RE-EXAMINING THE PSYCHOMETRIC PROPERTIES OF THE MASSACHUSETTS GENERAL HOSPITAL HAIRPULLING SCALE

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Christopher Charles Bauer

In Partial Fulfillment
for the Degree of
MASTER OF SCIENCE

Major Program:
Psychology

August 2014

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Re-examining the Psychometric Properties of the Massachusetts General Hospital Hairpulling Scale

By

Christopher Charles Bauer

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Paul Rokke
Chair

James Council

Robert Gordon

Kevin Thompson

Approved:

08/29/14
Date

James Council
Department Chair

ABSTRACT

Trichotillomania is characterized by recurrent hair pulling resulting in hair loss and has been shown to have a significant impact on the health and social functioning of individuals with the disorder. Despite a growing presence in clinical and research settings, there are relatively few tools available to assess the severity of Trichotillomania, and data examining the reliability and validity of those measures are sparse. The goal of this study was to replicate and expand on previous examinations of the reliability and validity of the Massachusetts General Hospital Hairpulling Scale, a measure of Trichotillomania severity. Similar to previous studies, results showed acceptable internal consistency and provided evidence for divergent validity however, test-retest reliability was not acceptable and evidence for convergent validity was mixed. Implications of these results and suggestions for the future assessment of Trichotillomania severity are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

BAI.....Beck Anxiety Inventory

BDI.....Beck Depression Inventory

BDI-II.....Beck Depression Inventory 2nd Edition

CGI-S .....Clinical Global Inventory Severity

DSM.....Diagnostic and Statistical Manual of Mental Disorders

FOCI .....Florida Obsessive Compulsive Inventory

HRT.....Habit Reversal Therapy

MGH-HS.....Massachusetts General Hospital Hairpulling Scale

NAC .....n-acetylcysteine

NIMH-TS.....National Institute of Mental Health Trichotillomania Scale

OCD .....Obsessive Compulsive Disorder

PITS .....Psychiatric Institute Trichotillomania Scale

QOLI.....Quality of Life Inventory

STAI-T.....State-Trait Anxiety Inventory-Trait

TTM .....Trichotillomania

Y-BOCS.....Yale-Brown Obsessive Compulsive Scale

INTRODUCTION

Trichotillomania (TTM) is a disorder that has been associated with significant health complications and functional impairment (Woods et al., 2006). In order to better research, assess, and understand the disorder, the development of reliable and valid measures of severity is vital. A brief review of the disorder and summary of effective treatments is provided before assessment tools are discussed. The purpose of this paper is to examine the reliability and validity of a particular self-report measure of TTM severity.

Background

Trichotillomania (TTM) is currently classified as an Obsessive-Compulsive Spectrum Disorder in the Diagnostic and Statistical Manual of Mental Disorders (5th ed; DSM-5; American Psychiatric Association, 2013). TTM is defined as the recurrent pulling of one's own hair, causing hair loss. Hair can be pulled from any site on the body but is most commonly pulled from the scalp, eyelashes, and eyebrows (Woods et al., 2006). Pulling sites may vary over time and can occur in brief episodes throughout the day or across sustained episodes lasting several hours. Individuals may pull from localized areas, creating bald patches or, they may distribute their pulling, resulting in less noticeable hair loss.

Additional diagnostic criteria require individuals to have made repeated unsuccessful attempts to stop pulling (criteria B) and the pulling or resulting hair loss must cause clinically significant distress or impairment (criteria C). Finally, the hair pulling or hair loss must not be attributable to another medical condition (e.g. dermatologic) (criteria D) and must not be better explained by the symptoms of another mental disorder (criteria E), such as Obsessive Compulsive Disorder (OCD).

Previous versions of the DSM required that the patient report rising tension prior to pulling along with pleasure, relief, or gratification during or immediately after pulling (4th ed., text rev.; DSM-IV-TR; American Psychiatric Association, 2000). Although individuals with TTM commonly report these feelings, this finding has not been consistent across individuals who pull hair (Lochner et al., 2011; Lochner et al., 2012; du Toit, van Kradenburg, Niehaus, & Stein, 2001). As a result, debates over the usefulness of these criteria led to their removal in the most recent DSM (Lochner et al., 2011; Lochner et al., 2012). It has also been argued that the word "noticeable", used to describe hair loss in previous versions of the DSM, could be overly restrictive for individuals who distribute their pulling in an attempt to conceal hair loss (Lochner et al., 2012). As a result, the "noticeability" criterion was omitted from the DSM 5.

TTM occurs in approximately 0.6% - 3.4% of the population and affects more women than men by a ratio of roughly 9:1 (Christenson, Pyle, & Mitchell, 1991; Duke, Keeley, Geffken, & Storch, 2010). A number of studies have shown individuals with TTM express functional impairment. (Woods et al., 2006; Diefenbach, Tolin, Hannan, Crocetto, & Worhunsky 2005; Stemberger, Thomas, Mansueto, & Carter, 2000). For example, in an internet sample of 1697 self-reporting hair pullers, individuals reported significant interference in social (e.g. avoiding social activities, ability to maintain close relationships), occupational/academic (e.g. interference with job duties, difficulties performing school responsibilities), and psychosocial domains (e.g. elevated levels of depression and anxiety) (Woods et. al. 2006).

TTM varies with respect to how hair is pulled and the rituals individuals perform after pulling. Most people tend to pull one hair at a time using their fingers; however, tools such as needles or tweezers may also be used. There does not appear to be any meaningful distinction between those who pull one hair at a time versus multiple hairs at one time (Duke et al., 2010).

After the hair is removed, various behaviors involving the pulled hair can occur. Some individuals discard the hair immediately, while others manipulate the hair in some way. Individuals may save the pulled hair, rub the hair between their thumb and forefinger, run the hair across their lips, chew or swallow the hair, or perform various other rituals (Christenson & Mansueto, 1999).

Additionally, research has addressed styles of hair pulling (Flessner, Woods, Franklin, Keuthen, & Piacentini, 2008; du Toit et al., 2001; Christenson, Ristvedt, & Mackenzie, 1993). Two of these styles, automatic and focused, refer to an individual's attentiveness to their pulling behavior. Automatic pulling occurs outside the awareness of the individual and usually occurs during sedentary activities or while attention is on another task (e.g. watching television, reading, talking on the phone) (Christenson et al., 1993). Focused pulling is characterized by attention to the pulling behavior and is thought to be an attempt to regulate negative private experiences (Begotka, Wetterneck, & Woods, 2004; Diefenbach, Mouton-Odum, & Stanley, 2002). Research investigating the extent to which individuals experience one style of pulling or the other has been limited. Although virtually all individuals report some level of both hair pulling styles (Flessner et al., 2008), estimates suggest between 15-47% of individuals characterize their pulling as primarily focused, whereas 19% report equal levels of automatic and focused pulling (Christenson, Mackenzie, & Mitchell, 1991; du Toit et al. 2001). The recognition that distinct styles exist has also led to an understanding that different interventions may be needed (Franklin, Tolin, & Diefenbach, 2006).

Treatment studies for TTM have evaluated the efficacy of pharmacotherapy, psychotherapy, or a combination of both. In adult populations, most pharmacological studies have evaluated antidepressants (Clomipramine, Desipramine, Fluoxetine), but opioid antagonists

(Naltrexone), atypical antipsychotics (Olanzapine), and a glutamate modulator (n-acetylcysteine; NAC) have also been studied (Christenson et al. 1991; Grant, Odlaug, & Kim, 2009; O'Sullivan & Christenson, 1999; Streichenwein & Thornby, 1995; van Ameringen, Mancini, Patterson, Bennett, & Oakman, 2010;). A meta-analysis by Bloch and colleagues (2007) reported no significant differences between antidepressants and control conditions. In an unpublished double-blind, placebo-controlled trial, Naltrexone did show significant change in one of three outcome measures (O'Sullivan & Christenson, 1999). However, the most promising medications are Olanzapine and NAC, which have both shown significant improvement in TTM symptoms when compared to placebo controls (van Ameringen et al., 2010; Grant et al., 2009). Specifically, Olanzapine showed improvement in 85% of participants compared to 17% who received the placebo, while NAC reported 56% of participants were "much" to "very much improved" compared to 16% who received the placebo.

Although examination of pharmacological treatments has shown promise, the most effective treatments for TTM have used some form of habit reversal therapy (HRT) (Bloch et al., 2007). HRT involves teaching individuals to use replacement behaviors termed competing responses (e.g. holding hands in lap, clenching fists) when pulling occurs or is about to occur. HRT for TTM also includes self-monitoring, awareness training, and stimulus control techniques to help individuals recognize situations where they are likely to pull and implement strategies to reduce pulling (e.g. removing tweezers from the bathroom, covering mirrors). In two separate trials where HRT was compared to medication (Clomipramine and Fluoxetine), HRT was found to be significantly more effective in reducing symptoms of hair pulling (Ninan, Rothbaum, Marsteller, Knight, & Eccard, 2000; van Minen, Hoogduin, Keijsers, Hellenbrand, & Hendriks, 2003).

Regardless of modality, the foundation of effective treatment lies in the ability to define the problem and evaluate change. Given the myriad of ways individuals can present with TTM, it is important clinicians and researchers use a comprehensive set of assessment methods and measures to best guide treatment and report findings. Goals of assessment should include establishing a diagnosis, developing a functional analysis to inform treatment planning, and establishing symptom severity to evaluate treatment progress (Diefenbach, Tolin, Crocetto, Maltby, & Hannan, 2005). Although a number of measures and methods have been developed to assess these aspects of TTM (Diefenbach, Reitman, & Williamson, 2000; Elliott & Fuqua, 2000; Rothbaum, Opdyke, & Keuthen, 1999), this paper focuses on measures of severity, particularly of a widely used self-report measure, the Massachusetts General Hospital Hairpulling Scale (MGH-HS).

Assessment of TTM Severity

Various methods of assessing TTM severity have been developed, but generally fall into one of two categories; direct and indirect. Direct methods rely on self-monitoring and videotaping, both of which have limitations. Self-monitoring requires the individual to record time spent, number of pulling episodes, or number of hairs pulled (Azrin, Nunn, & Frantz, 1980). Videotaping requires individuals to record themselves during times where they are likely to pull, usually at home, and give the tapes to a clinician. Unfortunately, individuals may not accurately record how much time they spent pulling during each episode or when recalling the day. Frequent episodes of pulling, pulling in situations where immediate recording is difficult (e.g. in the car), and pulling that occurs without awareness could hinder an individual's ability to accurately record their own hair pulling. Likewise, videotaping allows clinicians and researchers to accurately assess frequency and duration of hair pulling, but is time consuming and

confounded by the fact that most individuals do not pull in the presence of others (Rapp, Miltenberger, Long, Elliott, & Lumley, 1998; Mackenzie, Ristvedt, Christenson, Lebow, & Mitchell, 1995).

Indirect methods of severity assessment include photographing affected areas, collecting pulled hairs, clinician-rated measures, and self-report measures. Photographing affected areas relies on being able to show hair loss or regrowth by taking the same picture throughout treatment. This method could be difficult due to lack of noticeable hair loss (i.e. distributed pulling) or uncomfortable due to the location of the affected area (i.e. pubic, chest, or leg regions). Collecting pulled hairs requires the individual to save pulled hairs, which are then counted or weighed. This method may be uncomfortable and the accuracy may be questionable (Winchel et al., 1992a). For example, there is no guarantee that collected hairs were actually pulled or that they were saved during each pulling episode.

Clinician-rated measures include the Psychiatric Institute Trichotillomania Scale (PITS; Winchel et al., 1992b) and the National Institute of Mental Health Trichotillomania Scale (NIMH-TS; Swedo et al., 1989). The PITS is a 6-item scale measuring hair pulling sites, duration, frequency, interference, distress, and severity of hair loss. Each question is endorsed on an eight-point scale, with higher scores indicating more severe symptoms. The NIMH-TS is a clinical interview modeled scale that assesses time spent (previous week and yesterday), resistance, distress, and interference related to hair pulling. Resistance is measured on a five point response scale (0-4) and the other dimensions are measured on a six point response scale (0-5). These 5 items are totaled with higher scores indicating more severe symptoms. Both of these scales have demonstrated acceptable inter-rater agreement, but neither has shown acceptable internal consistency (Diefenbach et al., 2005). Likewise, although clinician rated

scales tend to be informative, their use in clinical and research settings could be limited by the time required for administration.

Self-report measures are beneficial in that they can serve as brief, easily distributed assessments. Reliable and validated measures can be indispensable tools in both clinic and research settings. To date, there are only three self-report measures of TTM severity, one for children and two for adults. The two measures for adults are the Florida Obsessive Compulsive Inventory (FOCI; adapted for hair pulling; Storch et al., 2007) and the Massachusetts General Hospital Hairpulling Scale (MGH-HS; Keuthen et al., 1995). The FOCI is comprised of two scales. The first is a 20 item checklist asking individuals to confirm or deny the presence of various obsessions and compulsions. The second scale uses a 0 – 4 point likert scale to assess symptom severity based on loss of control, time spent, interference, distress, and avoidance. This scale was modified to better reflect hair pulling, replacing obsessions and compulsions with urges to pull and hair pulling. The MGH-HS is currently the most widely used self-report measure of TTM severity, despite mostly un-replicated reliability and validity statistics usually based on small sample sizes.

Although the original FOCI has been established as a quality assessment of obsessions and compulsions (Walser, 2013), information regarding the FOCI adapted for TTM is lacking. Currently, only one study has used the adapted FOCI and no psychometric properties have been reported (Lochner et al., 2012). Because the assessment was derived from an assessment of OCD, similar to many TTM assessments, it overlaps considerably with the MGH-HS. Due to the use of the MGH-HS as the most common outcome measure of TTM, it is important to further explore the psychometric properties of the MGH-HS.

The Massachusetts General Hospital Hairpulling Scale

Initial Development

The MGH-HS was developed out of need for a brief self-report measure to track changes in symptom severity throughout treatement (Keuthen et al., 1995). The original scale was based on the 10-item Yale-Brown Obsessive Compulsive Scale (Y-BOCS; Goodman et al., 1989) which is widely used to assess symptoms of OCD. The Y-BOCS was chosen because of the overlap between certain features of OCD and TTM (e.g. the presence of an urge prior to pulling and the repetitive nature of the pulling). The first half of the Y-BOCS assesses obsessions while the second half assesses compulsions. Urges were substituted for obsessions and three questions were developed to assess the frequency, intensity and perceived control over urges. Pulling behavior replaced the items assessing compulsions and 5 questions were developed assessing the frequency, efforts to resist, and actual control over pulling as well as social impact and associated distress of pulling. The original product was an eight item measure assessing the dominant features of hair pulling.

After examining the original eight items, it was found that the social impact item was poorly correlated to the total score and had the lowest mean, indicating it may not be eliciting the intended information (Keuthen et al., 1995). As previously described, research has shown that individuals with TTM report interference in social, occupational, and psychological domains (Woods et al., 2006). However, due to the item being unable to capture the intended facet of TTM, it was removed. As a result, the final MGH-HS consisted of seven items, each rated on a scale from 0 (no symptoms) to 4 (extreme symptoms). Individuals completing the instrument report the severity of their symptoms on each item over the previous week. The seven items are then summed to produce a score ranging from 0-28.

Psychometrics

Psychometrics is a term used to describe the statistical evaluation of psychological tests. More specifically, psychometrics is concerned with determining the usefulness of a given instrument based on core psychological concepts that include reliability and validity (Martin & McGlynn, 2013). The reliability of a test describes the ability of that test to produce stable and consistent results when the same or similar individuals are examined under the same or similar conditions. The validity of a test describes how well that test measures what it is intended to measure. A measure that is valid must also be reliable but a reliable measure does not have to be valid. For example, a scale that consistently weighs an object as 10 lbs. over its true weight is reliable in that the same result is achieved each time; however it is not valid in that it does not measure true weight. Some well-established analyses used in psychometric research include, but are not limited to, internal consistency, test-retest reliability, and construct validity.

Internal consistency describes the relationship between a single item of a test and the other items in that test and is used to determine how well the items measure an underlying construct. Most commonly reported as a Cronbach's alpha, scores can range from zero to one. Tests that use multiple items to produce a summary score should have high internal consistency. Guidelines for interpreting and reporting Cronbach's alpha usually suggest a 0.7 or above is "acceptable" for internal consistency. An alpha of .8 or above is considered "good" and .9 or above is considered "excellent." As alpha approaches 1.0 it may become undesirable because of redundancy between items. It has been suggested that these guidelines may not always be appropriate and interpretation of Cronbach's alpha should consider that the number of items on a test influences the statistic (Martin & McGlynn, 2013, Lance, Butts, & Michels, 2006).

Particularly, Lance et al. (2006) reviews the origins of cut off criteria for reliability and reports that a minimal acceptability of .8 should be used.

As well as being internally consistent, a test should have high test-retest reliability or be relatively stable over repeated administrations. Test-retest reliability can be influenced by the time between administrations, the relative stability of the construct being measured, and environmental factors. There is no standard for describing how much time should come between test administrations, but too short a time may be considered inadequate due to memory effects, while too long may appear to reflect inadequate reliability in the instrument, but may actually be capturing instability of what is being measured. In TTM, symptoms may appear stable day to day, but have greater fluctuation with more time between tests. Care should be taken in determining an appropriate amount of time to assess test-retest reliability.

Test-retest reliability is calculated by comparing the scores from the two test administrations and is typically reported as a Pearson correlation coefficient (r), which can range between zero and one. Similar to internal consistency, a value of .7 or above is considered "acceptable" reliability. A value of .8 or above is considered "good" and .9 or above is considered "excellent." Again, caution should be taken as the coefficient approaches 1.0 in that the test may be resistant to change, which is problematic for a measure designed to detect treatment-related change.

Construct validity refers to the ability of a test to measure what it intends to measure. Evidence for or against construct validity can be found by examining the convergent and divergent validity of a test. Convergent validity refers to how well two tests that propose to measure the same thing are related. Divergent validity refers to how well two tests that propose to measure different things are unrelated. Theoretically, two tests that propose to measure the

severity of a disorder should be highly correlated without being so closely related that they are redundant. Likewise, two tests that propose to measure different constructs should have a low correlation. Replication of convergent and divergent validity continually supports or discourages the construct validity of a test.

Despite being used in clinical and research settings there is little research available on the psychometric properties of the MGH-HS. In the original development article, 119 consecutive chronic hair pullers either receiving treatment or participating in research protocols were given the MGH-HS (Keuthen et al., 1995). Internal consistency showed a coefficient $\alpha = .89$. This paper also examined the factor structure of the MGH-HS and found a 1 factor structure to fit best, accounting for 93% of the scale score variance.

In a related article, 26 consecutive TTM patients were given the MGH-HS (O'Sullivan et al., 1995). Twenty-two of these patients were given the MGH-HS at the beginning and end of one of their clinical sessions, approximately one hour apart to examine test-retest reliability. All 26 patients were given the MGH-HS 2-4 weeks apart to measure sensitivity to change. Additional measures included the PITS and Clinical Global Impression (CGI; Guy, 1976; a 7-point clinician-rated summative response scale detailed later) used to assess convergent validity, as well as the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) and Beck Anxiety Inventory (BAI; Beck, Epstein, Brown, & Steer, 1988) used to assess divergent validity. Test-retest reliability of the MGH-HS at one hour was .97 while test-retest reliability was not reported for the 2-4 week administrations. Strong positive correlations were found with the PITS ($r = .63$) and the CGI ($r = .75$) providing evidence for convergent validity. The MGH-HS was showed weak, non-significant correlations with the BDI ($r = .30$) or the BAI ($r = .10$) providing evidence for divergent validity.

Diefenbach and colleagues attempted to replicate and expand research on the psychometrics of various hair pulling severity measures (Diefenbach et al., 2005). Twenty-eight individuals diagnosed with TTM were given a variety of measures as part of an intake to a clinical trial. Measures included the MGH-HS, PITS, NIMH-TS, CGI, Beck Depression Inventory 2nd Edition (BDI-II; Beck, Steer, & Brown, 1996), and State-Trait Anxiety Inventory-Trait (STAI-T; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). In this study the MGH-HS had an internal consistency of α = .80 and moderately correlated with the NIMH-TS (r = .52) and the PITS (r = .55). Similar to the O'Sullivan et al. (1995), Diefenbach et al. (2005) found a weak and non-significant correlation between the MGH-HS and the BDI-II (r = .26). However, the MGH-HS did correlate moderately with their measure of anxiety (r = .53).

Across these studies, the MGH-HS seems to exhibit good internal consistency (α = .80-.89). Although promising, there is little evidence for convergent or divergent validity. Furthermore, evidence that is available has been calculated from studies with relatively small sample sizes. Although small sample sizes are common in TTM literature, additional reports would add credibility to the validity of the measure. Finally, only one example can be found of test-retest reliability for the MGH-HS, which was measured over approximately one hour instead of the more traditional and practical time span of a full week as is commonly practiced in treatment. This study seeks to replicate and continue expanding the literature examining the psychometric properties of the MGH-HS in a large clinical sample of adults with TTM.

The goals of this study are to test the internal consistency, test-retest reliability, and provide evidence for convergent and divergent validity of the MGH-HS. There are four corresponding hypotheses in this study: 1) The internal consistency of the MGH-HS will be acceptable with a Cronbach's alpha above .80, 2) Test-retest reliability will be acceptable with

correlations between Screen and baseline and Screen and session 1 above .70, 3) the MGH-HS will correlate moderately to strongly with other measures of TTM severity (i.e. NIMH-TS and CGI), providing evidence for convergent validity, and 4) the MGH-HS will have a low correlation (e.g. $r < .40$) with measures of quality of life (i.e. Quality of Life Inventory; QOLI; Frisch, 1994), anxiety (BAI), and depression (BDI-II), providing evidence for divergent validity.

METHODS

Participants

Participants for this study came from a randomized controlled treatment trial examining the difference between two therapies for TTM in adults. The study was approved by the University of Wisconsin – Milwaukee Institutional Review Board and funded by the National Institute of Health. Entry into the study used a multi-gate process that included a brief phone survey, an initial in person screening assessment, and a baseline assessment before starting a 10-week treatment protocol. Integral to study entry was a diagnosis of TTM based on DSM-IV criteria during the initial screening assessment. Also, participants must have been between 18 and 65 years old, scored above 12 on the MGH-HS, had an estimated IQ above 85, been un-medicated or on stable medication for at least 8 weeks, and not been receiving psychotherapy for TTM or other psychiatric condition. Participants were excluded if they met criteria for Bipolar I, Psychotic Disorder, current substance dependence, or minimum severity for hair pulling was not met.

A total of 117 treatment seeking adults completed the initial screening assessment. Of those who completed the screening, 26 either did not qualify for the study or did not return for the baseline assessment. Individuals who did not qualify either met exclusionary criteria (e.g. bipolar disorder), fell below the minimum severity for hair pulling, or did not receive a diagnosis of TTM. A total of 91 adults (M_{age} = 35.0, age range: 18-61) completed the baseline assessment. The baseline sample consisted of 84 females and 7 males. Seventy-six (83.5%) of the participants identified themselves as Caucasian, 11 (12.1%) identified themselves as African American, one (1.1%) identified themselves as Asian, and three (3.3%) chose not to respond.

Following the baseline assessment, eight additional participants did not receive the first session of therapy. One participant was dropped because of treatment assignment, one participant was dropped due to a primary mood disorder, contact was lost for one participant, and five others did not return for session 1 for unknown reasons. A total of 83 participants (M_{age} = 35.5, age range: 18-61) completed the self-report measures provided at the beginning of the first treatment session. The session 1 sample consisted of 76 females and 7 males. Seventy (84.3%) of the participants identified themselves as Caucasian, 10 (12.0%) identified themselves as African American, and 3 (3.6%) chose not to respond. Additional demographic information is provided in Table 1.

Table 1: Demographic Information

	Baseline		Session 1	
	(N)	(%)	(N)	(%)
Comorbid Diagnoses				
Skin Picking[+]	12	15.2	11	14.9
GAD	12	13.2	11	13.3
Specific Phobia	6	6.6	5	6.0
OCD	5	5.5	3	3.6
Major Depressive Episode	4	4.4	3	3.6
Social Phobia	4	4.4	4	4.8
Alcohol Abuse	2	2.2	2	2.4
Adjustment Disorder	2	2.2	2	2.4
Agoraphobia	2	2.2	2	2.4
Dysthymic Disorder	2	2.2	1	1.2
Panic Disorder	1	1.1	1	1.2
PTSD	1	1.1	1	1.2
Bipolar II	1	1.1	1	1.2
Marital Status				
Never Married	47	51.6	43	51.8
Married	35	38.5	32	38.6
Separated	1	1.1	*N/A*	*N/A*
Divorced	8	8.8	8	9.6
Highest Education				
Partial High School	6	6.6	5	6.0
High School Graduate	9	9.9	8	9.6
Technical College	7	7.7	6	7.2
Partial College	33	36.3	30	36.1
College Graduate	27	29.7	26	31.3
Professional Degree	9	9.9	8	9.6

Note: GAD = Generalized Anxiety Disorder; OCD = Obsessive Compulsive Disorder; PTSD = Post Traumatic Stress Disorder.
Comorbid Diagnoses were obtained using the Structured Clinical Interview for DSM-IV-TR during the initial screening assessment with the exception of Skin Picking (see below). Only current comorbid diagnoses at the time of the initial screen are listed.
[+] Skin Picking was assessed during the baseline assessment. Data is missing from 12 participants that completed the baseline assessment.

Procedure

The MGH-HS was given during the initial screen as well as the baseline assessment and the beginning of each treatment session. During the baseline assessment, participants completed a number of other measures assessing severity of hair pulling, depressive symptoms, anxiety symptoms, and quality of life. The measures used for analysis are detailed below. Data from the 91 participants who completed baseline was used to examine single time point psychometrics of the MGH-HS (i.e. internal consistency, convergent validity, and divergent validity). Using data from the individuals who completed baseline ensures a diagnosis of TTM with the minimal severity level required to enter the study. Test-retest reliability used data from the 91 participants who completed both screen and baseline and data from the 82 participants who completed both the initial screen and session one of treatment.

Materials

Massachusetts General Hospital – Hairpulling Scale (MGH-HS)

The MGH-HS is a 7-item self-report scale used to assess the severity of hair pulling in the previous week. The items specifically measure the frequency, intensity, and perceived control over urges, as well as the frequency, resistance, perceived control, and distress over pulling behaviors. Each item is scored from 0 (no symptoms) to 4 (extreme symptoms) and the items are summed to produce an overall severity score ranging from 0-28.

National Institute of Mental Health – Trichotillomania Scale (NIMH-TS)

The NIMH-TS is comprised of 5 items administered in a clinical interview style. It was derived from the Y-BOCS and assesses time spent pulling in the past week, time spent pulling the previous day, resistance to pulling, distress, and impairment. Resistance is scored on a 1-5 scale while the other items are scored on a 0-5 scale, with higher scores indicating more severe

symptoms. Scores are summed to give a measure of total severity. Previous research has shown the NIMH to have moderate test-retest reliability (r = .70) and good inter-rater reliability (p = .88) (Franklin, Edson, Ledley, & Cahill, 2011). It was also significantly correlated with the PITS (r = .75) and the CGI (r = .63; described below) (Diefenbach et al. 2005).

Clinical Global Inventory – Severity (CGI-S)

The CGI is a 7-item clinician-rated instrument that uses a 7-point scale to assess symptom severity and improvement (CGI-S and CGI-I). A single rating is made for each item on a 1 - 7 scale. Severity scores range from 1 = normal, not at all ill to 7 = extremely ill. Scores on the CGI-S have been shown to be significantly correlated with the distress (r = .58) and impairment (r = .76) ratings on the NIMH-TS as well as the Interference (r = .67) and hair loss severity (r = .58) ratings on the PITS.

Beck Anxiety Inventory (BAI)

The BAI is a 21 item self-report measure that assesses symptoms of anxiety experienced in the past week (e.g. heart pounding or racing, unable to relax). Each item is scored on a 4-point scale where 0 = not at all, 1 = mildly, 2 = moderately, and 3 = severely. Scores are totaled and can range from 0-63. The BAI has shown high internal consistency (α = 0.92) and one-week test–retest reliability (r = 0.75 Beck et al., 1988).

Beck Depression Inventory – II (BDI-II)

The BDI-II is a 21 item self-report measure that assesses symptoms of depression experienced in the past two weeks. Each item is scored on a 0-3 summative scale with higher scores indicating more severe symptoms. Scores can range from 0-63. The BDI-II has shown high internal consistency (α = .92) and test-retest reliability (r = .96; Sprinkle et al. 2002).

Quality Of Life Inventory (QOLI)

The QOLI is a 32 item self-report questionnaire that assesses satisfaction with 16 domains of life (e.g. health, money, work). Questions are paired asking individuals to rate the importance (rated 0-2) and satisfaction (rated -3 - +3; no zero) with each domain. Scores for each domain are then calculated by multiplying the importance by the satisfaction to receive a weighted score. These weighted scores are added together and divided by 16 to receive a total raw score for quality of life. T-Scores are identified using the norms table in the manual. Internal consistency for the QOLI was acceptable at $\alpha = .79$ and it also showed acceptable convergent validity. Divergent validity was significant but low ($r = .25$) when compared to a social desirability scale.

RESULTS

Internal Consistency

The means, standard deviations, and Cronbach's alpha are presented in Table II. The internal consistency of the MGH-HS was shown to be in the acceptable and "good" range (α = 0.83). Only the deletion of item 5, resistance to pulling, would increase internal consistency, though not enough to be considered "excellent."

Table 2: Descriptive Statistics and Internal Consistency of the MGH-HS

	Mean (SD)	Cronbach's α	Cronbach's α if Item deleted
MGH-HS (Total)	16.98(4.65)	0.83	
Frequency of Urges	2.23 (1.01)		0.78
Intensity of Urges	2.37 (0.87)		0.78
Control of Urges	2.54 (0.95)		0.80
Frequency of Pulling	2.10 (0.96)		0.80
Resistance to Pulling	2.22 (0.89)		0.87
Control of Pulling	3.21 (0.78)		0.80
Distress of Pulling	2.31 (1.11)		0.82

Note: MGH-HS = Massachusetts General Hospital Hairpulling Scale

Test-Retest Reliability

The initial screen and baseline assessment were an average of 13.3 days apart (SD = 9.9, range: 3–85). The correlation between the MGH-HS completed at screen and baseline was $r(89) = .482$, $p < .01$. When the participant with the greatest timespan between screen and baseline was removed (i.e. 85 days), the average number of days changed to 12.5 (SD = 6.38, range: 3–39), but the correlation remained identical. The average time between the screen and session 1 was 22.7 days (SD = 9.2, range: 6–56). The correlation between screen and session one was $r(79) = .328$, $p < .01$.

Construct Validity

Convergent and divergent validity were examined using the measures described above in the baseline assessment. A Pearson correlation was conducted comparing the MGH-HS to the continuous measures (NIMH-TS, BAI, BDI-II, and QOLI) and a Spearman correlation was used to compare the MGH-HS to the CGI-S. Results from the Pearson correlation are presented in the left column of Table III. A Shapiro-Wilk test of normality on the continuous measures revealed that the NIMH-TS, the measure of anxiety (BAI), and the measure of depression (BDI-II) were not normally distributed, which violates an assumption of the Pearson correlation. To correct for the non-normal distribution, a Spearman correlation was conducted (Hauke, & Kossowski, 2011) and results are presented in the right column of Table III with the CGI-S.

Table 3: Correlational Analysis of Construct Validity

	Pearson Correlation	Spearman Correlation
	MGH-HS	MGH-HS
NIMH-TS	0.57**	0.57**
BAI	0.17	0.22*
BDI-II	0.30**	0.28**
QOL	-0.16	-
CGI-S	-	0.27*

Note: MGH-HS = Massachusetts General Hospital Hairpulling Scale; NIMH = National Institute of Mental Health Trichotillomania Severity Scale; BDI-II = Beck Depression Inventory 2nd Edition; BAI = Beck Anxiety Inventory; QOL = Quality of Life Inventory; CGI-S = Clinical Global Impression Scale.

*$p < .05$ (2-tailed)

**$p < .01$ (2-tailed)

Results showed a moderate and statistically significant correlation between the MGH-HS and the NIMH-TS and a low, yet statistically significant correlation between the MGH-HS and CGI-S. The MGH-HS showed low correlations with the measures of anxiety and quality of life. The correlation between the MGH-HS and the BDI-II was also low, but statistically significant.

DISCUSSION

This study sought to examine the psychometric properties of the MGH-HS. The MGH-HS was chosen due to its wide use as a severity measure of TTM in clinical and research settings and the relative paucity of evidence demonstrating acceptable reliability and validity. A strength of this study was the relatively large and well defined sample. Also, due to the multi-gated design required prior to receiving treatment, this study allowed for a unique view of fluctuations in TTM severity in the absence of an active intervention. Below, the hypotheses pertaining to each domain of psychometric properties is reviewed and relevant follow-up discussed.

Internal Consistency

It was predicted that the internal consistency of the MGH-HS would be acceptable with a Cronbach's alpha above .80. Similar to previous studies examining the MGH-HS, Cronbach's alpha indicated good internal consistency. The deletion of any single item of the MGH-HS would result in a Cronbach's alpha ranging from .78 to .87. This suggests that the seven items that make up the MGH are a cohesive set that measure the same underlying construct.

Test-Retest Reliability

The MGH-HS was hypothesized to demonstrate acceptable test-retest reliability (i.e. $r > .70$) for administrations from screen to baseline and screen to session one of treatment. The proposed design of the randomized controlled trial intended to have a one week gap between the initial screening assessment, baseline assessment, and first treatment session. Under ideal circumstances, this would have allowed for a one-week and two-week examination of test-retest reliability following the procedures laid out by this study. However, the time between assessments led to a two-week and three-week examination of test-retest reliability. Results from

this study indicate that the MGH-HS does not possess strong temporal stability at roughly two and three week periods.

The interpretation of test-retest reliability should take into account three factors that can influence scores over repeated administrations. These factors include environmental influences, time between administrations, and the relative stability of the construct being measured. With respect to environmental influences, the discussion of hair pulling may have increased participant's self-awareness to hair pulling symptoms between assessments. When completing the MGH-HS the first time, participants are asked to retrospectively consider hair pulling symptoms they may not have focused on previously. Lack of awareness may have resulted in over or underestimation of hair pulling symptoms when completing the MGH-HS the first time. Increased awareness between assessments could have led to differences in ratings even if symptoms were largely unchanged.

As for the time between administrations, the two week and three week latency likely reduced the test-retest reliability in comparison to earlier studies. O'Sullivan and colleagues showed a 0.97 correlation when administrations of the MGH-HS were one hour apart. The strong correlation found in their study may be misleading due to participants remembering their responses. Given the decrease in reliability over time, it is possible that there is an amount of time between administrations of the MGH-HS that would demonstrate acceptable test-retest reliability while minimizing memory effects.

Lastly, test-retest reliability assumes that the construct being measured is relatively stable in the absence of an active intervention. However, research examining fluctuations of TTM symptoms across various age groups has suggested that this is not the case (Flessner et al., 2009). Specifically, Flessner and colleagues showed that changes in hair pulling symptoms appeared to

coincide with increased psychological distress or hormonal changes. Although their study examined hair pulling symptoms in age groups across the lifespan, results suggest hair pulling symptoms could fluctuate with smaller scale changes in psychological distress or hormonal changes that occur on a daily, weekly, or monthly basis.

It is likely that each of the factors mentioned above had some influence that led to the demonstration of weak temporal stability of the MGH-HS. Unfortunately, weak temporal stability impacts the interpretation of treatment gains as measured by the MGH-HS. If the MGH-HS is determined to be unreliable, it becomes difficult to conclude that changes in scores are the result of an intervention or natural variation in hair pulling symptoms.

Future research would benefit from further investigation of the temporal stability of the MGH-HS. One suggestion is to examine test-retest reliability with one week between administrations. Not only is a one week time frame consistent with the completion instructions of the MGH-HS, it conforms to a more realistic timeframe between clinical sessions. Another suggestion is to conduct multiple assessments of TTM severity and analyze their test-retest reliability simultaneously. If the MGH-HS and another measure were to show similar changes in scores between administrations it would suggest that both measures are reacting to variations in the behavior. If the other measure was to demonstrate more stability than the MGH-HS, it would suggest that the MGH-HS may be an unreliable measure of TTM severity.

Construct Validity

It was hypothesized that results would provide evidence for convergent validity by showing moderate to strong correlations between the MGH-HS and other measures of TTM severity (i.e. NIMH-TS and CGI-S). Similarly, it was hypothesized that results would provide

evidence for divergent validity by showing low correlations between the MGH-HS and measures of anxiety, depression, and quality of life (i.e. BAI, BDI-II, QOLI).

Results from this study showed a moderate correlation between the MGH-HS and the NIMH-TS which is consistent with previous research (Diefenbach et al., 2005) and provides additional evidence for convergent validity. Inconsistent with previous research was the low correlation between the MGH-HS and the CGI-S which does not provide support for convergent validity. One reason for the low correlation between the MGH-HS and the CGI-S could have been the difference in experience with hair pulling symptoms between raters. That is, the clinician may have compared the participant's symptoms to other cases while the participant may only have themselves as a reference point. Another reason may be differences between questionnaires in what is being considered in regards to hair pulling symptoms. For example, the MGH-HS asks participants to rate how frequent and distressing hair pulling symptoms have been in the past week, while clinicians take into account distress, impairment, and noticeability to others when rating the CGI-S.

Examining the evidence for divergent validity, the MGH-HS showed low correlations with the BAI, BDI-II, and QOLI. Previous studies have shown similar correlations for measures of anxiety and depression (Diefenbach et al., 2005; O'Sullivan et al., 1995) and evidence for divergent validity is expanded further by the addition of the QOLI.

Given the low and moderate correlations between the MGH-HS and other measures of hair pulling severity, evidence for convergent validity is weak overall. However, evidence for divergent validity is strong. While it is clear that the MGH-HS is measuring a construct different than anxiety, depression, or quality of life, it is unclear how well it is capturing hair pulling

severity. Due to the results of this study and the relatively few examinations of construct validity, it is important that researchers continue to gather evidence for the validity of the MGH-HS.

Evaluation of the MGH-HS

One interesting result from this study is the continuity of the psychometric properties of the MGH-HS with previous findings. Particularly, this study reported the third examination of internal consistency which showed an alpha coefficient between the previously reported .80 and .89 (Diefenbach et al., 2005; Keuthen et al., 1995). Evidence for convergent validity between the MGH-HS and NIMH-TS was replicated (Diefenbach et al. 2005) and evidence for divergent validity between the MGH-HS and measures of anxiety and depression were consistent (Diefenbach et al. 2005; O'Sullivan et al. 1995). When considered together, the consistency of these results allows researchers and clinicians to be increasingly confident in how the MGH-HS relates to itself as well as the NIMH-TS and measures of anxiety and depression.

Conversely, the MGH-HS demonstrated poor temporal stability as well as a weak relationship to the CGI-S. The implication of poor temporal stability is the inability to report with confidence, that changes in measurement scores are reflective of an intervention and not due to other influencing factors. This causes problems for researchers and clinicians tracking changes over time and making inferences about the effectiveness of interventions. Additionally, the weak relationship between the MGH-HS and CGI-S raises questions about the validity of the MGH-HS even though this is the first study to show this weak relationship.

Given the results of this study, it is recommended that the MGH-HS continue to be used, but with caution. The recommendation for the continued use of the MGH-HS is due to the limited psychometric data available and lack of other measures. First, The MGH-HS is the only widely distributed, brief, self-report measure of hair pulling severity for adults with TTM.

Second, its continued use allows for a common relationship between past and future studies that use the MGH-HS. Lastly, this study is only the second report on the psychometric properties of the MGH-HS after the initial development articles.

Further examinations of the MGH-HS may find increased test-retest reliability over shorter time spans and stronger relationships to other measures of hair pulling severity (the two problematic results of this study). It is also recommended that future researchers and clinicians use the MGH-HS alongside at least one other measure of hair pulling severity to aid in the interpretation of changes on the MGH-HS. Researchers and clinicians should be aware that the MGH-HS is not a comprehensive questionnaire of all hair pulling symptoms and contains ratings of symptoms that some individuals may not have or may not be aware of (i.e. urges to pull hair).

Future Directions

Besides expanding on the psychometric properties of the MGH-HS, researchers and clinicians should seek to improve upon current measures or develop new measures. To start, one aspect of TTM absent from the MGH-HS is a measure of interference in social, occupational, and psychological functioning. In the original development of the MGH-HS, there was an item that attempted to measure interference, but was discarded because it did not appear to be measuring what was intended. One solution could be to reintroduce a broad question related to interference phrased differently than before or with a list of examples. Another solution could be to develop a series of questions about specific interfering consequences common to individuals with TTM.

Additionally, the MGH-HS questions individuals about their experience with urges to pull hair. In an effort to improve measures of TTM severity, changes in our view of the construct of TTM should be reflected in the questions asked about that construct. As discussed earlier, the

latest version of the DSM saw the removal of criteria B (an increasing sense of tension) and C (pleasure, relief, or gratification when pulling), which refer to urges, from the diagnosis of TTM. Due to the recent changes in the DSM 5, and reports that not all individuals with TTM describe having urges to pull their hair, questions about urges may be unnecessary for a measure of TTM severity.

Furthermore, many of the available assessment measures of TTM severity have been developed based on measures of Obsessive Compulsive Disorder (OCD). This was a logical approach given the overlap in presentation of OCD and TTM (i.e. obsessive thoughts surrounding hair pulling, perceived inability to avoid performing a behavior, repetitive nature of the disorder) and that TTM was once thought to be a variant of OCD (Tynes, White, & Steketee, 1990). However, researchers have noted significant differences between the two disorders (Keuthen, O'Sullivan, & Sprich-Buckminister, 1998) and questioned the utility of OCD-based measures of TTM severity (Stanley, Prather, Davis, Wagner, & Swann, 1993; Stanley, Breckenridge, Snyder, & Novy, 1999). In response, a measure could be created that takes into account the specifics of the disorder to gather a more complete picture of TTM severity.

To date, a number of suggestions have been made for the improved assessment of TTM severity. For example, Stanley et al. (1999) included suggestions for an "ideal" clinician-rated measure of TTM severity. It was proposed that, in addition to frequency and duration of hair pulling, an assessment of TTM severity should include measurements of situational variables (e.g. places hair pulling is likely to occur), affective states (e.g. individual's mood before, during, or after a pulling episode), and sensory stimuli related to hair pulling behavior (e.g. itching at the pulling site). Likewise, Diefenbach et al. (2006) proposed developing an assessment along two dimensions. The first dimension would assess pulling behaviors such as frequency, duration, and

resistance of pulling in the previous week, while the second would assess the impact of pulling such as distress, hair loss, and interference.

Another consideration towards the development of a comprehensive measure of TTM severity is the instability of hair pulling symptoms. Although TTM is a chronic disorder, there is high variability in its presentation. One approach to improve reliability of a repeated severity measurement would be to identify and separate those variables that are relatively constant from those that have higher variability. For example, the lengths to which individuals go to conceal hair loss or their perceived control over hair pulling may be slow to change or require direct intervention to change. In contrast, hair pulling duration and interference can change daily depending on situational variables. Changes in these two hypothetical subscales would provide information about short and long term fluctuations in hair pulling severity.

Summary

This study was designed to expand the literature pertaining to the psychometric properties of the MGH-HS. As stated earlier, psychometrics is concerned with the statistical evaluation of the usefulness of an assessment instrument. Evidence for the reliability of the MGH-HS has been mixed, both in this study and when combined with previous research. Although the MGH-HS showed acceptable internal consistency, it was not shown to be stable over repeated administrations in the absence of an active intervention. The weak temporal stability of the MGH-HS suggests TTM severity may be inherently unstable or the MGH-HS does not fully capture the chronic characteristics of the disorder. The largest concern of inadequate reliability is the interpretation of treatment gains based on changes in MGH-HS scores.

The validity of the MGH-HS was also examined and demonstrated acceptable divergent validity from measures of anxiety, depression, and quality of life. However, when compared to

other measures of TTM severity, the convergent validity of the MGH-HS was shown to be inadequate. This suggests the MGH-HS may not be fully capturing a complete picture of TTM severity. It is recommended that the MGH-HS continue to be utilized and examined; however, caution should be used when interpreting changes in scores on the MGH-HS in both research and clinical settings. A multi-assessment approach should be utilized when interpreting results from a clinical trial or making clinical decisions. Despite the shortcomings of the MGH-HS, it remains an easily administered and brief measure of hair pulling and is the only widely used self-report measure available for adults. Future researchers should also consider the development of a new self-report measure that takes into account the high variability of hair pulling symptoms.

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*, Fourth Edition, Text Revision (DSM-IV-TR). Washington, DC: American Psychiatric Association.

Azrin, N. H., Nunn, R. G., & Frantz, S. E. (1980). Treatment of hairpulling (trichotillomania): A comparative study of habit reversal and negative practice training. *Journal of Behavior Therapy and Experimental Psychiatry, 11*(1), 13-20.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory* (2nd ed.). San Antonio, TX: Psychological Corporation, Harcour, Brace.

Beck A. T., Ward C. H., Mendelson M., Mock J., & Erbaugh J. (1961) An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571.

Beck A. T., Epstein N., Brown G., Steer R.A. (1988) An inventory for measuring clinical anxiety: psychometric properties. *Journal of Consulting and Clinical Psychology, 56*, 893–897.

Begotka, A. M., Wetterneck, C. T., & Woods, D. W. (2004). The relationship between experiential avoidance and the severity of trichotillomania in a nonreferred sample. *Journal of Behavior Therapy and Experimental Psychiatry, 35*(1), 17-24.

Bloch, M. H., Landeros-Weisenberger, A., Dombrowski, P., Kelmendi, B., Wegner, R., Nudel, J., . . . Coric, V. (2007). Systematic review: Pharmacological and behavioral treatment for trichotillomania. *Biological Psychiatry, 62*(8), 839-846.

Christenson, G. A., Ristvedt, S. L., & Mackenzie, T. B. (1993). Identification of trichotillomania cue profiles. *Behaviour Research and Therapy, 31*(3), 315-320.

Christenson, G. A., Pyle, R. L., & Mitchell, J. E. (1991). Estimated lifetime prevalence of trichotillomania in college students.*Journal of Clinical Psychiatry, 52*(10), 415-417.

Christenson, G. A., & Mansueto, C. S. (1999). *Trichotillomania: Descriptive characteristics and phenomenology*. In D.J. Stein, G.A. Christenson, & Hollander (Eds.), *Trichotillomania* (pp. 1-41). Washington, DC: American Psychiatric Press.

Christenson, G. A., Mackenzie, T. B., & Mitchell, J. E. (1991). Characteristics of 60 adult hair pullers. *The American Journal of Psychiatry, 148*(3), 365-370.

Diefenbach, G. J., Tolin, D. F., Hannan, S., Crocetto, J., & Worhunsky, P. (2005). Trichotillomania: Impact on psychosocial functioning and quality of life. *Behaviour Research and Therapy, 43*(7), 869-884.

Diefenbach, G. J., Tolin, D. F., Crocetto, J., Maltby, N., & Hannan, S. (2005). Assessment of trichotillomania: A psychometric evaluation of hair-pulling scales. *Journal of Psychopathology and Behavioral Assessment, 27*(3), 169-178. doi:http://dx.doi.org/10.1007/s10862-005-0633-7

Diefenbach, G. J., Reitman, D., & Williamson, D. A. (2000). Trichotillomania: A challenge to research and practice. *Clinical Psychology Review, 20*(3), 289-309.

Diefenbach, G. J., Mouton-Odum, S., & Stanley, M. A. (2002). Affective correlates of trichotillomania. *Behaviour Research and Therapy, 40*(11), 1305-1315. doi:http://dx.doi.org/10.1016/S0005-7967(02)00006-2

du Toit, P.L., van Kradenburg, J., Niehaus, D. J., & Stein, D. J. (2001). Characteristics and phenomenology of hair-pulling: An exploration of subtypes. *Comprehensive Psychiatry, 42*(3), 247-256.

Duke, D. C., Keeley, M. L., Geffken, G. R., & Storch, E. A. (2010). Trichotillomania: A current review. *Clinical Psychology Review, 30*(2), 181-193. doi:http://dx.doi.org/10.1016/j.cpr.2009.10.008

Elliott, A. J., & Fuqua, R. W. (2000). Trichotillomania: Conceptualization, measurement, and treatment. *Behavior Therapy,31*(3), 529-545.

Flessner, C. A., Conelea, C. A., Woods, D. W., Franklin, M. E., Keuthen, N. J., & Cashin, S. E. (2008). Styles of pulling in trichotillomania: Exploring differences in symptom severity, phenomenology, and functional impact. *Behaviour Research and Therapy, 46*(3), 345-357. doi:http://dx.doi.org/10.1016/j.brat.2007.12.009

Flessner, C. A., Woods, D. W., Franklin, M. E., Keuthen, N. J., & Piacentini, J. (2008). Styles of pulling in youths with trichotillomania: Exploring differences in symptom severity, phenomenology, and comorbid psychiatric symptoms.*Behaviour Research and Therapy, 46*(9), 1055-1061. doi:http://dx.doi.org/10.1016/j.brat.2008.06.006

Franklin, M. E., Edson, A. L., Ledley, D. A., & Cahill, S. P. (2011). Behavior therapy for pediatric trichotillomania: A randomized controlled trial. *American Academy of Child and Adolescent Psychiatry.Journal, 50*(8), 763-771.

Franklin, M. E., Tolin, D. F., & Diefenbach, G. J. (2006). *Trichotillomania.* In E. Hollander & D. J. Stein (Eds.), *Clinical manual of impulse-control disorders* (pp. 149-173). Arlington, VA: American Psychiatric Publishing, Inc, 2006.

Frisch, M. B. (1994). *Quality of Life Inventory manual and treatment guide*. Minneapolis, MN: NCS Pearson and Pearson Assessments

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., . . . Charney, D. S. (1989). The yale-brown obsessive compulsive scale. I. development, use, and reliability. *Archives of General Psychiatry, 46*(11), 1006-1011.

Grant J. E., Odlaug B. L., Won Kim S. (2009) N-Acetylcysteine, a Glutamate Modulator, in the Treatment of Trichotillomania: A Double-blind, Placebo-Controlled Study.*Archives of General Psychiatry, 66*(7), 756-763. doi:10.1001/archgenpsychiatry.2009.60

Guy, W. (1976). *Assessment manual for psychopharmacology*. Washington, DC: U.S. Government Printing Office.

Hauke, J., & Kossowski, T. (2011). Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae, 30*(2), 87-93. doi:http://dx.doi.org/10.2478/v10117-011-0021-1

Keuthen, N. J., O'Sullivan, R. L., Ricciardi, J. N., Shera, D., Savage, C. R., Borgmann, A. S., . . . Baer, L. (1995). The massachusetts general hospital (MGH) hairpulling scale: I. development and factor analyses. *Psychotherapy and Psychosomatics, 64*(3-4), 141-145.

Keuthen, N. J., O'Sullivan, R. L., & Sprich-Buckminster, S. (1998). Trichotillomania: Current issues in conceptualization and treatment. *Psychotherapy and Psychosomatics, 67*(4-5), 202-213.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria. *Organizational Research Methods, 9*(2), 202-220. doi:http://dx.doi.org/10.1177/1094428105284919

Lochner, C., Stein, D. J., Woods, D., Pauls, D. L., Franklin, M. E., Loerke, E. H., & Keuthen, N. J. (2011). The validity of DSM-IV-TR criteria B and C of hair-pulling disorder (trichotillomania): Evidence from a clinical study. *Psychiatry Research,189*(2), 276-280. doi:http://dx.doi.org/10.1016/j.psychres.2011.07.022

Lochner, C., Grant, J. E., Odlaug, B. L., Woods, D. W., Keuthen, N. J., & Stein, D. J. (2012). DSM-5 FIELD SURVEY: HAIR-PULLING DISORDER (TRICHOTILLOMANIA). *Depression and Anxiety, 29*(12), 1025-1031. doi:http://dx.doi.org/10.1002/da.22011

Mackenzie, T. B., Ristvedt, S. L., Christenson, G. A., Lebow, A. S., & Mitchell, J. E. (1995). Identification of cues associated with compulsive, bulimic, and hair-pulling symptoms. *Journal of Behavior Therapy and Experimental Psychiatry, 26*(1), 9-16.

Martin, C. R., & Savage-McGlynn, E. (2013). A 'good practice' guide for the reporting of design and analysis for psychometric evaluation. *Journal of Reproductive and Infant Psychology, 31*(5), 449-455. doi:http://dx.doi.org/10.1080/02646838.2013.835036

Ninan, P. T., Rothbaum, B. O., Marsteller, F. A., Knight, B. T., & Eccard, M. B. (2000). A placebo-controlled trial of cognitive-behavioral therapy and clomipramine in trichotillomania. *Journal of Clinical Psychiatry, 61*(1), 47-50.

O'Sullivan, R. L., Keuthen, N. J., Hayday, C. F., Ricciardi, J. N., Buttolph, L., Jeni, M. A., & Baer, L. (1995). The massachusetts general hospital (MGH) hairpulling scale: II. reliability and validity. *Psychotherapy and Psychosomatics, 64*(3-4), 146-148.

O'Sullivan, R.L., & Christenson, G.A. (1999). *Pharmacotherapy of Trichotillomania*. In D. J. Stein, G. A. Christenson, & E. Hollander (Eds.), *Trichotillomania* (pp. 93-124). Washington, DC: American Psychiatric Press.

Rapp, J. T., Miltenberger, R. G., Long, E. S., Elliott, A. J., & Lumley, V. A. (1998). Simplified habit reversal treatment for chronic hair pulling in three adolescents: A clinical replication with direct observation. *Journal of Applied Behavior Analysis, 31*(2), 299-302.

Rothbaum, B.O., Opdyke, D.C., & Keuthen, N.J. (1999). *Assessment of Trichotillomania*. In D.J. Stein, G.A. Christenson, & Hollander (Eds.), *Trichotillomania* (pp. 285-298). Washington, DC: American Psychiatric Press.

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, R. E., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Sprinkle, S. D., Lurie, D., Insko, S. L., Atkinson, G., Jones, G. L., Logan, A. R., & Bissada, N. N. (2002). Criterion validity, severity cut scores, and test-retest reliability of the beck depression inventory-II in a university counseling center sample. *Journal of Counseling Psychology, 49*(3), 381-385.

Stanley, M. A., Prather, R. C., Davis, M. L., Wagner, A. L., & Swann, A. C. (1993). Can the yale-brown obsessive compulsive scale be used to assess trichotillomania? A preliminary report. *Behaviour Research and Therapy, 31*(2), 171.

Stanley, M. A., Breckenridge, J. K., Snyder, A. G., & Novy, D. M. (1999). Clinician-rated measures of hair pulling: A preliminary psychometric evaluation. *Journal of Psychopathology and Behavioral Assessment, 21*(2), 157-170. doi:http://dx.doi.org/10.1023/A:1022160522908

Streichenwein, S. M., & Thornby, J. I. (1995). A long-term, double-blind, placebo-controlled crossover trial of the efficacy of fluoxetine for trichotillomania. *American Journal of Psychiatry, 152*(8), 1192-1196.

Stemberger, R. M., Thomas, A. M., Mansueto, C. S., & Carter, J. G. (2000). Personal toll of trichotillomania: Behavioral and interpersonal sequelae. *Journal of Anxiety Disorders, 14*(1), 97-104.

Storch, E. A., Kaufman, D. A. S., Bagner, D., Merlo, L. J., Shapira, N. A., Geffken, G. R., . . . Goodman, W. K. (2007). Florida obsessive-compulsive inventory: Development, reliability, and validity. *Journal of Clinical Psychology, 63*(9), 851-859.

Swedo, S. E., Leonard, H. L., Rapoport, J. L., Lenane, M. C., Goldberger, E. L., & Cheslow, D. L. (1989). A double-blind comparison of clomipramine and desipramine in the treatment of trichotillomania (hair pulling). *The New England Journal of Medicine, 321*(8), 497-501.

Tynes, L. L., White, K., & Steketee, G. S. (1990). Toward a new nosology of obsessive compulsive disorder. *Comprehensive Psychiatry, 31*(5), 465-480.

van Ameringen, M., Mancini, C., Patterson, B., Bennett, M., & Oakman, J. (2010). A randomized, double-blind, placebo-controlled trial of olanzapine in the treatment of trichotillomania. *Journal of Clinical Psychiatry, 71*(10), 1336-1343. doi:http://dx.doi.org/10.4088/JCP.09m05114gre

van Minnen, A., Hoogduin, K. A., Keijsers, G. P., Hellenbrand, I., & Hendriks, G. J. (2003). Treatment of Trichotillomania With Behavioral Therapy of Fluoxetine: A Randomized, Waiting-List Controlled Study. *Archives of General Psychiatry, 60*(5), 517-522. doi:10.1001/archpsyc.60.5.517

Walser, A. L. (2013). *Psychometric properties of the florida obsessive compulsive inventory and the schedule of compulsions, obsessions and pathological impulses.* (Order No. AAI3534950, *Dissertation Abstracts International: Section B: The Sciences and Engineering.* url=http://search.proquest.com/docview/1449311324?accountid=7082. (1449311324; 2013-99200-517).

Winchel, R. M., Jones, J. S., Molcho, A., Parsons, B., Stanley, B., & Stanley, M. (1992a). Rating the severity of trichotillomania: Methods and problems. *Psychopharmacology Bulletin, 28*(4), 457-462.

Winchel, R. M., Jones, J. S., Molcho, A., Parsons, B., Stanley, B., & Stanley, M. (1992b). The psychiatric institute trichotillomania scale (PITS). *Psychopharmacology Bulletin, 28*(4), 463-476.

Woods, D. W., Flessner, C. A., Franklin, M. E., Keuthen, N. J., Goodwin, R. D., Stein, D. J., & Walther, M. R. (2006). The trichotillomania impact project (TIP): Exploring phenomenology, functional impairment, and treatment utilization.*Journal of Clinical Psychiatry, 67*(12), 1877-1888. doi:http://dx.doi.org/10.4088/JCP.v67n1207