



Published in final edited form as:

Contemp Clin Trials. 2013 November ; 36(2): 414–420. doi:10.1016/j.cct.2013.08.012.

Detecting a Clinically Meaningful Change in Tic Severity in Tourette Syndrome: A Comparison of Three Methods

Sangchoon Jeon, PhD^a, John T Walkup, MD^b, Douglas W. Woods, PhD^c, Alan Peterson, PhD^d, John Piacentini, PhD^e, Sabine Wilhelm, PhD^f, Lily Katsovich, MS, MBA^g, Joseph F. McGuire, MA^h, James Dziura, PhDⁱ, and Lawrence Scahill, MSN, PhD^j

^a School of Nursing, Yale University, New Haven, CT

^b Weill Cornell Medical College, New York, NY

^c Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI

^d Department of Psychiatry, University of Texas Health Science Center at San Antonio, TX

^e Psychiatry and Behavioral Sciences, University of California at Los Angeles, CA

^f Massachusetts General Hospital/Harvard Medical School, Cambridge, MA

^g Yale Child Study Center, New Haven, CT

^h Department of Psychology, University of South Florida, Tampa, FL

ⁱ Yale Center for Clinical Investigation, New Haven, CT

^j Department of Pediatrics, Emory University, Atlanta, GA

Abstract

Objective—To compare three statistical strategies for classifying positive treatment response based on a dimensional measure (Yale Global Tic Severity Scale [YGTSS]) and a categorical measure (Clinical Global Impression-Improvement [CGI-I]).

Method—Subjects (N=232; 69.4% male; ages 9–69 years) with Tourette syndrome or chronic tic disorder participated in one of two 10-week, randomized controlled trials comparing behavioral treatment to supportive therapy. The YGTSS and CGI-I were rated by clinicians blind to treatment assignment. We examined the percent reduction in the YGTSS-Total Tic Score (TTS) against Much Improved or Very Much Improved on the CGI-I, computed a signal detection analysis (SDA) and built a mixture model to classify *dimensional response* based on the change in the YGTSS-TTS.

© 2013 Elsevier Inc. All rights reserved.

Corresponding author: Lawrence Scahill, MSN, PhD Professor of Pediatrics, Emory University School of Medicine; Marcus Center, 1920 Briarcliff Rd. Atlanta, GA 30029. lawrence.scahill@emory.edu.

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

Clinicaltrials.gov Identifiers: NCT00218777; NCT00231985

Results—A 25% decrease on the YGTSS-TTS predicted *positive response* on the CGI-I during the trial. The SDA showed that a 25% reduction in the YGTSS-TTS provided optimal sensitivity (87%) and specificity (84%) for predicting positive response. Using a mixture model without consideration of the CGI-I, the *dimensional response* was defined by 23% (or greater) reduction on the YGTSS-TTS. The odds ratio (OR) of *positive response* (OR=5.68, 95% CI=[2.99, 10.78]) on the CGI-I for behavioral intervention was greater than the *dimensional response* (OR=2.86, 95% CI=[1.65, 4.99]).

Conclusion—A twenty five percent reduction on the YGTSS-TTS is highly predictive of *positive response* by all three analytic methods. For trained raters, however, tic severity alone does not drive the classification of positive response.

Keywords

Tourette Syndrome; Yale Global Tic Severity Scale; Clinical Global Impression; Cognitive Behavioral Intervention; Signal detection analysis; mixture model; treatment response

INTRODUCTION

Tourette Syndrome (TS) is defined by persistent motor and vocal tics beginning before age 18 years. Motor tics are usually brief, rapid movements of the face, shoulders and upper extremities, but may involve more complex and purposeful movements. Common vocal tics include throat clearing, grunting or coughing; complex vocal tics such as shrieks, words, parts of words or cursing occur in a minority of patients. The prevalence of TS in school-age children is estimated at 6 per 1000 [1]. In community and clinically-ascertained samples, children with TS have high rates of disruptive behavior and attention-deficit/hyperactivity disorder (ADHD) [1,2].

Although several instruments [3,4] have been developed to measure tic severity, the Yale Global Tic Severity Scale (YGTSS) is the most commonly used measure in clinical trials [5]. This multi-dimensional, clinician-rated measure of tic severity has established reliability and validity [6,7]. In addition to tracking tic severity in clinical trials, investigators and clinicians may be interested in the proportion of subjects showing a positive response in the active treatment group compared to a control condition [8,9]. The clinician-rated Clinical Global Impression-Improvement (CGI-I) Scale is commonly used to assess overall change compared to baseline [10,11,12]. The CGI-I permits the rater to consider all available information (subjective report from the participant, reports from close family members and direct observation) in the assessment of change [13,14]. In multisite trials, it has been shown that CGI-I training can reduce variability due to raters [13,14,15].

Identifying the magnitude of tic symptom reduction on a dimensional measure associated with positive response on a categorical measure such as the CGI-I can facilitate comparison of treatments across clinical trials and also guide clinical assessment of interventions in routine patient care. We identified 18 randomized medication trials targeting tics that used the Yale Global Tic Severity-Score (YGTSS-TTS) as an outcome measure [5,16]. Of these, only 9 studies included more than 30 subjects and few used the CGI-I as an outcome measure. Thus, the opportunity to determine the change score on the YGTSS-TTS

associated with a positive response has been limited. Only one previous report explored the association of change on the YGTSS-TTS with improvement on the CGI-I in 108 youth with chronic tic disorders. This report identified a 35% reduction or an absolute drop of 6-7 points on the YGTSS-TTS corresponded to a rating of improvement on the CGI-I [5]. Although informative, this report compiled data from children who participated in randomized clinical trial showing no difference between drug and placebo, unblinded assessment of subjects in an open trial and unblinded assessment of patients in standard clinical care. The purpose of this study is examine the level of improvement on the YGTSS-TTS that predicts positive response on the CGI-I.

METHOD

Design

The sample was derived from two multisite, behavioral intervention trials that used the same randomized design to evaluate the efficacy of a Comprehensive Behavioral Intervention for Tics (CBIT) compared to Psychoeducation and Supportive Therapy (PST). The first trial enrolled 126 subjects (age 9 to 17 years) [17]; the second trial included 122 subjects (age 16 to 69 years)[18]. CBIT and PST each followed a structured therapy manual that was delivered in 8 sessions over 10 weeks by therapists trained to reliability [19]. The primary outcome measures were the YGTSS-TTS and the Improvement item of the Clinical Global Impression (CGI-I) scale. Treatment outcomes were assessed at baseline, Week 5 and Week 10 by an independent evaluator at each site who was blind to group assignment. Blinding of the therapist, subject and family was not possible.

Setting and Subjects

The child trial included 126 subjects (age 9 to 17 years) enrolled at Johns Hopkins University, University of Wisconsin at Milwaukee or University of California at Los Angeles. The adult trial enrolled 122 subjects (age 16 to 69 years) at Harvard University, University of Texas Health Sciences Center at San Antonio or Yale University. The trials were approved by the institutional review board at each site. Parental permission for minors or consent for adult participants was obtained prior to formal data collection.

Participants could be on medication for tics or for a co-occurring condition (e.g., attention deficit hyperactivity disorder, obsessive-compulsive-disorder (OCD) or generalized anxiety disorder) if the medication was stable for at least six weeks with no planned changes for the 10-week trial. Subjects with IQ below 80, with another psychiatric condition in need of treatment, a lifetime history of psychosis or pervasive developmental disorder were excluded. To be eligible, subjects had to have tics of at least moderate severity (CGI-Severity score of 4 or greater). In the child study, the YGTSS-TTS had to be > 13 for subjects with motor and vocal tics (subjects with motor or vocal tics only could enter with a YGTSS score > 9). In the adult trial, the YGTSS-TTS had to be > 14 for subjects with motor and vocal tics (subjects with motor or vocal tics only could enter with a YGTSS score > 10).

Measures

The pre-treatment eligibility assessment was similar in each trial and included collection of demographic information, medical history, age-appropriate structured psychiatric interviews, as well as symptom severity for tics, obsessions and compulsions and ADHD. This report focuses on the YGTSS and the CGI-I. .

Yale Global Tic Severity Scales—(YGTSS). The YGTSS is a clinician-rated scale used to assess tic severity over the prior week. It includes a checklist of motor and vocal tics followed by an assessment of the number, frequency, intensity, complexity, and inference of motor tics and phonic tics – scored separately. Each of these dimensions is scored on a 0 to 5 scale. The YGTSS provides three tic severity scores: Total Motor (0 to 25); Total Phonic (0 to 25) and the combined Total Tic Severity Score (0 to 50), as well as a separate Impairment dimension scored from 0 to 50. The Total Tic Score (YGTSS-TTS) was used in these analyses.

Clinical Global Impression scales—(CGI). The CGI includes a Severity scale (CGI-S) and an Improvement scale (CGI-I). The CGI-S is a seven-point scale that ranges from 1 (Normal) to 7 (Extreme) [10]. A score of 3 (Mild) reflects the presence of symptoms with little or no impairment; scores of 4 (Moderate) and higher (5=Severe; 6=Marked; 7=Extreme) reflect greater symptom acuity and impairment. Although tics were strongly weighted in the assessment of overall severity, raters used all available information to score the CGI-S. The CGI-I is also a seven-point scale that is used to rate overall change from baseline. Scores range from 1 (Very Much Improved) through 4 (No Change) and 7 (Very Much Worse). In this analysis, a score of 2 (Much Improved) or 1 (Very Much Improved) defined *positive response*.

Rater Training

Prior to subject enrollment, experienced clinical investigators (LS, JTW) trained the independent evaluators (IEs) for the study. The CGI (Severity and Improvement scales) and the YGTSS were described in detail. Case vignettes for the CGI and video recordings for the YGTSS were used for didactic purposes. As with the CGI-S, raters were instructed to weight tics in the assessment of change from baseline, but raters were also encouraged to consider all available information when scoring the CGI-I. Raters then independently scored the CGI (Severity and Improvement) on three new vignettes and the YGTSS on three new video recordings. To be considered reliable, raters had to be within one unit on the CGI (Severity and Improvement) of the gold standard rating without disagreement on eligibility or treatment response. For example, a disagreement on the CGI-S of 4 (Moderate) versus 3 (Mild) at baseline is one unit difference on the scale, but would affect study eligibility and was considered unreliable. On the CGI-Improvement, if the expert rater assigned a score of 2 (Much Improved) and the new rater gave a score of 3 (Minimally Improved), this would affect the classification of treatment response and was regarded as unacceptable. For the YGTSS, raters had to be within 15% of the experienced rater for the Total Motor Tic score, the total Vocal Tic score and the Total Tic Score on three recordings. IEs who did not meet criteria on the YGTSS or CGIs were given additional training and asked to score additional CGI vignettes or YGTSS video recordings until agreement criteria were met.

IEs also participated in monthly conference calls led by an experienced rater (LS). These calls provided a forum for discussing cases and developing a common approach to conducting assessments across sites. Raters who joined the trial in progress were trained in a similar manner (using email and telephone). New raters were also required to send their initial YGTSS recordings for qualitative review by an experienced rater (LS).

Interventions

CBIT is a practice-based cognitive behavioral intervention intended to improve the patient's ability to manage tics [17,18,19]. It consists of three central components: tic-awareness training, competing-response training, and functional analysis. Awareness training promotes early detection of tic signals (e.g., unwanted premonitory urges or warnings that often precede tics). Competing response training teaches the patient to engage in a deliberate behavior in place of the tic upon detection of the impending tic. Because the tic often relieves the premonitory sensation, blocking the tic with a voluntary action is intended to break the link between the unwanted sensation and performance of the tic. Functional analysis fosters identification of the situations and social reactions that heighten tic severity and development of management strategies to neutralize these events.

PST provides current information about tic disorders including the natural history, genetics, pathophysiology and available treatments for tics [17,18]. In the child study, parents were included in PST sessions. PST did not include any discussion of tic management strategies.

Data Analysis

Method 1, Area under the receiver operating curve—To examine the association between the YGTSS-TTS and the CGI-I, the change on the YGTSS-TTS from baseline to Week 10 was compared across the full range (Very Much Improved through Very Much Worse) of IE-rated CGI-I scores. The area under the receiver operating characteristic curve (AUC) was used to evaluate the discriminative accuracy of the change in the YGTSS-TTS to predict *positive response* (Much Improved or Very Much Improved on the CGI-I). An AUC of 0.5 indicates that a given percent reduction on the YGTSS-TTS is no better than chance for predicting positive response on the CGI-I. A score of 1 represents perfect classification. An AUC > 0.8 is considered excellent discrimination [20]. We also assessed the probabilities of four CGI-I ratings (Very Much Improved, Much Improved, Minimally Improved, or No Change and all ratings reflecting worsening) across a range of percent reductions on the YGTSS-TTS. The probabilities of these CGI-I ratings were estimated with generalized multinomial logit model using PROC LOGISTIC with GLOGIT option in SAS version 9.2.

Method 2, Signal detection analysis—Signal detection analysis (SDA) was used to identify the optimal percent reduction of YGTSS-TTS against the CGI-I score of 2 (Much Improved) and 1 (Very Much Improved). Starting with 15% reduction on the YGTSS-TTS and moving up in 5% increments, we calculated the sensitivity and specificity to predict *positive response*. The optimal cutoffs were determined by comparing the 2×2 weighted kappa statistic [21], for a given reduction of YGTSS-TTS score and *positive response* on the CGI-I. The weighted kappa statistic (ranges from 0 to 1) measures the level of agreement

between positive response on the CGI-I and the percent reduction on YGTSS-TTS. Higher scores on the kappa reflect better agreement level.

Method 3, Mixture model— We also constructed a *Response Model* based on change in the YGTSS-TTS regressed on baseline score and measurement errors to develop a *dimensional response* (Appendix A). The parameters in *Response Model* were estimated with a mixture model using PROC NLMIXED in SAS version 9.2 (Appendix B). Odds ratios were calculated to compare the relative sensitivities of the three approaches (CGI-I, SDA, and Response Model) for detecting the efficacy of CBIT against PST.

RESULTS

Complete baseline data and complete Week 10 CGI-I data were available on 232 of the 248 randomized participants. Demographic and clinical characteristics by treatment group are shown in Table 1. There were no significant differences between the two treatment groups. Half of the 232 subjects in this analysis were not on any medication; 73 (31.5%) were on a tic medication with no difference by treatment group. CGI-I data were missing for 16 subjects (7 in CBIT and 9 in PST) and were not included in the analysis. At endpoint, IEs were asked to guess the treatment assignment for each subject. The IEs guess was correct for 62% and 72% of subjects in CBIT and PST, respectively. This difference was not significant (Chi square =2.52, $p=0.11$).

To control for study, we used the Cochran-Mantel-Haenszel (CMH) Chi square test. On the CGI-I at Week 10, 56 subjects (47.9%) in CBIT showed a *positive response* (Much Improved or Very Much Improved) compared to 16 subjects (13.7%) in PST which was significantly different (CMH Chi square = 32.78; $p < .0001$). If the subjects with missing data were included and classified as “non-responders,” the difference between CBIT and PST remained significant (45.2% vs. 12.9%, CMH Chi square=32.40; $p<.0001$).

Considering the available sample as a whole ($N=232$), subjects rated with *positive response* on the CGI-I and those who did not show a positive response had similar YGTSS-TTS at baseline (24.5 ± 6.5 and 23.4 ± 6.4 , respectively, $p= .21$). In a regression model controlling for baseline score, age, gender, and tic medication status, subjects with a positive response had significantly greater reduction on the YGTSS-TTS (10.6 ± 5.0) compared to those with non-response (2.3 ± 4.4) ($p < .0001$).

Area under the receiver operating curve to evaluate the association of reduction in YGTSS-TTS and original CGI-I ratings

The AUC of 0.91 (95% CI=[0.88, 0.95]) indicates outstanding discriminative accuracy for the change in the YGTSS-TTS to predict CGI-I classification. Figure 1 shows the mean YGTSS-TTS at baseline and Week 10 against CGI-I ratings at Week 10. With the exception of YGTSS-TTS over 30, tic severity at baseline was not predictive of positive response on the CGI-I. The generalized logits multinomial model estimated probability for each CGI-I level against percent reduction on the YGTSS-TTS. The probability of *positive response* (i.e., Much Improved or Very Much Improved steadily increases as the change in YGTSS-TTS exceeds 25%. Moreover, the probability of “No Change” or “Worsening” on the CGI-I remains high with percent reductions less than 25% on the YGTSS-TTS.

Means and standard deviations of YGTSS-TTS at baseline and Week-10 are shown in Table 2. Except for a rating of Much Worse, the baseline YGTSS-TTS were not different by CGI-I rating. Not surprisingly, the largest reduction on the YGTSS-TTS was observed for subjects classified as Very Much Improved (from 22.81 ± 5.82 to 8.00 ± 6.68) followed by subjects classified as Much Improved (from 25.06 ± 5.14 to 16.22 ± 4.45). The average percent reductions were 67.1% (SD=19.3) and 35.2% (SD=14.3) in Very Much Improved and Much Improved, respectively.

Signal Detection Analysis—Table 3 presents the sensitivities and specificities for percent change (baseline to Week 10) on the YGTSS-TTS to predict *positive response* on the CGI-I across the entire sample (N=232). The weighted kappa statistics [$\kappa(0.5)$] used to identify the balance of false negatives and false positives showed maximum efficiency at 25% [$\kappa(0.5)=0.68$] reduction of YGTSS-TTS. As shown in Table 3, this cutoff resulted in a sensitivity of 87% and specificity was 84%.

Mixture model to define a dimensional response model—The mixture model was used to estimate the percent reduction on the YGTSS-TTS that defines positive response without reference to the CGI-I. After plugging in the YGTSS-TTS at baseline and Week 10 into the *Response Model* for each subject, 83 of 232 participants (47.0 %) met criterion for the *dimensional response*. The minimum percent reduction on the YGTSS-TTS for subjects meeting the *dimensional response* was 23%; the mean was 34%. By contrast, an 8.7% reduction on the YGTSS-TTS was associated with non-response. Figure 2 shows the scatter plots of the YGTSS-TTS against the baseline score for *positive response* (on the CGI-I) and *dimensional response* (*Response Model*) at Week 10. Recalling that raters considered all available information (i.e., tic severity, distress, impairment) when scoring the CGI-I, the overlap of the plots for *positive response* (black circles) and *non-response* (open circles) is not surprising. By contrast, the *dimensional response*, based only on change in tic severity, shows less overlap. Nonetheless, the slopes of YGTSS-TTS from baseline to Week 10 were not much different for the *dimensional response* (Slope=0.72, 95% CI=[0.54, 0.90]) and *positive response* (Slope=0.76, 95% CI=[0.54, 0.97]).

Table 4 shows the positive response rates and odds ratios (OR) with 95% confidence intervals for CBIT versus PST across the three classification methods. The OR of *positive response* (original CGI-I) for CBIT was 5.65 (95% CI=[2.97, 10.74]). The ORs were lower in the SDA approach (3.27, 95% CI=[1.87, 5.72]) and *Response Model* analysis (2.86, 95% CI=[1.65, 4.99]). These lower ORs are likely due to the higher of positive response rates for PST in the SDA and *Response Model* compared to the original CGI-I (see Table 4).

Sixty subjects (51.3%) in CBIT met the benchmark of 25% or greater improvement on the YGTSS-TTS at Week 10, compared to 24.3% (n=28) for PST (Chi-square = 17.87, $p < .0001$). In CBIT, 80% (48/60) of those who met the 25% improvement benchmark were rated as Much Improved or Very Much Improved on the CGI-I compared to 53.6% (15/28) of subjects in PST. If the CGI-I is conditioned on a 25% or greater decline on the YGTSS-TTS, subjects in CBIT were 3.47 times more likely (Chi square 6.56, $p = 0.01$; OR=3.47, 95% CI=[1.31, 9.20]) to be rated Much Improved or Very Much Improved than subjects in PST.

DISCUSSION

In addition to effect size and the proportion of subjects showing a positive response, investigators and clinicians may be interested in the change on a dimensional measure that predicts *positive response*. This information promotes comparison of results across trials and provides a benchmark for patients in clinical settings. This secondary analysis combined samples from two 10-week, multisite, randomized trials of compared the efficacy of CBIT to PST on tics. The data set of 232 subjects with complete data (93.5% of the total combined sample) was used to compare three methods of mapping change on the YGTSS-TTS and positive treatment response. In the behavior therapy group, 56 of 117 (47.9%) subjects were classified by blinded raters as Much Improved or Very Much Improved on the CGI-I compared to 16 of 115 (13.9%) in the comparison condition.

Across both treatment groups, subjects rated Much Improved or Very Much Improved on the CGI-I showed a greater than four-fold reduction in YGTSS-TTS (10.6 ± 5.0) compared to those who did not show a positive response (2.3 ± 4.4). These results indicate that positive response on the CGI-I was mostly explained by the improvement of YGTSS-TTS (AUC=91%). A rating of Much Improved or Very Much Improved on the CGI-I was strongly associated with at least a 25% reduction on the YGTSS-TTS (Figure 1). However, 12 subjects in CBIT and 13 subjects in PST (10.8% of the entire sample) with 25% or greater improvement on the YGTSS-TTS were not rated Much Improved or Very Much Improved. This finding suggests that change in tic severity alone did not determine CGI-I ratings. Nonetheless, the dimensional measure and the categorical measure were highly consistent in these trials and the study results are not ambiguous.

The SDA approach showed that a 25% reduction in the YGTSS-TTS also provided the optimal balance of sensitivity and specificity for predicting *positive response* on the CGI-I of Much Improved or Very Much Improved. This percent decline is similar to the reduction on the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) in sample of adults with OCD (30% reduction) and a sample of children assessed on the Children's Y-BOCS in OCD (25% reduction) [8, 9]. Our finding of 25% reduction is lower than the 35% reduction on the YGTSS-TTS reported by Storch and colleagues [22]. In that study of 108 pediatric subjects, about half were ascertained from a negative, placebo-controlled trial, about a third participated in an open trial and the remainder participated in regular clinical care. This heterogeneous sampling frame may explain the higher threshold of 35% reduction. In addition, unblinded assessments were conducted in about half of the sample, raising the possibility of bias in the rating. Not surprisingly, the sensitivity (87%) at the 25% reduction threshold on the YGTSS-TTS in the current study was higher than the sensitivity of 70% for the 35% reduction benchmark reported by Storch [22]. The specificities across the two studies (84% in the current study) and 87% in the prior report were similar.

The odds ratio of *positive response* for CBIT versus PST was 5.65 (95% CI=[2.97, 10.74]) for the CGI-I compared to 3.28 and 2.82 for SDA and the *Dimensional Response Model*, respectively. This was due, at least in part, to the lower *positive response* rate in the PST group on the original CGI-I. As suggested by the high specificity, participants with less than 25% reduction in the YGTSS-TTS were rarely classified as achieving *positive response* in

either intervention group. However, 53.6% (15 of 28) of subjects with 25% or greater reduction in YGTSS-TSS in PST were rated with positive response compared to the 80% (48 of 60) in the CBIT group. As a practice-based behavioral intervention, CBIT may have promoted greater self-efficacy and greater confidence in subjects – which was detected by raters. The blinded independent evaluators correctly guessed the treatment condition in 72% of PST participants (83 of 115) compared to 62% (73 of 117) for the CBIT condition. However, this difference was not significant, suggesting that raters did not simply associate lack of efficacy with PST and positive response with CBIT. This finding also suggests that treatment assignment was indeed blinded.

The mixture model was used to identify treatment response based only on the change of YGTSS-TTS (see Appendix A). This model provided consistent results with the other analytic approaches. Although 24.4% of subjects in PST achieved this dimensional response criterion compared to a positive response of 13.9% on the CGI-I, the rate of *positive response* in the CBIT group was similar across the three analytic approaches. Taken together, the 25% reduction in the YGTSS-TTS appears to be a valid benchmark for clinically meaningful improvement. These findings were based on a combined sample of two psychotherapy trials. Although each of these studies was larger than any prior drug study focused on tic severity, these results may not apply to clinical trials in psychopharmacology.

Acknowledgments

The authors acknowledge the efforts of Cynthia Brandt, MD, MPH, Stephanie Argraves, MS, Haibei Liu, PhD, Allison Gavaletz, BS for data management, and Heather Cowley, PhD for administrative support.

Financial Disclosures: Drs Wilhelm, Peterson, Piacentini, Woods, Walkup and Scahill report receiving royalties from Oxford University Press for treatment manuals on tic disorders. Drs Wilhelm, Peterson, Piacentini, Woods, Walkup, and Scahill report receiving honoraria for continuing education presentations from the Tourette Syndrome Association. Drs Piacentini, Woods, and Walkup receive royalties from Guilford Press for a book on Tourette disorder. Dr Wilhelm reports receiving support in the form of free medication and matching placebo from Forest Laboratories for clinical trials funded by the National Institutes of Health (NIH) and receiving book royalties from Guilford Publications, New Harbinger Publications, and Oxford University Press. Dr Piacentini reports receiving royalties from Oxford University Press for treatment manuals on child obsessive-compulsive disorder. Dr Woods reports receiving book royalties from New Harbinger and Springer Publications. Dr. Scahill has received royalties from Oxford University Press and American Psychiatric Press, has served as a consultant for BioMarin, Roche, Pfizer and Bracket and has had research support from Shire Pharmaceutical, Roche and Pfizer. Dr Walkup reports receiving consulting fees from Shire Pharmaceuticals. He reports receiving free drug and matching placebo from Pfizer and Lilly and free drugs from Abbott for NIMH-funded clinical trials. Dr. Jeon, Ms. Katsoyich and Mr McGuire report no financial interests.

Funding/Support: This work was supported by grants from the National Institute of Mental Health (NIMH) to Drs. Wilhelm (5R01MH069877), Scahill (R01MH069874), and Peterson (R01MH069875) with subcontracts to Drs. Piacentini and Woods. Dr. Walkup consulted on this grant (ClinicalTrials.gov identifier: NCT00231985).

This work was also supported by grant R01MH070802 (clinicaltrials.gov Identifier: NCT00218777) from the National Institute of Mental Health to Dr. Piacentini, with subcontracts to Drs. Woods, Scahill, Wilhelm, Peterson, and Walkup. Drs Scahill and Jeon receive support from the Yale University Clinical and Translational Sciences Award grant (UL1 RR024139) from the National Center for Research Resources, NIH. Role of the Sponsor: The funding organization was not involved in the design or conduct of the study; collection, management, analysis, and interpretation of the data or preparation, review, and approval of the manuscript. This study was also supported by the Tourette Syndrome Association to Dr. Scahill.

APPENDIX A

Response-Model For given the baseline YGTSS-TTS (y_1), Week-10 YGTSS Total Tic Score (y_2) is predicted as following models;

$$\begin{aligned} Y_2 &= (1 - \beta) y_1 + \varepsilon_N \quad \text{if } \textit{non-response}, \text{ and} \\ Y_2 &= (1 - \beta - \delta) y_1 + \varepsilon_R \quad \text{if } \textit{dimensional response}, \end{aligned} \quad \text{Eq. (A.1)}$$

where the error terms ε_N and ε_R have normal distribution with zero mean and standard deviations σ_N and σ_R . The location parameter β represents average reduction of Total Tic Score for *non-response* at Week 10 and other location parameter δ represents additional average reductions to be rated as *dimensional response*.

Then the YGTSS-TTS has a mixture of two normal distributions as following;

$$g(y_2|y_1) = \theta f[(1 - \beta) y_1, \sigma_N] + (1 + \theta) f[(1 - \beta - \delta) y_1, \sigma_R], \quad \text{Eq. (A.2)}$$

where θ is probability of being *non-response* and $f(a, b)$ is a normal distribution with mean a and standard deviation of b . The parameters can be estimated using Maximum Likelihood (ML) Method.

APPENDIX B

```
PROC NLMIXED DATA=Final;

PARMS b1=0.2 sigma1=3 sigma2=3 r1=0.1 a1=0.5 ;

t=YGTSS_total10-(1-b1)*YGTSS_total0 ;

p=a1*pdf('normal',t,0,sigma1)+(1-a1)*pdf('normal',t,-r1*YGTSS_Total0,sigma2);

ll=log(p);

model t~general(ll);

Predict p out=YGTSS10;

Estimate 'beta' b1;

Estimate 'beta+delta' b1+r1;

RUN;
```

REFERENCES

1. Tabori-Kraft J, Dalsgaard S, Obel C, Thomsen PH, Henriksen TB, Scahill L. Prevalence and clinical correlates of tic disorders in a community sample of school-age children. *Eur Child Adolesc Psychiatry*. 2012; 21(1):5–13. [PubMed: 22038343]
2. Sukhodolsky DG, Scahill L, Zhang H, Peterson BS, King RA, Lombroso PJ, Katsovich L, Findley D, Leckman JF. Disruptive Behavior in Children with Tourette's Syndrome: Association with

- ADHD Comorbidity, Tic Severity, and Functional Impairment. *J Am Acad Child Adolesc Psychiatry*. 2003; 42(1):98–105. [PubMed: 12500082]
3. Shapiro AK, Shapiro E. Controlled study of pimozide vs. placebo in Tourette syndrome. *J Am Acad Child Psychiatry*. 1984; 23:161–173. [PubMed: 6371107]
 4. Harcherik DF, Leckman JF, Detlor J, Cohen DJ. A new instrument for clinical studies of Tourette's syndrome. *J Am Acad Child Psychiatry*. 1984; 23:153–160. [PubMed: 6585417]
 5. Scahill, L.; King, RA.; Lombroso, P.; Sukhodolsky, DG.; Leckman, JF. Assessment and treatment of Tourette syndrome and other tic disorders.. In: Martin, A.; Scahill, L.; Kratochvil, CJ., editors. *Pediatric Psychopharmacology: Principles and Practice*. 2nd ed.. Oxford University Press, Inc.; New York, NY: 2011. p. 516-530.
 6. Leckman JF, Riddle MA, Hardin MT, Ort SI, Swartz KL, Stevenson J, Cohen DJ. The Yale Global Tic Severity Scale: Initial Testing of a Clinician-Rated Scale of Tic Severity. *J Am Acad Child Psychiatry*. 1989; 28(4):566–573.
 7. Storch EA, Murphy TK, Geffken GR, Sajid M, Pam A, Roberti J, Goodman WK. Reliability and validity of the Yale Global Tic Severity Scale. *Psychological Assessment*. 2005; 17(4):486–491. [PubMed: 16393016]
 8. Tolin DF, Abramowitz JS, Diefenbach GJ. Defining response in clinical trials for obsessive-compulsive disorder: a signal detection analysis of the Yale-Brown obsessive compulsive scale. *J Clin Psychiatry*. 2005; 66(12):1549–1557. [PubMed: 16401156]
 9. Storch EA, Lewin AB, De Nadai AS, Murphy TK. Defining Treatment Response and Remission in Obsessive-Compulsive Disorder: A Signal Detection Analysis of the Children's Yale Brown Obsessive Compulsive Scale. *J Am Acad Child Psychiatry*. 2010; 49(7):708–717.
 10. Guy, W. *ECDEU Assessment Manual for Psychopharmacology*, revised. National Institute of Mental Health; Rockville, MD: 1976. Clinical Global Impressions..
 11. Calabrese JR, Keck PE, Macfadden W, Minkwitz M, Ketter TA, Weisler RH, Cutler AJ, McCoy R, Wilson E, Mullen J. A Randomized, Double-Blind, Placebo-Controlled Trial of Quetiapine in the Treatment of Bipolar I or II Depression. *Am J Psychiatry*. 2005; 162:1351–1360. [PubMed: 15994719]
 12. Kumra S, Kranzler H, Gerbino-Rosen G, Kester HM, DeThomas C, Kafantaris V, Correll CU, Kane JM. Clozapine and “High-Dose” Olanzapine in Refractory Early-Onset Schizophrenia: A 12-Week Randomized and Double-Blind Comparison. *Biol Psychiatry*. 2008; 63:524–529. [PubMed: 17651705]
 13. Zaider TI, Heimberg RG, Rresco DM, Schneier FR, Liebowitz MR. Evaluation of the Clinical Global Impression Scale among individuals with social anxiety disorder. *Psychological Medicine*. 2003; 33:611–622. [PubMed: 12785463]
 14. Targum SD, Busner J, Young AH. Targeted scoring criteria reduce variance in global impressions. *Hum Psychopharmacol Clin Exp*. 2008; 23:629–633.
 15. Leon AC, Shear K, Klerman GL, Portera L, Rosenbaum JF, Goldenberg I. A comparison of symptom determinants of patient and clinician global ratings in patients with panic disorder and depression. *J Clin Psychopharmacol*. 1993; 13(5):327–331. [PubMed: 8227491]
 16. Kurlan R, Crespi G, Coffey B, Mueller-Vahl K, Koval S, Wunderlich G. A multicenter randomized placebo-controlled clinical trial of pramipexole for Tourette's syndrome. *Movement Disorder*. 2012; 27(6):775–778.
 17. Piacentini J, Woods DW, Scahill L, Wilhelm S, Peterson AL, Chang S, Ginsburg GS, Deckersbach T, Dziura J, Levi-Pearl S, Walkup JT. Behavior Therapy for Children With Tourette Disorder: A Randomized Controlled Trial. *J Am Med Assoc*. 2010; 303(19):1929–1937.
 18. Wilhelm S, Peterson AL, Piacentini J, Woods DW, Deckersbach T, Sukhodolsky DG, Change S, Liu H, Dziura J, Walkup JT, Scahill L. Randomized Trial of Behavior Therapy for Adults with Tourette's Disorder. *Arch Gen Psychiatry*. 2012; 69(8):795–803. [PubMed: 22868933]
 19. Woods DW, Piacentini J, Chang S, Deckersbach T, Ginsburg GS, Peterson AL, Scahill L, Walkup JT, Wilhelm S. *Managing Tourette Syndrome: A Behavioral Intervention for Children and Adults. Therapist Guide*. New York: Oxford University Press. 2008
 20. Hosmer, DW.; Lemeshow, S. *Applied logistic regression*. 2nd ed.. John Wiley & Sons, Inc.; Hoboken, NJ: 2000.

21. Gilchrist JM. Weighted 2×2 kappa coefficients: recommended indices of diagnostic accuracy for evidence-based practice. *J Clin Epidemiol.* 2009; 62:1045–1053. [PubMed: 19278830]
22. Storch EA, De Nadai AS, Lewin AB, McGuire JF, Jones AM, Mutch PJ, Shytle RD, Murphy TK. Defining Treatment Response in Pediatric Tic Disorders: A Signal Detection Analysis of the Yale Global Tic Severity Scale. *J Child Adolesc Psychopharm.* 2011; (6):621–627. 2011.

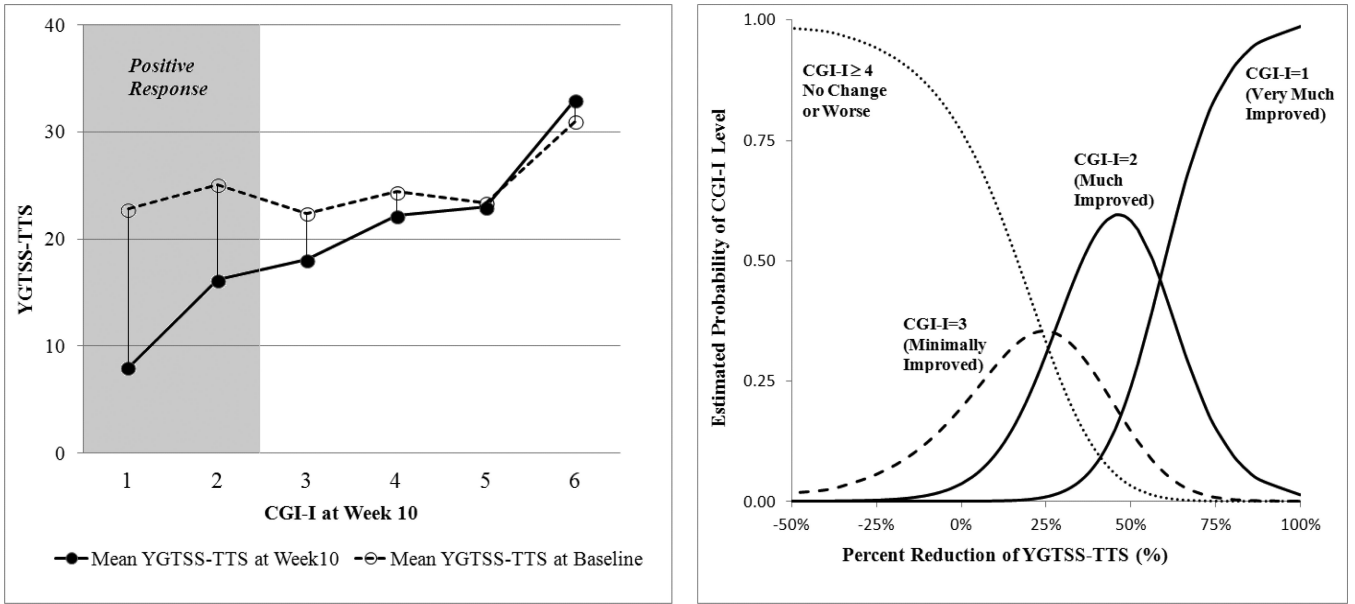
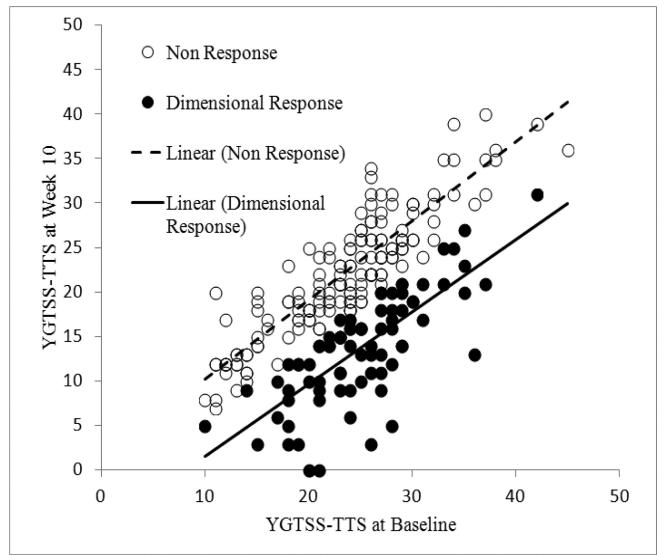
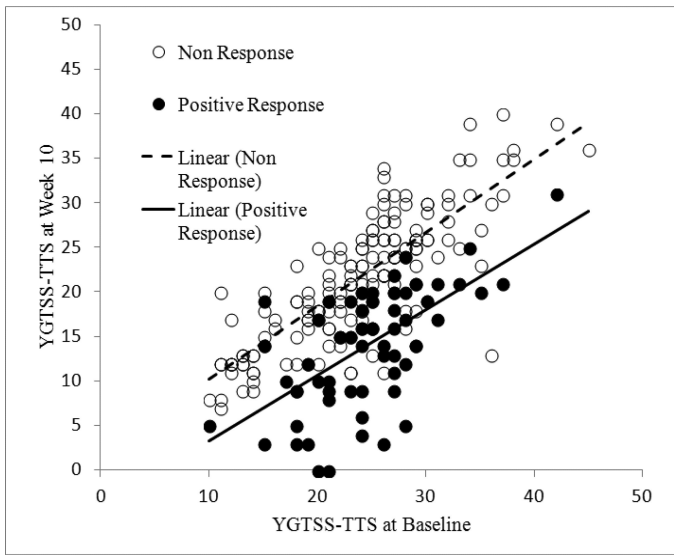


Figure 1. Mean YGTSS-TTS against CGI-I ratings and the estimated probability of CGI-I score based on the percent reduction on the YGTSS-TTS at Week 10 for the entire sample (N=232)

Figure on the left indicates the mean of YGTSS-TTS at baseline (open circles) and Week 10 (black circles). The figure on the right shows the probability of CGI-I score at Week 10 derived from generalized logits model based on the percent reduction on YGTSS-TTS from baseline. For example, a 25% reduction in the YGTSS-TTS has about a 30% chance of being rated Much Improved; a 50% reduction in the YGTSS-TTS has 60% chance of being rated Much Improved. Positive response was defined as Much Improved or Very Much Improved. Thus, the probability of positive response is equal to the probability of Much

Improved (CGI=2) plus the probability of Very Much Improved (CGI=1). the probability of CGI=2 decreases as the improvement in the YGTSS-TTS exceeds 50%, but the probability of CGI=1 incrementally increases with greater than 50% improvement in YGTSS. Therefore, the summed probabilities of the CGI=2 and CGI=1 increases as the decline in the YGTSS-TTS exceeds 50%.



(a) Positive Response based on CGI-I at Week10

(b) Dimensional Response based on Response Model at Week10

Figure 2. Treatment response based on (CGI-I) and the Response Model at Week 10
The low YGTSS-TTS at baseline come from subjects (31.5%) with only motor or vocal tics.

Table 1

Demographic and Clinical Characteristics of Combined CBIT Trials by Treatment Group (N=232)

		CBIT n=117	PST n=115
		N (%)	N (%)
Gender	Male	79 (67.5%)	85 (73.9%)
	Female	38 (32.5%)	30 (26.1%)
Tic Medication Use	No Medication	64 (54.7%)	60 (52.2%)
	Tic Medication	37 (31.6%)	36 (31.3%)
	Other Medication	16 (13.7%)	19 (16.5%)
		Mean ± SD	Mean ± SD
Age		22.4 ± 14.1	21.7 ± 14.2
Motor Tic Severity (YGTSS) at baseline		15.2 ± 4.0	14.8 ± 3.0
Vocal Tic Severity (YGTSS) at baseline		9.3 ± 4.9	8.7 ± 5.1
YGTSS-TTS		24.5 ± 6.5	23.4 ± 6.4
Impairment (YGTSS) at baseline		24.3 ± 8.4	23.6 ± 7.6

Positive Response ^(a) was defined by score of Much Improved or Very Much Improved on Clinical Global Impressions (CGI)-Improvement scale at 10 week.

Table 2

Means and standard deviation of YGTSS-TTS at baseline and Week-10 by CGI-I rating across the entire sample (N=232)

	N	YGTS-TTS at Baseline Mean \pm SD	YGTS-TTS at Week-10 Mean \pm SD	Percent Reduction at Week-10 Mean \pm SD %
Very Much Improved	21	22.81 \pm 5.82	8.00 \pm 6.68	67.1 \pm 19.3 %
Much Improved	51	25.06 \pm 5.14	16.22 \pm 4.45	35.2 \pm 14.3 %
Minimally Improved	53	22.40 \pm 6.28	18.08 \pm 5.65	19.2 \pm 13.3 %
No Change	58	24.45 \pm 7.33	\pm 6.82	12.0 \pm 13.8 %
Minimally Worsen	42	23.40 \pm 6.65	23.00 \pm 6.73	5.3 \pm 9.5 %
Much Worsen	7	31.00 \pm 6.11	33.29 \pm 5.79	0.7 \pm 1.9 %

Table 3

Signal Detection Analysis predicting *positive response* based on percent reduction on YGTSS-TTS from baseline to Week-10

YGTSS-TTS Reduction (%)	Sensitivity ^(a) Positive Response N=72	Specificity ^(b) Non-response N=160	$\kappa(0.5)$ ^(c)
15	0.96	0.67	0.53
20	0.90	0.76	0.59
25	0.87	0.84	0.68
30	0.76	0.89	0.65
35	0.67	0.91	0.60
40	0.49	0.96	0.50

^(a) Sensitivity is probability of exceeding YGTSS-TTS reduction cutoff among patients rated 1 or 2 on CGI-Improvement scale (Positive Response).

^(b) Specificity is probability of not exceeding YGTSS-TTS reduction cutoff among patients scored 3 or higher on CGI-Improvement (“Non Response”).

^(c) $\kappa(0.5)$ is weighted kappa statistic measuring accuracy of prediction.

Table 4
Odds Ratios for positive response in CBIT compared to PST across three classification methods

Intervention Group	CGI-Improvement Positive Response		Signal Detection Analysis ^(a)		Response Model Dimensional Response	
	Response N=72	Non-response N=160	Response N=88	Non-response N=144	Response N=83	Non-response N=149
CBIT	56 (47.9%)	61 (52.1%)	60 (51.3%)	57 (48.7%)	55 (47.0%)	62 (53.0%)
PST	16 (13.9%)	99 (86.1%)	28 (24.3%)	87 (75.6%)	28 (24.4%)	87 (75.6%)
Odds Ratio ^(b)	5.68 (95%CI=2.99, 10.78)		13.27 (95% CI=1.87, 5.72)		2.76 (95% CI=1.57, 4.82)	

^(a) Signal Detection Analysis defined positive response as 25% reduction on YGTSS-TTS at Week 10.

^(b) Odds Ratio reflects odds ratio for CBIT compared to PST without any adjustment (note: the odds ratios were not much different after adjusting for gender, age, medication, baseline YGTSS-TTS).