Journal of Obsessive-Compulsive and Related Disorders 16 (2018) 76–80

Contents lists available at ScienceDirect

Journal of Obsessive-Compulsive and Related Disorders

journal homepage: www.elsevier.com/locate/jocrd

Psychometric properties of the Hoarding Rating Scale-Interview

David F. Tolin[a,b,*], Christina M. Gilliam[a], Elizabeth Davis[a], Kristen Springer[a], Hannah C. Levy[a], Randy O. Frost[c], Gail Steketee[d], Michael C. Stevens[a,b]

[a] Institute of Living, Hartford, CT, United States
[b] Yale University School of Medicine, United States
[c] Smith College, United States
[d] Boston University School of Social Work, United States

ARTICLE INFO

Keywords:
Hoarding disorder
Assessment
Interview
Psychometrics

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 ($d = 1.28–6.58$). 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.

1. Introduction

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. Each item is rated on a 9-point scale from 0 to 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 to 0.94. The HRS-I correlated significantly with a self-report measure 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 to 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,

* Correspondence to: Institute of Living, 200 Retreat Avenue, Hartford, CT 06106, United States.
E-mail address: david.tolin@hhchealth.org (D.F. Tolin).

https://doi.org/10.1016/j.jocrd.2018.01.003
Received 30 November 2017; Received in revised form 17 January 2018; Accepted 18 January 2018
Available online 02 February 2018

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.

2. Method

2.1. 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 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).

2.2. 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 (κ = 0.86), very good test-retest reliability (κ = 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 = 0.98), difficulty discarding (α in present sample = 0.96), and acquiring (α in present sample = 0.94). 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 & 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 (α = 0.89 − 0.96) and good discriminant and divergent validity (Brown, Chorpita, Korotitsch, & Barlow, 1997); internal consistency was excellent in the present study (DASS-D, α = 0.95; DASS-A, α = 0.91; DASS-S, α = 0.95).

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 (α = 0.68).

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?"

2.3. 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