12-1-2015

Defining Treatment Response in Trichotillomania: A Signal Detection Analysis

David C. Houghton
Texas A&M University - College Station

Matthew R. Capriotti
University of California, San Francisco

Alessandro S. De Nadai
University of South Florida

Scott N. Compton
Duke University School of Medicine

Michael P. Twohig
Utah State University

See next page for additional authors

Accepted version. *Journal of Anxiety Disorders*, Vol. 36 (December 2015): 44-51. DOI.
Douglas W. Woods was affiliated with Texas A&M University at the time of publication.

Defining Treatment Response in Trichotillomania: A Signal Detection Analysis

David C. Houghton
Department of Psychology, Texas A&M University, College Station, TX
Matthew R. Capriotti
Department of Psychiatry, University of California San Francisco, San Francisco, CA
Alessandro S. De Nadai
Department of Psychology, University of South Florida, Tampa, FL
Scott N. Compton
Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC
Michael P. Twohig
Department of Psychology, Utah State University, Logan, UT
Angela M. Neal-Barnett
Department of Psychological Sciences, Kent State University, Kent, OH

Journal of Anxiety Disorders, Vol 36 (December 2015): pg. 44-51. DOI.

Stephen M. Saunders
Department of Psychology, Marquette University, Milwaukee, WI
Martin E. Franklin
Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA
Douglas W. Woods
Department of Psychology, Texas A&M University, College Station, TX

Abstract: The Massachusetts General Hospital Hairpulling Scale (MGH-HPS) and the NIMH Trichotillomania Severity Scale (NIMH-TSS) are two widely used measures of trichotillomania severity. Despite their popular use, currently no empirically-supported guidelines exist to determine the degrees of change on these scales that best indicate treatment response. Determination of such criteria could aid in clinical decision-making by defining clinically significant treatment response/recovery and producing accurate power analyses for use in clinical trials research. Adults with trichotillomania (N = 69) participated in a randomized controlled trial of psychotherapy and were assessed before and after treatment. Response status was measured via the Clinical Global Impressions-Improvement Scale, and remission status was measured via the Clinical Global Impressions-Severity Scale. For treatment response, a 45% reduction or 7-point raw score change on the MGH-HPS was the best indicator of clinically significant treatment response, and on the NIMH-TSS, a 30–40% reduction or 6-point raw score difference was most effective cutoff. For disorder remission, a 55–60% reduction or 7-point raw score change on the MGH-HPS was the best predictor, and on the NIMH-TSS, a 65% reduction or 6-point raw score change was the best indicator of disorder remission. Implications of these findings are discussed.

Keywords: Hair pulling, Trichotillomania, Obsessive-compulsive disorder, Signal detection, Psychotherapy

1. Introduction

Researchers have demonstrated the efficacy of various treatments for reducing hair pulling in adults with Trichotillomania (TTM; Bloch et al., 2007). Such studies typically utilize psychometrically-validated measures of pulling severity (Grant et al., 2009, Keuthen et al., 2012 and Woods et al., 2006), the most common of which are the Massachusetts General Hospital Hairpulling Scale (MGH-HPS; Keuthen et al., 1995) and the National Institutes of Mental Health Trichotillomania Severity Scale (NIMH-TSS; Swedo et al., 1989).

The MGH-HPS is a self-report measure and the NIMH-TSS is clinician-rated. Both are dimensional scales that possess sensitivity to change in TTM treatment studies (Diefenbach et al., 2005 and Swedo et al., 1989). Although existing treatments have yielded statistically significant changes in scores on both measures (Woods et al., 2006), the magnitude of reductions needed to signify clinically significant change is unclear.

When no clear cutoffs exist for a primary outcome measure, establishing the clinical significance of change requires the incorporation of additional information. For instance, clinicians might rely on a combination of qualitative and quantitative data to gauge improvement, thereby interpreting scores based on clinical judgment. An example of this type of measurement is the Clinical Global Impressions Scale (CGI; Guy, 1976), which consists of a severity index (CGI-S) and treatment improvement index (CGI-I). The CGI is a clinician-rated measure that incorporates multiple sources of data and provides a clearly interpretable metric of holistic disorder severity and treatment response. The CGI is also widely used in clinical trials (Bandelow et al., 2006, Leon et al., 1993, Leucht and Engel, 2006, Leucht et al., 2005, Spielmans and McFall, 2006 and Zaider et al., 2003) and has been used for trichotillomania (e.g., Keuthen et al., 2011 and Keuthen et al., 2012). To best determine the level of symptom reduction as measured by popular assessments of hair pulling severity, one could measure the points at which score reductions on dimensional measures (i.e., MGH-HPS and NIMH-TSS) converge best with the thresholds of clinical significance on the CGI-I and CGI-S.

Developing guidelines for clinically significant change on the MGH-HPS and NIMH-TSS would have numerous benefits in both research and clinical practice. When designing a randomized controlled trial (RCT), one ensures that the study is adequately powered to detect the desired effect size (Cohen, 1988 and Kraemer and Thiemann, 1987). Recent recommendations by Kraemer and Kupfer (2006) suggest that the level of power needed in studies be based on the determination of clinically significant effects. The current study attempts to identify clinically significant cutoff criteria in commonly used TTM outcome measures, so that future studies can better approximate the power needed to identify clinically significant effects. These guidelines will also have clinical utility, as a clinically meaningful change score can give therapists a target for change and can indicate the point at which change has become significant.

A recent study examined the ability of changes in the MGH-HPS and another clinician-rated measure of hair pulling severity, the Psychiatric Institute Trichotillomania Scale (PITS; Winchel et al., 1992) to predict various meaningful outcomes (Nelson et al., 2014). Various potential clinical predictors were used, including Jacobson and Truax's (1991) clinically significant change criteria (i.e., 1.96 times the reliable change index plus a post-treatment score that was two standard deviations below the dysfunctional population mean), complete abstinence from pulling (defined as a score of 0 on MGH-HPS item 4), 25% reduction on the MGH-HPS or PITS, and the recovery criterion alone (e.g., score of ≤9 on the MGH-HPS or ≤14 on the PITS). Post-treatment abstinence from hair pulling and the MGH-HPS 25% reduction predicted several positive outcomes (i.e., decision to successfully end treatment at step 2 in the stepped-care clinical trial, treatment satisfaction, and quality of life at 3-month follow-up), but the Jacobson and Truax clinically significant change criteria on the MGH-HPS predicted only quality of life at 3-month follow-up. The 25% PITS reduction predicted no outcomes, whereas the PITS-based recovery criterion predicted decision to end treatment and the Jacobson and Truax clinically significant change criteria on the PITS predicted absence of TTM diagnosis at 3-month follow-up. As such, it appears that the ways of defining different clinical predictors leads to differential prediction of various indices of treatment response. However, no cutoff stands out as the most efficient indicator of treatment response. Determining more efficient cutoffs might be

achieved through approaches that are not constrained by rigid definitions of these cutoffs, such as by testing the validity and efficiency of numerous score reductions as they converge with well-defined measures of clinically significant change (i.e., the CGI).

Indeed, five studies have performed signal detection analyses to determine such cutoffs with related conditions, such as obsessive-compulsive disorder and tic disorders. Investigators found that a 25% decrease on the Children's Yale-Brown Obsessive-Compulsive Scale was most efficient at predicting treatment response in childhood OCD, as measured by the CGI-I and the Child Obsessive-Compulsive Impact Scale (Storch, Lewin, De Nadai, & Murphy, 2010), while others found between 30 and 35% reductions on the Yale-Brown Obsessive-Compulsive Scale were most efficient in predicting adult OCD treatment response as measured by the CGI-I (Lewin et al., 2011 and Tolin et al., 2005). Likewise, a 35% reduction or 6–7 point raw score decrease on the Yale Global Tic Severity Scale (YGTSS) was found to best predict treatment response in Tourette syndrome as measured by the CGI-I (Storch et al., 2011), whereas Jeon et al. (2013) found that a 25% reduction on the YGTSS optimally predicted positive response as measured by the CGI-I in both children and adults with tic disorders. Although these studies allow clinicians to accurately predict which clients demonstrate clinically significant treatment *response*, no studies have determined reductions on dimensional measures of obsessive-compulsive related disorders that optimally predict disorder *recovery*. As was done in the Nelson et al. study on measures of treatment response in TTM, researchers have argued that estimates of clinical significance should calculate the propensity of a treatment to facilitate a decrease in symptoms within clinical individuals to those resembling normative levels (Jacobson & Truax, 1991). Thus, it would be useful to determine if certain levels of symptom reduction on dimensional scales correspond to both reliable change and recovery of normal functioning.

The present study sought to replicate the methods of previous signal detection analyses in defining treatment response for adults with TTM using both the MGH-HPS and the NIMH-TSS. In order to determine clinically significant treatment response, we used the CGI-I as the criterion measure. Similarly, the CGI-S was used as the

criterion measure of TTM recovery. No a priori hypotheses were made with regard to optimal cutoff points on the measures analyzed.

2. Method

2.1. Participants

Although 85 participants were randomized into the clinical trial, only those who completed treatment were included in the present study. Participants were 69 adults (62 females) diagnosed with TTM whose ages ranged from 18 to 61 (*M* = 35.86, SD = 13.05). The sample was 85.5% Caucasian, 11.6% African–American, and 2.9% "other." Data were collected as part of a randomized controlled trial for psychotherapy for adults with TTM (Woods et al., in preparation). Both therapeutic conditions tested in the trial (i.e., Acceptance-Enhanced Behavior Therapy and psychoeducation plus supportive psychotherapy) are included in these analyses. Also, only participants who completed both the baseline and post-treatment assessments were included. At baseline, mean scores on the MGH-HPS and NIMH-TSS were 16.99 (SD = 4.68, Range = 8–26) and 14.54 (SD = 3.72, Range = 6–21), respectively.

Inclusion criteria were: (1) a current DSM-IV-TR diagnosis of TTM (2) an MGH-HPS score of >12, (3) a Wechsler Test of Adult Reading score of >85, (4) age 18–65, (5) English fluency, (6) able to maintain outpatient status, (7) no initiation or change in psychotropic medication status or dosage for eight weeks preceding participation or during the study, (8) not currently receiving psychotherapy for any condition, and (9) completed all 10 sessions of treatment.

Exclusion criteria included: (1) diagnosis of bipolar disorder, psychotic disorder, substance dependence (except nicotine dependence), or pervasive developmental disorder, and (2) severe mood or anxiety problems with potential suicidality. In addition, individuals who endorsed ingesting their hair after pulling were eligible for participation only after they had received a physical exam from their primary care physician.

2.2. Treatment

Participants were randomized to receive either Acceptance-Enhanced Behavior Therapy (AEBT; *n* = 35) or psychoeducation and supportive psychotherapy (PST; *n* = 34) control. For a detailed description of AEBT therapeutic techniques, see Woods and Twohig (2008). The PST protocol was derived from Pinsker (1997). Inclusion criteria mandated that participants maintain a stable dose on any medications for the 8 weeks prior to and during the study. In total, 29% were currently taking a psychotropic medication during the study, but only 2.9% were prescribed medication for TTM. Of the sample, 21.7% were taking selective serotonin reuptake inhibitors, 7.2% were taking other antidepressants (e.g., tricyclics), 7.2% were taking psychostimulants, 2.9% benzodiazepines, 2.9% reported taking atypical neuroleptics, and 1 person (1.4%) was taking Hydroxyzine (an antihistamine) for anxiety. One-fifth of the total sample (20.3%) were taking only one medication, while 4.3% were taking two medications and 4.3% were taking three or four medications.

2.3. Measures

The Structured Clinical Interview for DSM-IV Axis-I Disorders, Patient Edition (SCID-P; First, Spitzer, Gibbon, & Williams, 1996) was used to screen for psychiatric comorbidities. Additionally, the Trichotillomania Diagnostic Interview (TDI; Rothbaum & Ninan, 1994) was employed for obtaining TTM diagnosis.

The MGH-HPS has demonstrated adequate psychometric properties (Diefenbach et al., 2005, Keuthen et al., 1995 and O'Sullivan et al., 1995). It consists of seven items that are scored on a 0–4 Likert scale, resulting in total scores ranging from 0 to 28, with higher scores indicating greater hair pulling severity. The MGH-HPS was administered at baseline and post-treatment.

The NIMH-TSS has demonstrated adequate psychometric properties in adults (Diefenbach et al., 2005 and Swedo et al., 1989). Interviewers using the NIMH-TSS ask questions about time spent pulling, resistance to urges, distress, and impairment, resulting in total scores that range from 0 to 25. The NIMH-TSS was also administered at baseline and post-treatment.

The CGI was developed to provide a brief, stand-alone measure of clinician-rated global treatment response and disorder severity in NIMH-sponsored clinical trials (Guy, 1976). The CGI has evidence of convergent validity on many symptom severity scales across many psychiatric conditions in both pharmacological and psychosocial treatment paradigms (Bandelow et al., 2006, Leon et al., 1993, Leucht and Engel, 2006, Leucht et al., 2005, Spielmans and McFall, 2006 and Zaider et al., 2003) and has been used for TTM (e.g., Keuthen et al., 2011 and Keuthen et al., 2012). The CGI-I is a single-item clinician-rated measure that assesses the overall improvement of a person's condition throughout treatment on an 8-point Likert scale (Range = 1–8). Scores of 1 and 2 (*very much improved* and *much improved*) are indicators of treatment response while all greater scores indicate treatment non-response. Similarly, the CGI-S is a single-term clinician rated measure that assesses the overall severity of a person's condition on an 8-point Likert scale (Range = 1–8). Scores of 1 and 2 (*normal, not at all ill* and *borderline ill*) are indicators of no TTM diagnosis or mild TTM symptoms, while all greater scores indicate significant TTM symptoms. The CGI-I and CGI-S were administered at post-treatment. To ensure the validity of CGI ratings, masked independent evaluators were trained in CGI administration and met weekly with the Principal Investigator (D.W.W.) to discuss and review taped assessments.

2.4. Procedure

Adults with TTM were recruited to participate in a randomized controlled trial of psychotherapy for TTM via local newspaper ads, public transportation flyers, newsletter and website advertisements via the Trichotillomania Learning Center (www.trich.org), and clinic referrals at a TTM specialty clinic.

Potential participants were screened by telephone. All callers to a TTM clinic were provided information about the study and screened for possible participation. If the participant appeared to be eligible and interested, he or she was scheduled for an initial clinic visit, during which consent was obtained and inclusion/exclusion criteria checked. Participants deemed ineligible or those not wishing to participate were referred for standard clinical services. Potential participants (N = 274)

were screened via telephone. The baseline sample consisted of 91 persons, of which 85 were randomized and 16 participants were lost throughout treatment, resulting in a post-treatment sample of 69 persons. For additional details regarding screening, exclusions, and attrition, see Woods et al. (in preparation). Additionally, all clinician-rated instruments were administered by masters- and doctoral-level independent evaluators who were masked to treatment condition. The CGI scales and the NIMH-TSS are rated using a semi-structured procedure.

IRB approval for this project was obtained at Texas A&M University (IRB2013-3025) and the University of Wisconsin-Milwaukee (IRB09.039). The study is publicly listed on ClinicalTrials.gov (#NCT00872742), and was performed in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.5. Analyses

The goal of the current study was to find the levels of symptom reduction needed on the MGH-HPS and NIMH-TSS that most optimally predicted treatment response (i.e., CGI-I < 3) and disorder recovery (i.e., CGI-S < 3). Both percent reductions and raw score reductions (from baseline to post-treatment) on each measure were used to predict the CGI-I and CGI-S. The authors chose not to define clinically significant treatment response as meeting both significant change on the CGI-I and significant recovery on the CGI-S. This decision was due to the fact that although many individuals with TTM wish to achieve complete abstinence from pulling, many others are satisfied with a significant reduction in hair pulling (Woods & Houghton, 2014). Thus, persons with severe TTM who show clinically meaningful symptom reductions but do not achieve complete recovery should not be discounted as having not responded to treatment, whereas those persons would be ignored by definitions of clinically significant treatment response that require both change and recovery. Additionally, performing such analyses separately allows a more detailed interpretation of the assessment of change in treatment for TTM.

Receiver operating characteristic (ROC) methods (Swets & Pickett, 1982), which have been previously used for these purposes

(Storch et al., 2011), were used in the present study. ROC methods focus on the predictive validity of psychological tests, using statistics such as number of true positives, true negatives, false positives, and false negatives. In testing percent reduction cutoffs, we created cutoffs at every 5% interval between 5% and 70%. For raw scores, point reductions between 1 and 11 were evaluated. Following the methodology of Storch et al. (2011), our analysis operationalized score reductions as raters, then tested which reduction (or "rater") has the best psychometric efficiency for detecting clinical response to treatment.

ROC analyses incorporate several psychometric properties of assessments, including sensitivity, specificity, positive predictive power, negative predictive power, and efficiency. Sensitivity is defined as a measure's ability to detect the presence of a given criterion (in this study, clinically significant treatment response or disorder recovery). Alternatively, specificity is defined as a test's ability to detect the absence of a given criterion. Positive predictive power (PPP) reflects the proportion of correctly predicted positive results provided by a diagnostic test, whereas negative predictive power (NPP) reflects the proportion of correctly predicted negative results. Efficiency can be described as the accuracy of a test, such that a given cutoff or rating on a test "agrees" with another definitive test.

Even the most psychometrically sound tests contain at least minimal measurement error (in this study, the CGI-I and CGI-S). Therefore, a weighted Kappa statistic was used to correct for such error when assessing the quality of efficiency (Kraemer, 1992 and Kraemer et al., 2002). Weighted kappa statistics examine the agreement between measures but correct for measurement error in a manner similar to the method by which Cohen's Kappa accounts for chance agreement in inter-observer reliability. For this analysis, the *K*(0.5) statistic was used, which ranges from 0.00–1.00. A value of 0 is indicative of agreement purely by chance, and a value of 1 reflects perfect classification (i.e., all true positives and true negatives). The *K*(0.5) statistic measures the quality of efficiency while weighing sensitivity and specificity equally, and was used in order to generalize results across contexts, following the approach of Storch et al. (2011).

3. Results

3.1. Adequacy of measures for signal detection analysis

Reliability analyses were performed in order to determine whether the MGH-HPS and NIMH-TSS were suitable for signal detection analysis. Test-retest reliability correlations were computed from the screening assessment date to the baseline assessment date, a time period that lasted, on average, 11.81 days (SD = 6.04). The MGH-HPS test-retest reliability coefficient was 0.45 ($p < 0.001$) and the NIMH-TSS reliability coefficient was 0.65 ($p < 0.001$), which are comparable to reliability coefficients of TTM severity instruments at similar intervals (McGuire et al., 2012 and Stanley et al., 1993). Because these measures assess hair-pulling severity during the previous week, and because hair pulling is a constantly fluctuating behavior, we deemed these reliability coefficients to be acceptable and that the measures were suitable for signal detection analysis.

3.2. Determining treatment response and recovery based on MGH-HPS percentage reduction

Table 1 shows ROC and quality assurance statistics for assessing performance of MGH-HPS percent reduction cutoffs in detecting clinical response and recovery. Results showed that 45% reductions optimally predicted treatment response (as measured by the *K*(0.5) statistic), with the predictive value of a positive test at 0.90 and predictive value of a negative test at 0.79. Recovery from TTM was optimally predicted by 55–60% reductions, which showed predictive values of a positive test at 0.79 and 0.83 and predictive values of negative tests at 0.86 and 0.83, respectively.

Table 1. Signal detection analysis of the prediction of clinical response and recovery at increasing Massachusetts General Hospital Hairpulling Scale (MGH-HPS) total percent reduction cutoff scores.

MGH-HPS reduction (%)	Sensitivity	Specificity	Positive predictive power	Negative predictive power	Efficiency	*K*(0.5)
Predicting treatment response (based on CGI-I)						
≥5	0.98	0.31	0.70	0.89	0.72	0.33
≥10	0.95	0.31	0.70	0.80	0.71	0.3

Journal of Anxiety Disorders, Vol 36 (December 2015): pg. 44-51. DOI.

MGH-HPS reduction (%)	Sensitivity	Specificity	Positive predictive power	Negative predictive power	Efficiency	*K*(0.5)
≥15	0.93	0.39	0.72	0.77	0.72	0.35
≥20	0.93	0.46	0.74	0.80	0.75	0.43
≥25	0.91	0.54	0.77	0.78	0.77	0.47
≥30	0.91	0.62	0.80	0.80	0.80	0.55
≥35	0.88	0.73	0.84	0.79	0.83	0.62
≥40	0.88	0.81	0.88	0.81	0.86	0.69
≥45	0.86	0.85	0.90	0.79	0.86	0.70
≥50	0.84	0.85	0.90	0.76	0.84	0.67
≥55	0.70	0.89	0.91	0.64	0.77	0.54
≥60	0.63	0.92	0.93	0.60	0.74	0.50
≥65	0.47	0.96	0.95	0.52	0.65	0.37
≥70	0.35	1	1	0.48	0.59	0.29
			Predicting recovery (based on CGI-S)			
≥5	1	0.24	0.52	1	0.58	0.22
≥10	1	0.26	0.53	1	0.59	0.24
≥15	1	0.34	0.55	1	0.64	0.32
≥20	1	0.40	0.57	1	0.67	0.37
≥25	0.97	0.45	0.59	0.94	0.68	0.39
≥30	0.97	0.50	0.61	0.95	0.71	0.44
≥35	0.94	0.58	0.64	0.92	0.74	0.49
≥40	0.94	0.63	0.67	0.92	0.77	0.55
≥45	0.94	0.68	0.71	0.93	0.80	0.60
≥50	0.94	0.71	0.73	0.93	0.81	0.63
≥55	0.84	0.82	0.79	0.86	0.83	0.65
≥60	0.77	0.87	0.83	0.83	0.83	0.65
≥65	0.58	0.92	0.86	0.73	0.77	0.52
≥70	0.045	0.97	0.93	0.69	0.74	0.45

3.3. Determining treatment response and recovery based on MGH-HPS raw score reduction

Table 2 shows ROC and quality assurance statistics for assessing performance of MGH-HPS point reduction cutoffs in detecting clinical response and recovery. These results indicate that a seven-point raw score reduction was most efficient at identifying treatment response. PPP at the seven-point level was 0.82 while NPP was 0.72. Similarly, the seven-point raw score reduction was most efficient at identifying recovery, with PPP and NPP at 0.64 and 0.88, respectively. Of note, the *K*(0.5) values reflect agreement that is not as strong as when the MGH-HPS percent reductions are used, and the peak *K*(0.5) value for raw score reductions predicting recovery (0.46) is lower than the peak

raw score reductions predicting response (0.53). As such, raw score reductions, particularly predicting recovery, might not be very efficient prediction tools.

Table 2. Signal detection analysis of the prediction of clinical response and recovery at increasing Massachusetts General Hospital Hairpulling Scale (MGH-HPS) total raw score cutoff scores.

MGH-HPS reduction (%)	Sensitivity	Specificity	Positive predictive power	Negative predictive power	Efficiency	*K*(0.5)
Predicting treatment response (based on CGI-I)						
≥1	0.98	0.31	0.70	0.89	0.72	0.33
≥2	0.93	0.35	0.70	0.75	0.71	0.31
≥3	0.93	0.39	0.71	0.77	0.72	0.35
≥4	0.88	0.54	0.76	0.74	0.75	0.45
≥5	0.88	0.54	0.76	0.74	0.75	0.45
≥6	0.88	0.62	0.79	0.76	0.78	0.52
≥7	0.84	0.69	0.82	0.72	0.78	0.53
≥8	0.70	0.77	0.83	0.61	0.72	0.44
≥9	0.70	0.85	0.88	0.63	0.75	0.51
≥10	0.61	0.89	0.9	0.58	0.71	0.44
≥11	0.54	0.89	0.89	0.54	0.66	0.37
Predicting recovery (based on CGI-S)						
≥1	1	0.24	0.52	1	0.58	0.22
≥2	1	0.32	0.54	1	0.62	0.29
≥3	1	0.34	0.55	1	0.64	0.32
≥4	0.94	0.45	0.58	0.90	0.67	0.36
≥5	0.94	0.45	0.58	0.90	0.67	0.36
≥6	0.94	0.50	0.60	0.91	0.70	0.42
≥7	0.90	0.58	0.64	0.88	0.72	0.46
≥8	0.71	0.63	0.61	0.73	0.67	0.34
≥9	0.71	0.68	0.65	0.74	0.70	0.39
≥10	0.61	0.74	0.66	0.74	0.68	0.35
≥11	0.55	0.76	0.65	0.67	0.67	0.32

3.4. Determining treatment response and recovery based on NIMH-TSS percent reduction

Table 3 shows ROC and quality assurance statistics for assessing performance of NIMH-TSS percent reductions cutoffs in detecting treatment response and recovery. Results indicate that a 30–40% reduction in scores maximally predict clinical response, with PPP and NPP at 0.89 and 0.84 for all percentiles within that range. Recovery

from TTM was optimally predicted by a much higher percentile reduction, 65%, which showed PPP of 0.96 and NPP of 0.84.

Table 3. Signal detection analysis of the prediction of clinical response and recovery at increasing National Institutes of Mental Health Trichotillomania Severity Scale (NIMH-TSS) total percent reduction cutoff scores.

NIMH-TSS reduction (%)	Sensitivity	Specificity	Positive predictive power	Negative predictive power	Efficiency	*K*(0.5)
Predicting treatment response (based on CGI-I)						
≥5	1	0.27	0.69	1	0.72	0.32
≥10	1	0.42	0.74	1	0.78	0.48
≥15	0.98	0.50	0.76	0.93	0.80	0.53
≥20	0.98	0.54	0.78	0.93	0.81	0.56
≥25	0.93	0.58	0.78	0.83	0.80	0.54
≥30	0.91	0.81	0.89	0.84	0.87	0.72
≥35	0.91	0.81	0.89	0.84	0.87	0.72
≥40	0.91	0.81	0.89	0.84	0.87	0.72
≥45	0.88	0.81	0.88	0.81	0.86	0.69
≥50	0.86	0.85	0.90	0.79	0.86	0.70
≥55	0.72	0.89	0.91	0.67	0.78	0.57
≥60	0.63	0.96	0.96	0.61	0.75	0.53
≥65	0.56	0.96	0.96	0.57	0.71	0.46
≥70	0.42	0.96	0.95	0.50	0.62	0.32
Predicting recovery (based on CGI-S)						
≥5	1	0.18	0.50	1	0.55	0.17
≥10	1	0.30	0.53	1	0.61	0.27
≥15	0.97	0.34	0.55	0.93	0.62	0.29
≥20	0.97	0.37	0.56	0.93	0.64	0.32
≥25	0.97	0.45	0.59	0.94	0.68	0.39
≥30	0.97	0.63	0.68	0.96	0.78	0.58
≥35	0.97	0.63	0.68	0.96	0.78	0.58
≥40	0.97	0.63	0.68	0.96	0.78	0.58
≥45	0.94	0.63	0.67	0.92	0.77	0.55
≥50	0.94	0.68	0.71	0.93	0.80	0.60
≥55	0.87	0.82	0.79	0.89	0.84	0.68
≥60	0.81	0.92	0.89	0.85	0.86	0.73
≥65	0.77	0.97	0.96	0.84	0.88	0.76
≥70	0.61	1	1	0.76	0.82	0.64
≥75	0.48	1	1	0.70	0.77	0.51

Journal of Anxiety Disorders, Vol 36 (December 2015): pg. 44-51. DOI.

3.5. Determining treatment response and recovery based on NIMH-TSS raw score reduction

Table 4 shows ROC and quality assurance statistics for assessing performance of NIMH-TSS raw score reduction cutoffs in detecting treatment response and recovery. Results show that a six-point reduction on this measure maximally predicts clinical response, with PPP and NPP at 0.88 and 0.78, respectively. Similarly, the six-point reduction also optimally predicted recovery, with PPP at 0.69 and NPP 0.93. Of note, the *K*(0.5) values in this analysis are considerably lower than those shown when using the NIMH-TSS percent reductions, appearing to behave similarly to the relationship between percent reductions and raw cutoffs on the MGH-HPS.

Table 4. Signal detection analysis of the prediction of clinical response and recovery at increasing National Institutes of Mental Health Trichotillomania Severity Scale (NIMH-TSS) total point reduction cutoff scores.

NIMH-TSS reduction (%)	Sensitivity	Specificity	Positive predictive power	Negative predictive power	Efficiency	*K*(0.5)
Predicting treatment response (based on CGI-I)						
≥1	1	0.27	0.69	1	0.72	0.32
≥2	0.98	0.46	0.75	0.92	0.78	0.49
≥3	0.98	0.50	0.76	0.93	0.80	0.53
≥4	0.93	0.62	0.80	0.84	0.81	0.58
≥5	0.88	0.73	0.84	0.79	0.83	0.62
≥6	0.86	0.81	0.88	0.78	0.84	0.66
≥7	0.77	0.81	0.87	0.68	0.78	0.55
≥8	0.70	0.81	0.86	0.62	0.74	0.48
≥9	0.61	0.85	0.87	0.56	0.70	0.41
≥10	0.42	0.92	0.90	0.49	0.61	0.29
≥11	0.33	0.92	0.88	0.45	0.55	0.21
Predicting recovery (based on CGI-S)						
≥1	1	0.18	0.5	1	0.55	0.17
≥2	0.97	0.32	0.54	0.92	0.61	0.26
≥3	0.97	0.34	0.55	0.93	0.62	0.29
≥4	0.97	0.47	0.6	0.95	0.7	0.42
≥5	0.94	0.58	0.64	0.92	0.74	0.49
≥6	0.94	0.66	0.69	0.93	0.78	0.58
≥7	0.84	0.68	0.68	0.84	0.75	0.51
≥8	0.74	0.68	0.66	0.77	0.71	0.42
≥9	0.68	0.76	0.7	0.74	0.72	0.44
≥10	0.48	0.87	0.75	0.67	0.7	0.36

Journal of Anxiety Disorders, Vol 36 (December 2015): pg. 44-51. DOI.

NIMH-TSS reduction (%)	Sensitivity	Specificity	Positive predictive power	Negative predictive power	Efficiency	*K*(0.5)
≥11	0.42	0.92	0.81	0.66	0.7	0.36

4. Discussion

The purpose of this investigation was to identify maximally efficient cutoff levels of two widely used measures of TTM severity. For predicting treatment response, the most efficient reductions on the MGH-HPS were found to be 45% or seven-point reductions, and the most efficient reductions on the NIMH-TSS were found to be 30–40% or six-points. For predicting recovery from TTM, the most efficient reductions on the MGH-HPS were found to be 55–60% or seven-point reductions, and the most efficient reductions on the NIMH-TSS were found to be 65% or six-points. We offer these empirically derived cutoffs so that future researchers and clinicians might utilize them to maximize their predictive validity in labeling TTM patients as clinically significant treatment responders. Likewise, researchers who develop clinical trials using these outcome measures should power their studies to ensure that these clinically meaningful effect sizes can be detected.

It might be expected that the degree of score reduction needed to achieve response might be less than that needed to achieve recovery from TTM. However, results showed that while percentage reductions where higher when predicting recovery than when predicting response, the raw score cutoffs did not change. This result might be explained by several factors. When the CGI-I is rated, trained evaluators consider the degree of change shown by the individual with reference to their baseline disorder severity. Conversely, the CGI-S ratings are static categories of TTM severity. An individual who enters treatment in the severe range of severity but exits treatment in the moderate range could be seen as having the same degree of improvement as an individual who enters treatment in the moderate range and exits in the mild, or undiagnosed, range. However, achieving recovery requires a greater degree of change for those who start treatment in the severe range as compared to those who start in the moderate range, meaning that individuals in the latter group are more likely to achieve recovery than those in the former group. For example, an MGH-HPS reduction from 12 to 5 conveys a very different

Journal of Anxiety Disorders, Vol 36 (December 2015): pg. 44-51. DOI.

clinical picture than one from 26 to 19. Both are 7-point reductions, but the former would be considered to have significantly improved and recovered while the latter has just significantly improved. If we were instead measuring the same two hypothetical individuals' response by percent reduction, the first would constitute a 58% reduction while the second would only represent a reduction of 27%. As such, the relatively high percent changes but small raw score change shown by those who recover from moderately severe pulling would be over-represented in those who recover from TTM and cause the difference in percentile reductions seen between predictions of response and recovery.

The same problem could have also conversely influenced the finding that the efficiency of both measures was notably higher when using percent reductions rather than raw score reductions. Indeed, while floor effects do bias the interpretation of percent reductions as predicting response versus recovery, they do provide an index of the degree of change relative to baseline levels. Raw score differences contain no information about baseline disorder severity, and thus might be less efficient at predicting criterion indices of treatment response and recovery. Still, additional factors might also influence the effectiveness of both raw score and percent reductions in predicting treatment response and recovery from TTM, such as regression toward the mean. With these scaling limitations in mind, clinicians should consider both raw score and percent reductions when determining whether a particular client has significantly responded to treatment or recovered from TTM. Researchers should also consider which criterion of improvement is most important to use when powering a study, as judgments of treatment efficacy could be expected to significantly vary depending on this question (Nelson et al., 2014).

This study identified the most optimal cutoffs based on their agreement with a criterion outcome measure, but the cutoff percentages and score reductions surrounding the most optimal cutoffs did not drop off steeply. This suggests that the incremental efficiency of this study's proposed cutoffs relative to nearby cutoffs is low, and other studies might find similar but not exact replications. In order to determine if these cutoffs generalize to other samples, multiple replications are required. It is also important that the cutoffs recommended in the current study be placed into clinical context. Hair

pulling is a constantly fluctuating behavior, and scores on the MGH-HPS and NIMH-TSS can vary over the passage of time. Our analyses indicated that certain raw score and percentage reductions could best identify persons who responded to treatment, but clinicians who treat TTM should consider these cutoffs alongside other clinical data.

In addition to the cutoff scores generated, the study had a number of methodological strengths, including a relatively large sample (for a disorder of low prevalence), administration of multiple treatments, and the use of multiple measures with separate response formats. Furthermore, we examined two widely used measures of TTM symptom severity, one being self-report and the other clinician-administered. Results of this analysis are thus applicable in a variety of assessment contexts, whether one wishes to use only one method or collect multiple sources of information.

The study had several limitations. First, the analysis could have been strengthened through the inclusion of additional therapeutic conditions, such as pharmacotherapy. Given that meta-analyses have consistently shown that drug treatments of TTM are less effective than behavioral treatments (Bloch et al., 2007 and McGuire et al., 2014), this might be considered a minor limitation. Second, the findings could have been enhanced by an analysis of moderating variables, such as gender or age. It is possible that such factors might predict different degrees of symptom reduction necessary for clinical response. However, the sample was heavily biased towards females (89.85%), as is common in treatment trials of TTM (Christenson, Mackenzie, & Mitchell, 1994). The highly unequal cell sizes would have made such analyses inappropriate for gender. Third, adults were the only age group studied in this intervention, and clinically significant symptom reductions might be different in children and adolescents. Comparable analyses within pediatric populations are needed to examine the generalizability of these cutoffs for all age groups.

Despite these limitations, the current study represents the first effort at quantifying clinically significant dimensional reductions in hair pulling severity. Given that the MGH-HPS and NIMH-TSS are widely used in TTM research and treatment, researchers and providers can make use of the guidelines for assessing treatment response outlined in this study.

Acknowledgements

Research reported in this paper was supported by the NIMH of the National Institutes of Health under Awards number R01MH080966 (Woods; PI) and F31MH094095 (De Nadai; PI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The authors would like to thank the Trichotillomania Learning Center for assisting in recruiting as well as the participants in this study. The authors would also like to thank Steve Hayes, Thilo Deckersbach, Flint Espil, Mike Walther, Chris Bauer, Shawn Cahill, Jason Levine, Emily Ricketts, Bryan Brandt, Zach Hosale, Joe Rohde, Valerie Esser, and Olivia Smith for their assistance on this project.

References

Bandelow et al., 2006. B. Bandelow, D.S. Baldwin, O.T. Dolberg, D.F. Andersen, D.J. Stein. What is the threshold for symptomatic response and remission for major depressive disorder, panic disorder, social anxiety disorder, and generalized anxiety disorder? *Journal of Clinical Psychiatry*, 67 (9) (2006), pp. 1428–1434

Bloch et al., 2007. M.H. Bloch, A. Landeros-Weisenberger, P. Dombrowski, B. Kelmendi, R. Wegner, J. Nudel, *et al.* Systematic review: pharmacological and behavioral treatment for trichotillomania. *Biological Psychiatry*, 62 (8) (2007), pp. 839–846 http://dx.doi.org/10.1016/j.biopsych.2007.05.019

Christenson et al., 1994. G.A. Christenson, T.B. Mackenzie, J.E. Mitchell. Adult men and women with trichotillomania—a comparison of male and female characteristics. *Psychosomatics*, 35 (2) (1994), pp. 142–149

Cohen, 1988. J. Cohen. *Statistical power analysis for the behavioral sciences.* Routledge (1988)

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

First et al., 1996. M.B. First, R.L. Spitzer, M. Gibbon, J. Williams. Structured clinical interview for DSM-IV axis I disorders patient-version (SCID-I/P) NYSPI. Biometrics Research Department, New York State Psychiatric Institute, New York, NY (1996)

Grant et al., 2009.J.E. Grant, B.L. Odlaug, S.W. Kim. *N*-Acetylcysteine, a glutamate modulator, in the treatment of trichotillomania. *Archives of General Psychiatry*, 66 (7) (2009), pp. 756–763

Guy, 1976. W. Guy. The clinical global impression scale. *The ECDEU assessment manual for psychopharmacology* revised (Vol. DHEW Publ. No. ADM 76-338), Public Health Service (1976), pp. 218–222

Jacobson and Truax, 1991. N.S. Jacobson, P. Truax. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59 (1) (1991), pp. 12–19

Jeon et al., 2013. S. Jeon, J.T. Walkup, D.W. Woods, A. Peterson, J. Piacentini, S. Wilhelm, *et al.* Detecting a clinically meaningful change in tic severity in Tourette syndrome: a comparison of three methods. *Contemporary Clinical Trials*, 36 (2) (2013), pp. 414–420

Kraemer, 1992. H.C. Kraemer. *Evaluating medical tests: objective and quantitative guidelines*. Sage, Newbury Park, CA (1992)

Kraemer and Kupfer, 2006. H.C. Kraemer, D.J. Kupfer. Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry*, 59 (11) (2006), pp. 990–996

Kraemer and Thiemann, 1987. H.C. Kraemer, S. Thiemann. *How many subjects?* Sage, Newbury Park, CA (1987)

Kraemer et al., 2002. H.C. Kraemer, V.S. Periyakoil, A. Noda. Kappa coefficients in medical research. *Statistics in Medicine*, 21 (14) (2002), pp. 2109–2129 http://dx.doi.org/10.1002/sim.1180

Keuthen et al., 1995. N.J. Keuthen, R.L. O'Sullivan, J.N. Ricciardi, D. Shera, C.R. Savage, A.S. Borgmann, *et al.* The Massachusetts General Hospital (MGH) Hairpulling Scale: 1. Development and factor analyses. *Psychotherapy and Psychosomatics*, 64 (3–4) (1995), pp. 141–145

Keuthen et al., 2012. N.J. Keuthen, B.O. Rothbaum, J. Fama, E. Altenburger, M.J. Falkenstein, S.E. Sprich, *et al.* DBT-enhanced cognitive-behavioral treatment for trichotillomania: a randomized controlled trial. *Journal of Behavioral Addictions*, 1 (3) (2012), pp. 106–114 http://dx.doi.org/10.1556/jba.1.2012.003

Keuthen et al., 2011. N.J. Keuthen, B.O. Rothbaum, M.J. Falkenstein, S. Meunier, K.R. Timpano, M.A. Jenike, *et al.* DBT-enhanced habit reversal treatment for trichotillomania: 3-and 6-month follow-up results. *Depression and Anxiety*, 28 (4) (2011), pp. 310–313

Lewin et al., 2011. A.B. Lewin, A.S. De Nadai, J. Park, W.K. Goodman, T.K. Murphy, E.A. Storch. Refining clinical judgment of treatment outcome in obsessive-compulsive disorder. *Psychiatry Research*, 185 (3) (2011), pp. 394–401 http://dx.doi.org/10.1016/j.psychres.2010.08.021

Leon et al., 1993. A.C. Leon, M.K. Shear, G.L. Klerman, L. Portera, J.F. Rosenbaum, I. Goldenberg. A comparison of symptom determinants of patient and clinician global ratings in patients with panic disorder and depression. *Journal of Clinical Psychopharmacology*, 13 (5) (1993), pp. 327–331

Leucht and Engel, 2006. S. Leucht, R.R. Engel. The relative sensitivity of the clinical global impressions scale and the brief psychiatric rating scale in antipsychotic drug trials. *Neuropsychopharmacology*, 31 (2) (2006), pp. 406–412

Leucht et al., 2005. S. Leucht, J.M. Kane, W. Kissling, J. Hamann, E. Etchel, R. Engel. Clinical implications of brief psychiatric rating scale scores. *British Journal of Psychiatry*, 187 (2005), pp. 366–371

McGuire et al., 2012. J.F. McGuire, B.B. Kugler, J.M. Park, B. Horng, A.B. Lewin, T.K. Murphy, *et al.* Evidence-based assessment of compulsive skin picking, chronic tic disorders, and trichotillomania in children. *Child Psychiatry and Human Development*, 43 (2012), pp. 855–883

McGuire et al., 2014. J.F. McGuire, D. Ung, R.R. Selles, O. Rahman, A.B. Lewin, T.K. Murphy, *et al.* Treating trichotillomania: a meta-analysis of treatment effects and moderators for behavior therapy and serotonin reuptake inhibitors. *Journal of Psychiatric Research*, 58 (2014), pp. 76–83 http://dx.doi.org/10.1016/j.jpsychires.2014.07.015

Nelson et al., 2014. S.O. Nelson, K. Rogers, N. Rusch, L. McDonough, E.J. Malloy, M.J. Falkenstein, *et al.* Validating indicators of treatments response: application to Trichotillomania. *Psychological Assessment*, 26 (3) (2014), pp. 857–864

O'Sullivan et al., 1995. R.L. O'Sullivan, N.J. Keuthen, C.F. Haydah, J.N. Ricciardi, M.L. Buttolph, M.A. Jenike, *et al.* The Massachusetts-General-Hospital (MGH) Hairpulling Scale. 2. Reliability and validity. *Psychotherapy and Psychosomatics*, 64 (3–4) (1995), pp. 146–148

Pinsker, 1997. H. Pinsker. *A primer on supportive psychotherapy*. The Analytic Press, Hillsdayle, NJ (1997)

Rothbaum and Ninan, 1994. B.O. Rothbaum, P.T. Ninan. The assessment of trichotillomania. *Behaviour Research and Therapy*, 32 (6) (1994), pp. 651–662

Storch et al., 2011. E.A. Storch, A.S. De Nadai, A.B. Lewin, J.F. McGuire, A.M. Jones, P.J. Mutch, *et al.* Defining treatment response in pediatric tic disorders: a signal detection analysis of the yale global tic severity scale. *Journal of Child and Adolescent Psychopharmacology*, 21 (6) (2011), pp. 621–627 http://dx.doi.org/10.1089/cap.2010.0149

Storch et al., 2010. E.A. Storch, A.B. Lewin, A.S. De Nadai, T.K. Murphy. Defining treatment response and remission in obsessive-compulsive disorder: a signal detection analysis of the Children's Yale-Brown Obsessive Compulsive Scale. *Journal of the American Academy of Child*

and Adolescent Psychiatry, 49 (7) (2010), pp. 708–717 http://dx.doi.org/10.1016/j.jaac.2010.04.005

Spielmans and McFall, 2006. G.I. Spielmans, J.P. McFall. A comparative meta-analysis of the clinical global impressions change in antidepressant trials. *Journal of Nervous and Mental Disease*, 194 (11) (2006), pp. 845–854

Swets and Pickett, 1982. J.A. Swets, R.M. Pickett. *Evaluation of diagnostic systems: methods from signal detection theory*. Academic Press, New York, NY (1982)

Stanley et al., 1993. M.A. Stanley, R.C. Prather, A.L. Wagner, M.L. Davis, A.C. Swann. Can the yale-brown obsessive-compulsive scale be used to assess trichotillomania? A preliminary report. *Behaviour Research and Therapy,* 31 (2) (1993), pp. 171–177

Swedo et al., 1989. S.E. Swedo, H.L. Leonard, J.L. Rapoport, M.C. Lenane, E.L. Goldberger, D.L. Cheslow. A double-blind comparison of clomipramine and desipramine in the treatment of trichotillomania (hair pulling). *The New England Journal of Medicine*, 321 (8) (1989), pp. 497–501 http://dx.doi.org/10.1056/nejm198908243210803

Tolin et al., 2005. D.E. Tolin, J.S. Abramowitz, G.J. Diefenbach. Defining response in clinical trials for obsessive-compulsive disorder: a signal detection analysis of the yale-brown obsessive compulsive scale. *The Journal of Clinical Psychiatry*, 66 (12) (2005), pp. 1549–1557

Woods et al., 2015. Woods, D. W., Ely, L. J., Bauer, C. C., Twohig, M. P., Saunders, S. S., Compton, S. N., . . . and Franklin, M. E., Acceptance-enhanced behavior therapy for trichotillomania in adults: a randomized clinical trial, in preparation.

Woods and Houghton, 2014. D.W. Woods, D.C. Houghton. Diagnosis, evaluation, and management of trichotillomania. *Psychiatric Clinics of North America*, 37 (3) (2014), pp. 301–317

Woods and Twohig, 2008 .D.W. Woods, M.P. Twohig. *Trichotillomania: an ACT-enhanced behavior therapy approach therapist guide*. Oxford University Press (2008)

Woods et al., 2006. D.W. Woods, C.T. Wetterneck, C.A. Flessner. A controlled evaluation of acceptance and commitment therapy plus habit reversal for trichotillomania. *Behaviour Research and Therapy*, 44 (5) (2006), pp. 639–656

Winchel et al., 1992. R.M. Winchel, J.S. Jones, A. Molcho, B. Parsons, B. Stanley, M. Stanley. The psychiatric institute trichotillomania scale (PITS). *Psychopharmacology Bulletin*, 28 (1992), pp. 463–476

Zaider et al., 2003. T.I. Zaider, R.G. Heimberg, D.M. Fresco, F.R. Schneier, M.R. Liebowitz. Evaluation of the clinical global impression scale among individuals with social anxiety disorder. *Psychological Medicine*, 33 (4) (2003), pp. 611–622