75%) because of distress		Extreme, feels extremely distressed by discarding or avoids discarding altogether because of distress

3. To what extent do you currently have a problem with collecting free things or buying more things than you need, or can use, or can afford?

Supplemental questions:

- a. How often do you acquire things that you don't really need, can't use, or can't afford? Do you sometimes feel like your buying or collecting is out of control?
- b. Have you tried to resist the urge to acquire things? When you try to resist acquiring, how hard is it? How much discomfort do you feel?
- c. Do you often buy or pick up free things even though you intended not to?
- d. Do you have to avoid certain places because you would be unable to control your desire to acquire things? If so, what kinds of places do you have to avoid? If you were in that place, how hard would it be to resist the urge to acquire?

0	1	2	3	4	5	6	7	8
No problem		Mild, acquires a few items that are not needed or affordable or feels a slight loss of control		Moderate, acquires a moderate number of items that are not needed or affordable or feels a moderate loss of control		Severe, acquires a large number of items that are not needed or affordable or feels a strong loss of control		Extreme, acquires an extreme amount of items that are not needed or affordable or feels completely out of control

4. To what extent do you experience emotional distress because of clutter, difficulty discarding or problems with buying or acquiring things?

Supplemental questions:

- a. How often do you feel distressed by the condition of your home, or by your saving and acquiring?
- b. When you feel distressed by the condition of your home, or by your saving and acquiring, how strong is that distress? Can you manage it?
- c. When you feel distressed by the condition of your home, or by your saving and acquiring, how long does that distress? Does it last for a few minutes, or all day?
- d. Do you avoid certain activities or places because it would be too distressing? For example, do you stay away from your home, or certain places in your home, because of distressed feelings? If you were in that place, how distressed would you become?

0	1	2	3	4	5	6	7	8
No problem		Mild, occasionally feels distressed or feels mildly distressed but distress is brief and not severe or engages in minimal avoidance to manage distress		Moderate, regularly feels distressed or distress is moderately severe or distress lasts for more than a few minutes or engages in moderate avoidance to manage distress		Severe, frequently feels distressed or feels severely distressed with a noticeable intensity or distress lasts for more than an hour or engages in substantial avoidance to manage distress		Extreme, nearly constantly feels distressed or feels extremely distressed to point of being completely unable to cope or distress lasts for more than a few hours or engages in extreme avoidance to manage distress

5. To what extent do you experience impairment in your life (daily routine, job / school, social activities, family activities, financial difficulties) because of clutter, difficulty discarding, or problems with buying or acquiring things?

Supplemental questions:

- a. Do clutter, difficulty discarding, or acquiring affect your ability to work? How so?
- b. Do clutter, difficulty discarding, or acquiring affect your physical health? How so?
- c. Do clutter, difficulty discarding, or acquiring affect you financially? How so?
- d. Do clutter, difficulty discarding, or acquiring affect your relationships with neighbors? How so?
- e. Do clutter, difficulty discarding, or acquiring create legal problems for you? How so?
- f. Do clutter, difficulty discarding, or acquiring affect your social life? How so?
- g. Do clutter, difficulty discarding, or acquiring affect your relationships with family members? How so?
- h. Is the home in disrepair because of clutter? What are the specific problems?

0	1	2	3	4	5	6	7	8
No problem		Mild, slight impairment in work, social or family activities or slight financial impact but for the most part functioning is intact		Moderate, noticeable impairment in work, social or family activities or moderate financial impact or some areas of disrepair but many areas of functioning are intact		Severe, substantially reduced capacity to work and/or have good social or family activities, or significant financial problems due to hoarding or significant health consequences or problems with neighbors or the legal system or severe conditions of disrepair		Extreme, virtually unable to perform any work, virtually no social or family activities or major financial problems due to hoarding or home is not habitable or major legal or health consequences

Highlights

- Tested the psychometric properties of an expanded version of the *Hoarding Rating Scale* (HRS-I), a semistructured interview for hoarding disorder (HD).
- The HRS-I showed excellent internal consistency and good inter-rater reliability and test-retest reliability.
- The HRS-I correlated strongly with self-report measures of hoarding symptom severity.
- The HRS-I showed excellent sensitivity and specificity in distinguishing HD patients from healthy controls.

Table 1

Sample characteristics.

	HC	HD	χ^2	<i>t</i>
Female [N(%)]	33 (75.0%)	72 (82.8%)	1.11	
Age [M(SD)]	53.41 (7.21)	53.99 (9.24)		0.36
Hispanic [N(%)]	4 (9.1%)	3 (3.4%)	1.84	
Nonwhite [N(%)]	10 (22.7%)	8 (9.2%)	4.51	
SI-R Clutter [M(SD)]	2.47 (3.20)	25.84 (5.64)		25.39**
SI-R Difficulty Discarding [M(SD)]	3.57 (2.85)	19.94 (3.68)		25.68**
SI-R Acquiring [M(SD)]	2.80 (2.16)	15.67 (5.54)		14.81**
SI-R total [M(SD)]	8.84 (6.75)	61.46 (11.03)		28.86**
DASS-D [M(SD)]	0.98 (2.60)	9.71 (8.40)		6.71**
DASS-A [M(SD)]	0.82 (2.05)	5.84 (6.55)		4.95**
DASS-S [M(SD)]	2.66 (3.70)	13.20 (8.48)		7.85**
SF-36 Emotional Role Functioning [M(SD)]	98.55 (6.95)	51.02 (36.71)		6.14**
Comorbid depressive disorder [N(%)]	0 (0%)	45 (52.9%)		
Comorbid anxiety disorder [N(%)]	0 (0%)	31 (36.0%)		
Comorbid obsessive-compulsive disorder [N(%)]	0 (0%)	10 (11.5%)		

HC = Healthy control group. HD = Hoarding disorder group. SI-R = Saving Inventory-Revised. DASS-D = Depression Anxiety Stress Scales, Depression subscale. DASS-A = Depression Anxiety Stress Scales, Anxiety subscale. DASS-S = Depression Anxiety Stress Scales, Stress subscale. SF-36 = 36-Item Short Form Health Survey.

* $p < .05$.

** $p < .01$.

Table 2

Inter-rater reliability (intra-class coefficients), and test-retest reliability (intra-class coefficients) for scores on the Hoarding Rating Scale-Interview (HRS-I) in patients with hoarding disorder.

HRS-I Item	IRR	TRR
1. Clutter	0.71	0.88
2. Difficulty discarding	0.75	0.94
3. Acquiring	0.90	0.72
4. Distress	0.75	0.75
5. Impairment	0.82	0.80
Total score	0.81	0.85

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3
 First-order correlations between HRS-I items with measures of hoarding symptoms, mood, and functioning (top), and partial correlations between HRS-I items with measures of hoarding symptoms, mood, and functioning, controlling for other subscales (bottom).

HRS-I score	SLR score			DASS			SF-36		
	Clutter	Difficulty Discarding	Acquiring	Total	Total	Emotional Role Funct.	Total	Total	Emotional Role Funct.
1. Clutter	.92**	.85**	.74**	.90**	.51**	-.57**			
2. Difficulty Discarding	.91**	.90**	.78**	.92**	.52**	-.61**			
3. Acquiring	.79**	.78**	.81**	.84**	.49**	-.54**			
4. Distress	.90**	.89**	.79**	.92**	.60**	-.65**			
5. Impairment	.88**	.87**	.78**	.90**	.52**	-.54**			
Total	.93**	.91**	.82**	.94**	.56**	-.62**			

HRS-I score	SLR score			DASS			SF-36		
	Clutter (controlling for Difficulty Discarding and Acquiring)	Difficulty Discarding (controlling for Clutter and Acquiring)	Acquiring (controlling for Clutter and Difficulty Discarding)	Total (controlling for Clutter and Acquiring)	Total (controlling for SF-36 Emotional Role Functioning)	Emotional Role Funct. (controlling for DASS total)			
1. Clutter	0.67**	0.12	-0.01	0.15		-0.33**			
2. Difficulty Discarding	0.48**	0.42**	0.06	0.20		-0.34**			
3. Acquiring	0.27**	0.07	0.42**	0.12		-0.31**			
4. Distress	0.48**	0.35**	0.12	0.24*		-0.36**			
5. Impairment	0.45**	0.29**	0.12	0.21		-0.26**			

* $p < .05$.

** $p < .01$.

Table 4

Mean (SD) scores on the Hoarding Rating Scale-Interview (HRS-I) for hoarding disorder (HD) vs. healthy control (HC) participants.

HRS-I Item	HC	HD	<i>t</i>	<i>d</i>
1. Clutter	0.09 (0.29)	5.52 (1.39)	25.49**	4.75
2. Difficulty discarding	0.09 (0.29)	6.25 (1.13)	35.38**	6.58
3. Acquiring	0.11 (0.39)	4.48 (1.82)	15.68**	2.91
4. Distress	0.05 (0.21)	5.65 (1.33)	27.63**	5.13
5. Impairment	0.02 (0.15)	5.84 (5.57)	6.91**	1.28
Total	0.36 (0.78)	27.73 (6.64)	27.19**	5.05

HC = Healthy control. HD = hoarding disorder.

**
p < .001.

Table 5

Receiver operating characteristic (ROC) analysis of the Hoarding Rating Scale-Interview (HRS-I) to predict hoarding disorder diagnosis.

HRS-I Item	Optimal Cutoff			
	AUC	Score	Sensitivity	Specificity
1. Clutter	1.00 **	2	0.98	1.00
2. Difficulty discarding	1.00 **	3	1.00	1.00
3. Acquiring	0.98 **	2	0.95	0.98
4. Distress	1.00 **	2	1.00	1.00
5. Impairment	1.00 **	2	1.00	1.00
Total	1.00 **	11	1.00	1.00

AUC = area under the curve.

** $p < .001$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript