Journal of Psychopathology and Behavioral Assessment, Vol. 27, No. 3, September 2005 (© 2005)
DOI: 10.1007/s10862-005-0633-7

Assessment of Trichotillomania: A Psychometric Evaluation of Hair-Pulling Scales

Gretchen J. Diefenbach,[1,2] David F. Tolin,[1] Johanna Crocetto,[1] Nicholas Maltby,[1] and Scott Hannan[1]

Accepted July 3, 2004

Currently there is no "gold standard" treatment outcome measure for trichotillomania (TTM), a disorder characterized by repetitive hair pulling resulting in noticeable hair loss. The goal of the current study was to evaluate the psychometric properties of TTM measures of differing assessment methods: self-report, clinician-rated summary scales, and clinician-rated global severity scales. Data collected from 28 patients with TTM indicated mixed psychometric properties for current scales. Internal consistency was strong for self-report but not clinician-rated summary scales. One clinician-rated summary scale total and subjective ratings of hair loss demonstrated good interrater agreement. Although convergent validity was good within measurement type, self-report did not correlate with clinician-rated global severity scales, perhaps because of absence of hair loss severity assessment on the self-report measure. A multimethod assessment including one of each type of measure reviewed in this paper, along with self-monitoring and measures of hair loss severity, is recommended for a comprehensive best practice approach to TTM assessment.

KEY WORDS: trichotillomania; hair pulling; assessment; measurement.

Trichotillomania (TTM), characterized by repetitive hair pulling resulting in noticeable hair loss, has been gaining recognition as a common and distressing psychiatric disorder (Diefenbach, Reitman, & Williamson, 2000). Despite increased awareness of the clinical need presented by patients with TTM, research on the psychopathology and treatment of this disorder has progressed slowly. One of the obstacles impeding further development of the TTM literature, and in particular treatment outcome research, is the lack of well-validated measures for TTM assessment.

Goals of TTM assessment include establishing diagnoses, developing a functional analysis of hair pulling to inform treatment planning, and establishing a baseline severity of symptoms to evaluate treatment progress. The severity of hair pulling, as well as the functional and emotional impact of pulling, are important variables to include in outcome assessments of TTM treatments. Variables assessing severity of hair pulling include frequency (i.e., how often pulling occurs), duration (i.e., how long pulling episodes last), number of hairs pulled, and number of hair-pulling sites. In addition to actual hair-pulling episodes, the frequency, intensity, and resistance of hair-pulling urges can be included in assessment. Variables assessing the emotional and functional impact of hair pulling include physical damage resulting from pulling (i.e., hair loss severity), interference with daily functioning, and the distress associated with hair pulling.

To date a variety of methods have been developed for TTM assessment (see Diefenbach et al., 2000; Elliott & Fuqua, 2000; Rothbaum, Opdyke, & Keuthen, 1999 for reviews). Some methods of assessment involve direct observation (e.g., videotaping children at play and coding number of hair-pulling episodes) or response product measures (e.g., saving pulled hairs to be counted). The strength of these methods is the objective and concrete nature of the variables. However, the reliability and validity of these methods is nonetheless limited by factors such as behavioral reactivity, compromised generalizability, problems

[1]Anxiety Disorders Center, The Institute of Living, Hartford Hospital, Connecticut.
[2]To whom correspondence should be addressed at Anxiety Disorders Center, The Institute of Living, Hartford Hospital, 200 Retreat Avenue, Hartford, Connecticut; e-mail: gdiefen@harthosp. org.

with logistical practicality, and/or unreliable saving or return of response products.

Assessment of alopecia areas is particularly important, given that noticeable hair loss is a cardinal criterion of TTM. Subjective ratings of hair loss severity may be made within the context of clinician-rated scales (discussed later) or as social validity measures (Tarnowski, Rosen, McGrath, & Drabman, 1987). Objective measurements can be made for hair length, hair density, and area of hair loss (e.g., Barabasz, 1987; Bornstein & Rychtarik, 1978; Dahlquist & Kalfus, 1984; Winchel et al., 1992b). However, the long latency period and variability of hair regrowth may cloud true treatment progress. In addition, ratings of alopecia are complicated by multiple hair-pulling sites, pulling in pubic regions, or other areas, which are not easily assessed, and diffuse pulling styles (i.e., pulling strategically all over the head to avoid creating an area of alopecia). Nonetheless, hair loss severity is an important construct in TTM assessment.

Self-monitoring is another common method of hair-pulling assessment that can provide a wealth of information by recording the patterns of hair-pulling episodes. Self-monitoring data are essential to developing a functional analysis of hair pulling and behavioral treatment plan. However, these data are limited in treatment outcome assessments given the unreliable recording and therapeutic reactivity of the intervention itself. In addition, self-monitoring only assesses a limited range of variables suggested earlier in a comprehensive treatment outcome assessment.

Given the weaknesses of direct observation and self-monitoring, treatment outcome assessment traditionally rely on alternative methods, primarily self-report and clinician-rated scales. Although these "paper and pencil" measures depend on subjective ratings, the benefits include ability to assess a wider range of hair-pulling correlates (e.g., severity, interference, distress), ease of administration, control over frequency and timing of assessments, and decreased reliance on patients completing monitoring reliably. However, few studies have explored psychometric properties of TTM self-report and clinician-rated scales.

Two papers described the development and validation of the Massachusetts General Hospital Hairpulling Scale (MGH-HPS), the only widely used self-report measure of TTM (Keuthen et al., 1995; O'Sullivan et al., 1995). The MGH-HPS was modeled after the Yale–Brown Obsessive–Compulsive Scale (Goodman et al., 1989), with items revised to better reflect phenomenology of hair pulling. MGH-HPS items assess frequency, intensity, and control of hair-pulling urges; frequency, resistance, and control of hair-pulling behaviors; and distress associated with hair pulling. These items demonstrate good internal consistency and a single factor structure (Keuthen et al., 1995) in a large sample ($n = 119$) of chronic hair pullers. Administration with a smaller sample ($n = 26$) demonstrated good test–retest reliability (over 1 h), convergent validity with two clinician-rated scales, divergent validity with self-report measures of depression and anxiety, and sensitivity to change over time (over a 2–4 week period) (O'Sullivan et al., 1995). Thus, the MGH-HPS demonstrates strong psychometric properties. However, no measures of alopecia, and only two clinician-rated measures, were included for convergent validity with the MGH-HPS.

Clinician-rated scales for TTM include global ratings of severity as well as summary scales comprised of several items added together for a total score. Although clinician-rated scales are one of the most common and efficient means of assessing treatment outcome, only limited support has been provided for use of these scales with TTM. To date, only one study has explored the psychometric properties of clinician-rated scales, and findings provided mixed support (Stanley, Breckenridge, Snyder, & Novy, 1999). In this study, four clinician-rated scales were administered to a sample of 22 TTM patients. Generally, internal consistency of summary scales was poor, and some of the clinician-rated scales failed to show adequate interrater agreement. Concurrent validity with clinician-rated TTM severity, number of hairs pulled, and time pulling demonstrated mixed results. In addition, this study was limited by the absence of a standardized self-report measure, such as the MGH-HPS, in analyses of convergent validity.

The present study provides both a replication and extension of previous psychometric research on hair-pulling measures. Replication is particularly important given that previous validity research included small sample sizes. Small sample sizes are typical of TTM research, and although the current study includes a similarly sized sample, it nonetheless provides an opportunity to substantiate previous findings through replication. In addition, the current study extends previous research, which focused predominantly on validation of specific methods of TTM assessment, by evaluating the psychometric properties of several commonly used TTM measures of differing assessment methods: self-report, clinician-rated summary scales, and clinician-rated global severity scales. A comprehensive analysis such as conducted in the present study can support conclusions regarding best practice guidelines for TTM assessment. Additional goals of this paper were to clarify the TTM characteristics assessed by different scales and discuss directions for the improvement of TTM assessment measures.

METHOD

Participants

Participants were 28 individuals with a primary diagnosis of TTM. As is customary in TTM research, diagnosis was based on *DSM-IV* criteria (American Psychiatric Association, 1994) with or without the endorsement of increasing and decreasing tension associated with pulling (criteria B and C).[3] TTM diagnosis was assessed using the Trichotillomania Diagnostic Interview (Rothbaum & Ninan, 1994). The sample included 26 women (92.9%) and two men (7.1%), with a mean age of 38.64 ($SD = 12.17$). Twenty two (78.6%) of the participants identified themselves as Caucasian, four (14.3%) as African American, and two (7.1%) as Other ethnicity. Additional demographic information is presented in Table I. The Structured Clinical Interview for *DSM-IV* (First, Spitzer, Gibbon, & Williams, 1995) was administered to evaluate comorbid disorders. Comorbid diagnoses were as follows: 21.4% ($n = 6$) major depression or dysthymic disorder, 17.9% ($n = 5$) specific phobia, 10.7% ($n = 3$) pain disorder, 10.7% ($n = 3$) social phobia, and 3.6% ($n = 1$) each of posttraumatic stress disorder, hypochondriasis, panic disorder, and anxiety disorder NOS.

All clinical assessments and self-report measures were completed as part of an intake evaluation for a TTM treatment outcome study (Diefenbach, Tolin, Maltby, Hannan, & Crocetto, 2003). Participants were recruited through advertisements in local media, community referrals, and Internet resources. A primary diagnosis of TTM with minimum symptom duration of 6 months was required for inclusion. All participants were between the ages of 18–65 years. Exclusion criteria were lifetime bipolar disorder, lifetime pervasive developmental disorder, lifetime psychotic disorder, current attention-deficit disorder, current alcohol, or substance abuse (within past 3 months), mental retardation, and current active suicidal or homicidal ideation. Given that participants were recruited within the context of a treatment outcome study, individuals were also excluded if they were unwilling to discontinue current psychotherapy. For those taking medications, individuals were only enrolled when medication was stabilized for at least 2 months.[4]

Table I. Demographic Information

Marital status		
Single	12	(42.9%)
Married	12	(42.9%)
Cohabitating	4	(14.3%)
Employment		
Unemployed	4	(14.3%)
Part-time	7	(25.0%)
Full-time	15	(53.6%)
Student	2	(7.1%)
Education		
MA or MS	4	(14.3%)
Some graduate school	2	(7.1%)
BA or BS	8	(28.6%)
AA or some college	9	(32.1%)
High school graduate	3	(10.7%)
Some school	2	(7.1%)
Household income		
70,000 and up	9	(32.1%)
60,000–70,000	2	(7.1%)
50,000–60,000	1	(3.6%)
40,000–50,000	3	(10.7%)
30,000–40,000	1	(3.6%)
20,000–30,000	7	(25.0%)
10,000–20,000	1	(3.6%)
10,000 and less	1	(3.6%)
Data not available	3	(10.7%)

Measures

Three types of TTM measures were administered: self-report, clinician-rated summary scales, and clinician-rated global severity measures. The MGH-HPS (Keuthen et al., 1995) was the self-report measure administered. Clinician-rated summary scales are interviews that include several items added together to create a total score. Clinician-rated summary scales included the Psychiatric Institute Trichotillomania Scale (PITS; Winchel et al., 1992a) and the NIMH Trichotillomania Severity Scale (Swedo, Rapoport, Leonard, Lenane, & Cheslow). Clinician-rated global severity measures are one-item measures assessing the overall severity of the patient's presentation. Clinician-rated global severity measures included the NIMH Trichotillomania Impairment Scale (NIMH-TIS, Swedo et al., 1989), the Clinical Global Impression Severity Scale (CGI; Guy, 1976), and a global Alopecia rating. In addition, the Beck Depression Inventory-II (BDI-II, Beck, Steer, & Brown, 1996) and State-Trait Anxiety

[3] Endorsement of these criteria was not required as they have been found to exclude participants with clinically significant hair pulling (Christenson, MacKenzie, & Mitchell, 1991). In the current sample 53.6% ($n = 15$) endorsed full diagnostic criteria.

[4] Fifty-two participants were excluded at the phone screen: 23.1% for diagnostic considerations (e.g., comorbid substance abuse; TTM not primary), 21.2% for logistical problems or unwilling to make the time commitment, 17.2% due to treatment considerations (e.g., did not want group treatment; unwilling to discontinue other treatment), 13.5% due to age exclusion or calling for another person, and 25.0% for not returning phone calls. Five participants were excluded during intake assessments, all due to diagnostic considerations (e.g., TTM not primary, comorbid bipolar disorder).

Inventory—Trait (STAI-T, Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) were administered to assess depressive and anxiety symptoms, respectively. Each of these measures is described in detail later.

MGH Hairpulling Scale (MGH-HPS)

The MGH-HPS is a 7-item self-report instrument that assesses frequency, intensity, and control of hair-pulling urges; frequency, resistance, and control of hair-pulling behaviors; and distress associated with hair pulling. Items are assessed using a 5-point Likert scale, ranging from 0 to 4, with higher scores indicating more severe symptoms. Total scores are calculated by summing the responses from all seven items. The MGH-HPS has demonstrated strong internal consistency ($\alpha = .89$) and test-retest reliability ($r = .97$; Keuthen et al., 1995; O'Sullivan et al., 1995). The MGH-HPS has demonstrated acceptable convergent and divergent validity. It correlated positively with two clinician-rated measures of hair-pulling severity (r range $= .63$–$.75$, $p < .001$), and did not correlate with self-report measures of depression ($r = .30$, $p = ns$) or anxiety ($r = .10$, $p = ns$) [O'Sullivan et al., 1995]. The MGH-HPS is also sensitive to changes in hair pulling, correlating positively with changes on clinician-rated measures (r range $= .50$–$.83$, $p < .02$) [O'Sullivan et al., 1995].

Psychiatric Institute Trichotillomania Scale (PITS)

The PITS is a 6-item clinician-rated measure, which assesses the following symptoms: number of hair-pulling sites, duration of time spent pulling or thinking about pulling, frequency of resisting hair-pulling urges, interference, distress, and severity of hair loss. Items are rated on an 8-point Likert scale, ranging from 0 to 7, with higher scores indicating more severe symptoms. The total score is calculated by summing the responses of all six items. Previous research conducted by Stanley et al. (1999) found that the internal consistency for the total score is poor ($\alpha = .59$). In addition, although the PITS demonstrated strong interrater agreement for most items (sites $r = .55$; duration $r = .92$; resistance $r = .95$; interference $r = .93$; distress $r = .95$, severity $r = 1.00$), there was only moderate agreement for total scores ($r = .60$; Stanley et al., 1999). Evidence for convergent validity also was mixed. Clinician ratings of TTM severity correlated significantly with PITS assessment of interference ($r = .75$, $p < .001$), distress ($r = .63$, p <.01), hair loss severity ($r = .73$, $p < .001$), and total scores ($r = .64$, $p < .001$). However, items that assess number of pulling sites, duration, and resistance did not correlate with clinician ratings of TTM severity, and the number of hairs pulled did not correlate significantly with any of the PITS items.

NIMH Trichotillomania Severity Scale (NIMH-TSS)

The NIMH-TSS is a clinical interview that was derived from the Y-BOCS (Goodman et al., 1989). The NIMH-TSS is comprised of five items that assess time spent pulling in the past week, time spent pulling the previous day, resistance to pulling, distress, and interference. Resistance is rated on a scale ranging from 0 to 4, and the four other items are rated on a scale ranging from 0 to 5, with higher scores indicating greater symptom severity. The total severity score is calculated by summing the five items. The NIMH-TSS has demonstrated mixed psychometric properties in a previous investigation (Stanley et al., 1999). On the NIMH-TSS, items assessing duration of hair pulling (past week $r = .87$; yesterday $r = 1.00$), as well as the NIMH-TSS total score ($r = .85$) have demonstrated adequate interrater agreement. However, the internal consistency for the NIMH-TSS total score was inadequate (total score $\alpha = .63$). The convergent validity of the NIMH-TSS was assessed through correlations with global TTM severity, number of hairs pulled, and amount of time spent pulling. For the NIMH-TSS, the ratings of duration showed significant correlations with number of hairs pulled (past week $r = 63$; yesterday $r = .72$) and with time spent pulling (past week $r = .67$; yesterday $r = .78$). Total NIMH-TSS scores correlated only with the number of hairs pulled ($r = .58$). However, no items in the NIMH-TSS correlated with global ratings of TTM severity.

NIMH Trichotillomania Impairment Scale (NIMH-TIS)

The NIMH-TIS is a global TTM severity measure, which provides an impairment score based on the degree of hair loss, money and time expended in pulling or concealing hair loss, and patient's sense of control over the behavior. Ratings range from 0 = *no impairment*; 1–3 = *minimal impairment*; 4–6 = *mild impairment*; 7–10 = *moderate/severe impairment*. The NIMH-TIS score demonstrates adequate interrater reliability ($r = .71$), and correlates significantly with alternative clinician-ratings of global TTM severity ($r = .87$; Stanley et al., 1999).

Clinical Global Impression (CGI)

The CGI is a clinician-rated assessment of global illness including both TTM as well as comorbid psychopathology. CGI severity is assessed using a 7-point Likert scale, from 1 = *normal, not at all ill* to 7 = *extremely ill*. The CGI is a common measure of treatment outcome for many psychological disorders, including TTM (Ninan, Rothbaum, Marsteller, Knight, & Eccard, 2000; Tolin, Franklin, Diefenbach, & Gross, 2002). Previous research using the CGI has found interrater reliability of $r = .66$ for severity scores (Dahlke, Lohaus, & Gutzmann, 1992). Validity of the CGI is demonstrated by its strong positive correlations with clinician-rated panic and depression symptoms (Leon et al., 1993), and these correlations remain consistent throughout an 8-week treatment period, suggesting that the CGI scores are sensitive to change over time.

Alopecia Ratings

A clinician-rated measure of hair loss severity was administered using a 7-point Likert scale (1 = *no evidence of hair pulling* to 7 = *large bald spots that are difficult to conceal)*. This scale has been used in previous research, and has demonstrated good interrater reliability ($r = .87$) when rating photographs of alopecia areas (Tolin et al., 2002). In the current study, data for Alopecia ratings were missing for one participant who pulled solely from the pubic area.

Beck Depression Inventory-II (BDI-II)

The BDI-II is a widely used self-report scale that assesses the severity of depressive symptoms. Internal consistency is strong ($\alpha = .92$) and test-retest reliability is high ($r = .96$; Sprinkle et al., 2002). The BDI-II also correlates with the number of depressive symptoms endorsed in a clinical interview (Sprinkle et al., 2002).

State-Trait Anxiety Inventory-Trait (STAI-T)

The STAI-T is a self-report measure demonstrating strong internal consistency ($\alpha = .90$), and relatively good test-retest reliability (r range = 65–.86; Spielberger et al., 1983). Scores on the STAI-T also discriminate clinical anxiety patients from nonanxious individuals (Spielberger et al., 1983).

Procedure

Data were collected during intake evaluations for a TTM treatment outcome study comparing group cognitive-behavioral therapy with group supportive therapy (Diefenbach, Tolin, Maltby et al., 2003). Written informed consent was obtained, and self-report measures were completed prior to clinical interviews. Either a licensed psychologist or a postdoctoral fellow under supervision of a licensed psychologist completed diagnostic interviews and Alopecia ratings. Alopecia ratings completed by diagnostic interviewers were used for interrater reliability with ratings made by independent evaluators. Immediately following the diagnostic assessment, participants met with independent evaluators, who completed clinician-rated scales. Both independent evaluators were licensed psychologists, who had prior training and experience in administration of the clinician-rated scales. In addition, two calibration meetings (one prior to data collection and one following) were completed for training.

A subset of interviews ($n = 12$; 42.9% of interviews) was randomly selected to assess interrater agreement of clinician-rated summary scales. Random selection was completed until the target subsample was reached of tapes with adequate audio quality. To calculate interrater agreement, independent evaluators listened to audiotaped interviews completed by the alternate independent evaluator. Given that interrater agreement was completed using audiotapes, direct inspection of hair loss severity by the second independent evaluator was not possible. Thus, global severity measures (CGI; NIMH-TIS) relying in part on direct inspection of hair loss, were excluded from interrater analyses. In addition, given that the PITS item rating severity of hair loss relies on direct inspection of the alopecia area, the PITS item rating severity of hair loss was excluded from interrater analyses. Interrater agreement of PITS total score was recalculated excluding the PITS hair loss item. However, interrater agreement of the Alopecia rating was possible by comparing the independent evaluator rating with the rating of the diagnostic interviewer, and those results are presented here.

RESULTS

Descriptive Statistics and Internal Consistency

Means, standard deviations, and coefficient alphas are presented in Table II. The MGH-HPS demonstrated good internal consistency; however, internal consistency for clinician-rated summary scales was poor. Deletion of

Table II. Descriptive Statistics and Internal Consistency

	Mean (SD)	α	Item-remainder correlation	α without item
MGH-HPS	16.71 (4.78)	0.80		
Frequency-urges	2.35 (0.87)		0.55	0.77
Intensity-urges	2.39 (0.92)		0.73	0.74
Control-urges	2.57 (0.96)		0.64	0.76
Frequency-pulling	2.14 (0.93)		0.73	0.74
Resistance-pulling	1.96 (0.92)		0.10	0.84
Control-pulling	3.11 (0.96)		0.70	0.75
Distress	2.18 (1.42)		0.44	0.81
NIMH-TSS	12.45 (4.56)	0.65		
Duration, past week	2.93 (1.63)		0.45	0.58
Duration, yesterday	1.86 (1.48)		0.51	0.54
Resistance	2.00 (1.28)		0.39	0.61
Distress	3.38 (1.26)		0.37	0.62
Interference	2.29 (1.38)		0.31	0.64
PITS	25.89 (5.97)	0.60		
Sites	4.46 (1.91)		0.13	0.64
Duration	4.14 (1.94)		0.37	0.53
Resistance	5.39 (1.03)		0.31	0.57
Interference	3.07 (2.07)		0.57	0.42
Distress	4.04 (1.29)		0.35	0.55
Hair loss severity	4.79 (1.85)		0.33	0.55
CGI	4.64 (0.83)	—	—	—
NIMH-TIS	6.61 (1.37)	—	—	—
Alopecia rating	5.41 (1.58)	—	—	—

Note. MGH-HPS = Massachusetts General Hospital Hairpulling Scale; NIMH-TSS = National Institute of Mental Health Trichotillomania Severity Scale; PITS = Psychiatric Institute Trichotillomania Scale; CGI = Clinical Global Impression; NIMH-TIS = National Institute of Mental Health Trichotillomania Impairment Scale. CGI, NIMH-TIS, and Alopecia rating are single item scales, thus coefficient α data are not available

any individual item did not improve internal consistency to adequate levels.

Interrater Reliability

Interrater reliability data for item and total scores of the clinician-rated summary scales and the Alopecia rating are presented in Table III. Spearman rank correlation coefficients and paired t-tests were used to assess agreement. The criteria of Spearman coefficients of at least .70 along with nonsignificant t-test results were used to determine adequate interrater agreement. These criteria have been used in previous psychometric analyses of clinician-rated hair-pulling scales (Stanley et al., 1999). Results indicated that the Alopecia rating was stable across interviewers. In addition, the NIMH-TSS items assessing duration, resistance, and interference demonstrated adequate interrater agreement, however, the distress item and total score did not. With respect to the PITS, the duration, resistance, and interference items, as well as the total score, demonstrated good interrater agreement. However, the PITS items assessing hair-pulling sites and distress were unstable across raters.

Table III. Interrater Agreement Based on Spearman Rank Correlation Coefficients and Paired t-Tests

Scale/item	Spearman coefficient	t
NIMH-TSS		
Duration, past week	0.81***	0.32
Duration, yesterday	0.98***	1.00
Resistance	0.95***	1.48
Distress	0.56	1.86
Interference	0.85***	0.94
Total	0.92***	2.40*
PITS		
Sites	0.96***	2.35*
Duration	0.93***	1.00
Resistance	0.90***	1.00
Interference	0.81***	0.76
Distress	0.45	2.77*
Hair loss severity	—	—
Total (items 1–5)	0.90***	0.72
Alopecia rating	0.93***	0.30

Note. NIMH-TSS = National Institute of Mental Health Trichotillomania Severity Scale; PITS = Psychiatric Institute Trichotillomania Scale. For PITS and NIMH-TSS, $n = 12$; $df = 11$. For Alopecia rating, $n = 27$; $df = 26$.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Construct Validity

Total Score Correlations

Correlational analyses of self-report, clinician-rated summary scales total scores, and clinician-rated global severity measures are presented in Table IV. Given the large number of correlations in this analysis, an alpha correction of $p = .003$ (.05/15) was applied to interpretation of statistically significant results in order to protect against alpha inflation. Given the small sample size, correlations indicative of a moderate effect size ($\geq$.40) are also highlighted to enhance interpretation of these data. Results indicated moderately strong and statistically significant relationships among measures within the same type of assessment method. However, correlations between assessment method types were less consistent. In general, both types of clinician-rated measures (summary scales and global severity) demonstrated moderately strong relationships with each other, with most of these correlations reaching statistical significance. However, the Alopecia rating did not correlate significantly with clinician-rated summary scales (although the correlation with the

Table IV. Correlational Analyses of Self-Report, Clinician-Rated Summary Scales, and Clinician-Rated Global Severity Scales

	Self-report	Clinician-rated summary		Clinician-rated global severity		
	MGH-HPS	NIMH-TSS	PITS	CGI	NIMH-TIS	Alopecia rating
MGH-HPS						
NIMH-TSS	.52					
PITS	.55***	.75***				
CGI	.13	.63***	.63***			
NIMH-TIS	.19	.39	.64***	.69***		
Alopecia rating	.10	.25	.53	.55***	.77***	

Note. Correlations indicative of a moderate effect size ($r \geq .40$) are in italic. MGH-HPS = Massachusetts General Hospital Hairpulling Scale; NIMH-TSS = National Institute of Mental Health Trichotillomania Severity Scale; PITS = Psychiatric Institute Trichotillomania Scale; CGI = Clinical Global Impression; NIMH-TIS = National Institute of Mental Health Trichotillomania Impairment Scale.
*** $p < .003$.

PITS approached significance, $p < .004$), and the NIMH-TSS did not correlate significantly with the NIMH-TIS. In addition, the MGH-HPS correlated moderately with clinician-rated summary scales, with the PITS correlation reaching statistical significance and the NIMH-TSS correlation approaching statistical significance ($p < .005$). In contrast, the MGH-HPS demonstrated low and nonsignficant correlations with all clinician-rated global severity measures.[5]

Item Score Correlations

Given the low internal consistency of clinician-rated summary scales, the relationship between these scales' item scores and total scores for the self-report and clinician-rated global severity measures was explored (see Table V). The MGH-HPS correlated moderately and significantly with the interference and distress items from the PITS.[6] With respect to clinician-rated global severity measures, the CGI demonstrated moderate to high and significant correlations with the NIMH-TSS items assessing distress and interference, and PITS items assessing interference and hair loss severity. In contrast, the NIMH-TIS and Alopecia rating correlated highly and significantly only with the PITS hair loss severity item.

[5]Analyses were rerun exploring correlations for the MGH-HPS urges (items #1-3) and pulling (items #4-6) subscales, as well as the distress item (item #7). Generally, the urges and pulling subscales correlational patterns were similar to each other and were similar to those for the total score. Thus, only data for the total score is presented in the manuscript. The distress item demonstrated a different correlational pattern from those for the subscales or total scores, showing moderate correlations with only the distress items on the PITS ($r = .58$, $p < .001$) and NIMH-TSS ($r = .46$, $p < .02$).

[6]See footnote 5.

Correlations with Mood Symptoms

Correlational analyses of TTM measures with mood symptoms are presented in Table VI. Again, an alpha correction of $p = .003$ (.05/15) was applied to interpretation of statistical significance, whereas correlations with a moderate effect size ($\geq .40$) are also highlighted. The BDI-II demonstrated moderate correlations with only the distress and interference items from the NIMH-TSS, as

Table V. Correlational Analyses of Items From Clinician-Rated Summary Scales With Self-Report and Clinician-Rated Global Severity Measures

	Self-report	Clinician-rated global severity		
	MGH-HPS	CGI	NIMH-TIS	Alopecia rating
NIMH-TSS				
Duration, past week	.34	.17	.17	.12
Duration, yesterday	.40	.17	.08	.13
Resistance	.33	.39	.06	.14
Distress	.28	.58***	.39	.04
Interference	.31	.76***	.51	.34
PITS				
Sites	−.08	.25	.21	.08
Duration	.40	.13	.20	−.03
Resistance	.51	−.00	−.04	−.05
Interference	.57***	.67***	.49	.48
Distress	.58***	.46	.45	.25
Hair loss severity	.13	.58***	.78***	.97***

Note. Correlations indicative of a moderate effect size ($r \geq .40$) are in italic. MGH-HPS = Massachusetts General Hospital Hairpulling Scale; NIMH-TSS = National Institute of Mental Health Trichotillomania Severity Scale; PITS = Psychiatric Institute Trichotillomania Scale; CGI = Clinical Global Impression; NIMH-TIS = National Institute of Mental Health Trichotillomania Impairment Scale.
*** $p < .003$.

Table VI. Correlations Between Hair-Pulling Measures and Mood Symptoms

	BDI-II	STAI-T
MGH-HPS	.26	*.53*
NIMH TSS Total	.39	*.59****
Duration, past week	.21	.24
Duration, yesterday	.09	.34
Resistance	.15	.28
Distress	*.45*	*.55****
Interference	*.41*	*.51*
PITS Total	.30	*.54****
Sites	−.20	−.15
Duration	.22	.30
Resistance	.31	.29
Interference	.34	*.63****
Distress	.35	*.54****
Severity	.14	.37
CGI	.47	*.50*
NIMH-TIS	.35	*.59****
Alopecia rating	.21	*.40*

Note. Correlations indicative of a moderate effect size ($r \geq .40$) are in italic. BDI-II = Beck Depression Inventory-II; STAI-T = State-Trait Anxiety Inventory-Trait; MGH-HPS = Massachusetts General Hospital Hairpulling Scale; NIMH-TSS = National Institute of Mental Health Trichotillomania Severity Scale; PITS = Psychiatric Institute Trichotillomania Scale; CGI = Clinical Global Impression; NIMH-TIS = National Institute of Mental Health Trichotillomania Impairment Scale.

*** $p < .003$.

well as the CGI. There were no statistically significant correlations of hair-pulling measures with the BDI-II. The STAI-T demonstrated moderate correlations with the MGH-HPS, NIMH-TSS total, PITS total, CGI, NIMH-TIS, and Alopecia rating, with several of these correlations reaching statistical significance. The NIMH-TSS and PITS items correlating most highly with the STAI-T assessed interference and distress.

DISCUSSION

Reliability and validity of three categories (self-report, clinician-rated summary scales, clinician-rated global severity scales) of commonly used hair pulling measures were evaluated. Results indicated the importance of using a multimethod assessment for TTM, given inconsistent convergent validity among the three types of assessment tools. The most prominent inconsistency was found for the MGH-HPS self-report measure and global severity scales, which both demonstrated convergent validity with clinician-rated summary scales, but not with each other. A review of item correlations indicated that a lack of hair loss severity assessment on the MGH-HPS may explain the inconsistent correlational patterns. Although the MGH-HPS demonstrated low correlations with hair loss severity, the CGI and NIMH-TIS correlated significantly with this item. In addition, the CGI demonstrated moderate to high correlations with interference and distress items from clinician-rated summary scales. Thus, the CGI may be a useful adjunct to self-report measures to assess clinically significant (distress, interference, hair loss) domains as required by a *DSM-IV* diagnosis.

In addition, hair loss severity is an important variable to include as a separate measure in treatment outcome assessment, especially given that hair loss severity appears independent of otherwise logically associated hair pulling variables (e.g., pulling frequency). For example, current data suggest low correlations between the Alopecia rating and most items assessed on clinician-rated summary scales (e.g., frequency, duration, resistance, number of sites). Subjective measures of hair loss are more commonly used than are objective measures, and the subjective Alopecia rating administered in the current study did demonstrate adequate interrater reliability. However, additional research comparing subjective with objective measurements of hair loss would bolster the validation of subjective hair loss assessment for use in future treatment outcome assessments.

Of the three categories of measures, clinician-rated summary scales (NIMH-TSS; PITS) demonstrated the weakest psychometric properties. Only the PITS demonstrated adequate interrater agreement for the total score, and neither measure demonstrated adequate internal consistency. Thus, current results are in agreement with previous findings suggesting superior psychometric properties for the PITS over the NIMH-TSS (Stanley et al., 1999). However, research employing the PITS could rely more heavily on item than total score analysis, given low internal consistency. The problem of low internal consistency has been reported in previous research (Stanley et al., 1999) and is consistent with conceptualizations of TTM emphasizing the diverse presentation of hair pulling patterns (Franklin, Tolin, & Diefenbach, in press; Mansueto, Stemberger, Thomas, & Golomb, 1997). Thus, it is unclear whether low internal consistency is a reflection of poor measurement or alternatively accurate measurement of an inherently inconsistent phenomenon.

Post hoc analyses were used to explore which item combination of the PITS may improve internal consistency. Findings indicated that summing three of the seven items (distress, interference, and hair-loss severity) raised internal consistency to adequate levels ($\alpha = .72$), with convergent validity patterns similar to those for the total score (data not shown). In addition, these items show good face validity as they are the most relevant

to *DSM-IV* diagnostic criteria for TTM. However, development of new clinician-rated summary scales might be a better approach to improving TTM assessment. New scales are best constructed using homogenous subscales for each of the domains of interest (e.g., distress, interference, frequency), rather than single items. Similar subscales were incorporated into a new self-report measure for pediatric TTM, which demonstrated good internal consistency for the total score (Diefenbach, Tolin, Franklin, & Anderson, 2003). Perhaps the same approach would lead to reliable measurement using clinician ratings.

Another possible approach is to develop assessment scales along two dimensions: current hair pulling behaviors (e.g., frequency, duration, resistance of pulling over the past week); and impact of current and past pulling behaviors (e.g., distress, hair loss, interference). Results of the current study suggest that these two dimensions are not consistently related to one another. However, both of these dimensions are important in terms of assessing treatment outcome. For example, although global measures (e.g., hair loss, distress, interference) are clearly relevant to treatment outcome assessment, the *process* of treatment effects is best measured through snapshots of behavioral changes. Thus, it is important to assess the process of behavioral change (e.g., frequency of pulling) even though the outcome of that change (e.g., hair growth, decreased interference) may not come to fruition for a longer period of time. Current summary scales appear to combine these two dimensions into a single rating, which may be limiting the internal consistency and clouding treatment outcome assessment.

The development of new psychometrically valid TTM scales would be an important contribution to TTM clinicians and researchers. However, it is also important to continue treatment and research efforts during the interim period. Although no gold standard measure is yet available, a multimethod approach is recommended. Assessment including self-monitoring, self-report using the MGH-HPS, global clinician-rated scale such as the CGI, items from clinician-rated summary scale such as the PITS, and subjective and objective ratings of hair loss severity provides a comprehensive best practice approach to assess hair pulling. Given the impact of hair pulling on psychosocial functioning (Diefenbach, Tolin, Hannan, Crocetto, & Worhunsky, 2003), it is also recommended that measures of mood symptoms and quality of life be administered to complement TTM assessment.

These findings and recommendations are considered with regard for study limitations. For example, data were collected on a small sample of participants. However, this limitation is mitigated somewhat, given that previous research on psychometric validation of TTM scales was similarly limited by sample size (O'Sullivan et al., 1995; Stanley et al., 1999). Thus, the importance of replication of previous findings is increased, and current data contribute replication of this work. In addition, participants in this study were preselected for involvement in a treatment outcome study, and it is unclear how results generalize to unselected samples of TTM patients. Another weakness of the study was the lack of self-monitoring or objective hair loss measurements to provide a basis for validity comparisons with self-report and clinician-rated scales. Nevertheless, these data contribute to a sparse literature on TTM assessment by providing replication of previous psychometric findings, as well as new data on a multimethod analysis defining the relationships among different types of TTM measures. Taken together these results support recommendations for enhancing clinical assessments as well as developing improved TTM scales.

ACKNOWLEDGMENTS

This research was funded by a grant from Hartford Hospital (Grant #126073) to Dr Diefenbach. The authors express their appreciation to Melinda Stanley, PhD for her helpful comments on an earlier draft of this manuscript.

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.

Barabasz, M. (1987). Trichotillomania: A new treatment. *The International Journal of Clinical and Experimental Hypnosis*, *35,* 146–154.

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

Bornstein, P. H., & Rychtarik, R. G. (1978). Multicomponent behavioral treatment of trichotillomania: A case study. *Behaviour Research and Therapy*, *16,* 217–220.

Christenson, G. A., MacKenzie, T. B., & Mitchell, J. E. (1991). Characteristics of 60 adult chronic hair pullers. *American Journal of Psychiatry*, *148,* 365–370.

Dahlke, F., Lohaus, A., & Gutzmann, H. (1992). Reliability and clinical concepts underlying global judgments in dementia: Implications for clinical research. *Psychopharmacology Bulletin*, *28,* 425–432.

Dahlquist, L. M., & Kalfus, G. R. (1984). A comparison of relaxation training and competing response training to eliminate hair pulling and nail biting. *Journal of Behavior Therapy and Experimental Psychiatry*, *15,* 67–70.

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

Diefenbach, G. J., Tolin, D. F., Franklin, M. E., & Anderson, E. R. (2003, November). *The Trichotillomania Scale for Children (TSC): A new self-report measure to assess pediatric hair pulling*. Paper presented at the Annual Meeting of the Association for Advancement of Behavior Therapy, Boston, MA.

Diefenbach, G. J., Tolin, D. F., Hannan, S. E., Crocetto, J. S., & Worhunsky, P. (2003, November). *Trichotillomania: Impact on*

Daily Functioning and Quality of Life. Paper presented at the Annual Meeting of the Association for Advancement of Behavior Therapy, Boston, MA.

Diefenbach, G. J., Tolin, D. F., Maltby, N., Hannan, S., & Crocetto, J. S. (2003, November). Group treatment for trichotillomania: CBT versus supportive therapy. In M. E. Franklin & N. J. Keuthen (Chairs), *New Developments in Trichotillomania Research*. Symposium conducted at the Annual Meeting of the Association for Advancement of Behavior Therapy, Boston, MA.

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

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995). *Structured Clinical Interview for DSM-IV Axis I Disorders—Patient Edition (SCID I/P, version 2.0)*. New York: Biometrics Research Department.

Franklin, M. E., Tolin, D. F., & Diefenbach, G. J. (in press). Trichotillomania. In E. Hollander & D. J. Stein (Eds.), *Handbook of impulse control disorders*. Washington, DC: American Psychiatric Press.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., et al. (1989). The Yale–Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Archives of General Psychiatry*, *46,* 1006–1011.

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

Keuthen, N. J., O'Sullivan, R. L., Ricciardi, J. N., Shera, D., Savage, C. R., Borgmann, A. S., et al. (1995). The Massachusetts General Hospital (MGH) Hairpulling Scale: 1. Development and factor analyses. *Psychotherapy and Psychosomatics*, *64,* 141–145.

Leon, A. C., Shear, M. K., Klerman, G. L., Portera, L., Rosenbaum, J. F., & Goldenberg, I. (1993). A comparison of symptom determinants of patient and clinician global ratings in patients with panic disorder and depression. *Journal of Clinical Psychopharmacology*, *13,* 327–331.

Mansueto, C. S., Stemberger, R. M. T., Thomas, A. M., & Golomb, R. G. (1997). Trichotillomania: A comprehensive behavioral model. *Clinical Psychology Review*, *17,* 567–577.

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,* 47–50.

O'Sullivan, R. L., Keuthen, N. J., Hayday, C. F., Ricciardi, J. N., Buttolph, M. L., Jenike, M. A., et al. (1995). The Massachusetts General Hospital (MGH) Hairpulling Scale: 2. Reliability and validity. *Psychotherapy and Psychosomatics*, *64,* 146–148.

Rothbaum, B. O., & Ninan, P. T. (1994). The assessment of trichotillomania. *Behaviour Research and Therapy*, *32,* 651–662.

Rothbaum, B. O., Opdyke, D. C., & Keuthen, N. J. (1999). Assessment of trichotillomania. In D. J. Stein, G. A. Christenson, & E. 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., Daphne, L., Insko, S. L., Atkinson, G., Jones, G. L., Logan, A. R., et al. (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,* 381–385.

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,* 157–170.

Swedo, S. E., Rapoport, J. L., Leonard, H., Lenane, M., & Cheslow, D. (1989). Obsessive-compulsive disorder in children and adolescents. Clinical phenomenology of 70 consecutive cases. *Archives of General Psychiatry, 46*, 335–341.

Tarnowski, K. J., Rosen, L. A., McGrath, M. L., & Drabman, R. S. (1987). A modified habit reversal procedure in a recalcitrant case of trichotillomania. *Journal of Behavior Therapy and Experimental Psychiatry*, *18,* 157–163.

Tolin, D. F., Franklin, M. E., Diefenbach, G. J., & Gross, A. (2002, November). *Cognitive-behavioral therapy for pediatric trichotillomania: An open trial*. Paper Presented at the Association for Advancement of Behavior Therapy, Reno, NV.

Winchel, R. M., Jones, J. S., Molcho, A., Parsons, B., Stanley, B., & Stanley, M. A. (1992a). The Psychiatric Institute Trichotillomania Scale (PITS). *Psychopharmacology Bulletin*, *28,* 463–476.

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