UCLA

UCLA Previously Published Works

Title

Benchmarking Treatment Response in Tourette's Disorder: A Psychometric Evaluation and Signal Detection Analysis of the Parent Tic Questionnaire.

Permalink

https://escholarship.org/uc/item/0d05g9fn

Journal

Behavior therapy, 49(1)

ISSN

0005-7894

Authors

Ricketts, Emily J
McGuire, Joseph F
Chang, Susanna
et al.

Publication Date

2018

DOI

10.1016/j.beth.2017.05.006

Peer reviewed

Benchmarking Treatment Response in Tourette's Disorder: A Psychometric Evaluation and Signal Detection Analysis of the Parent Tic Questionnaire

Emily J. Ricketts,
University of California, Los Angeles

Joseph F. McGuire,
University of California, Los Angeles

Susanna Chang,
University of California, Los Angeles

Deepika Bose,
University of California, Los Angeles

Madeline M. Rasch,
University of California, Los Angeles

Douglas W. Woods,
Marquette University

Matthew W. Specht,
Johns Hopkins University

John T. Walkup,
Weill Cornell Medical College

Lawrence Scahill,
Emory University School of Medicine

Sabine Wilhelm,
Massachusetts General Hospital/Harvard Medical School

Alan L. Peterson, and
University of Texas Health Science Center at San Antonio

John Piacentini
University of California, Los Angeles

Abstract

Address correspondence to Emily J. Ricketts, Ph.D., University of California, Los Angeles, Department of Psychiatry and Biobehavioral Sciences, 760 Westwood Plz., Los Angeles, CA; ericketts@mednet.ucla.edu.

This study assessed the psychometric properties of a parent-reported tic severity measure, the Parent Tic Questionnaire (PTQ), and used the scale to establish guidelines for delineating clinically significant tic treatment response. Participants were 126 children ages 9 to 17 who participated in a randomized controlled trial of Comprehensive Behavioral Intervention for Tics (CBIT). Tic severity was assessed using the Yale Global Tic Severity Scale (YGTSS), Hopkins Motor/Vocal Tic Scale (HMVTS) and PTQ; positive treatment response was defined by a score of 1 (*very much improved*) or 2 (*much improved*) on the Clinical Global Impressions – Improvement (CGI-I) scale. Cronbach's alpha and intraclass correlations (ICC) assessed internal consistency and test-retest reliability, with correlations evaluating validity. Receiver- and Quality-Receiver Operating Characteristic analyses assessed the efficiency of percent and raw-reduction cutoffs associated with positive treatment response. The PTQ demonstrated good internal consistency (α = 0.80 to 0.86), excellent test-retest reliability (ICC = .84 to .89), good convergent validity with the YGTSS and HM/VTS, and good discriminant validity from hyperactive, obsessive-compulsive, and externalizing (i.e., aggression and rule-breaking) symptoms. A 55% reduction and 10-point decrease in PTQ Total score were optimal for defining positive treatment response. Findings help standardize tic assessment and provide clinicians with greater clarity in determining clinically meaningful tic symptom change during treatment.

Keywords

Chronic tic disorders (CTDS), including Tourette's disorder (TD), are characterized by involuntary, repetitive movements (i.e., motor tics) and/or vocalizations (i.e., vocal tics) that have persisted for more than 1 year (American Psychiatric Association [APA], 2013). Tics generally first emerge in early childhood, peaking in severity in early adolescence, and, in many cases, steadily declining through early adulthood (Hallett, 2015). Among youth, CTDs are more common in males, with a ratio as high as 4:1 (Hallett, 2015; Robertson, 2012), and are prevalent at rates ranging from 0.4% to 3.8% (Knight et al., 2012; Scahill, Specht, & Page, 2013). In addition to tics, youth with CTDs commonly present with attention-deficit/ hyperactivity disorder (ADHD) and obsessive- compulsive disorder (OCD; Cavanna & Rickards, 2013). Although CTDs are often associated with diminished quality of life (Cavanna et al., 2013), behavioral and pharmacological interventions have both demonstrated efficacy in reducing tic severity (Murphy, Lewin, Storch, & Stock, 2013; Piacentini et al., 2010).

In clinical research, tics are most commonly assessed using the Yale Global Tic Severity Scale (YGTSS; Leckman et al., 1989). The YGTSS is a clinician-rated interview measure of tic severity that takes 30 to 45 minutes to complete. Available evidence suggests a reduction of 25% to 35% or decrease of 6 to 7 points on the YGTSS is associated with positive treatment response to empirically supported interventions among children and adults with CTDs (Jeon et al., 2013; Storch et al., 2011). Although informative, the YGTSS is less commonly used in clinical practice as it requires administration by a trained rater and may be time consuming (Chang, Himle, Tucker, Woods, & Piacentini, 2009). Indeed, clinician-rated measures are less favored in clinical practice due to the time burden (Boswell, Kraus,

Miller, & Lambert, 2015; Hatfield & Ogles, 2007) of administration, scoring, and interpretation (Garland, Kruse, & Aarons, 2003).

A time-efficient alternative to clinician ratings of tic severity are parent and self-report rating scales. These scales can be completed quickly in the waiting room prior to treatment visits. Although there are several options for parent and/or self-report rating scales (see McGuire et al. 2012 for a review), most have noted limitations that constrain their use in either research or clinical practice (e.g., minimal psychometric evaluation, lack of specificity to tic symptoms, absence of individual tic ratings, etc.). One promising parent-report measure of tic severity is the Parent Tic Questionnaire (PTQ; Chang et al., 2009). The PTQ assesses tic severity in the past week, allowing for individual parent ratings of tic presence or absence for 14 vocal tics and 14 motor tics. Additionally, the measure allows for separate ratings of tic frequency and intensity, completed for each tic. Frequency ratings range from 1 to 4 with the following anchors: weekly, daily, constantly, hourly. Intensity ratings range from 1 to 4, with higher scores indicative of greater tic intensity. The frequency and intensity ratings can be summed to yield a severity score ranging from 0 (i.e., tic is absent, thus no frequency or intensity ratings are given) to 8 for each tic. The PTQ includes subtotals for motor and vocal tic severity, which are summed to produce a total tic score. In the only prior psychometric evaluation (Chang et al., 2009), the PTQ exhibited fair to excellent internal consistency (α = .79 to .90), good to excellent 2-week test-retest reliability (Interclass correlation coefficient; ICC = .72 to .84), strong convergent validity with other measures of tic severity (r = .54 to .73), and discriminant validity, with correlations between the PTQ and YGTSS remaining strong after controlling for symptoms of inattention (rs = .53 to .70) and OCD (rs = .45 to .66).

Although providing initial evidence for using the PTQ, this initial report had some limitations, including a relatively small sample size (n = 40) and limited attention to discriminant validity of the PTQ. The latter point is particularly important due to the common presence of co-occurring psychiatric conditions (e.g., OCD, ADHD, disruptive behavior) among youth with tic disorders. It is important to ensure the PTQ distinguishes between tics and potentially co-occurring behaviors (e.g., compulsions, hyperactivity, aggression) that are distinct from tics but may appear similar in topography (Cath et al., 2011; Schapiro, 2002). Finally, the prior study did not provide guidelines for using the PTQ to determine a clinically meaningful response to treatment. The utilization of evidence-based assessment is important for a variety of reasons, including standardization of assessment, sharing treatment progress with patients, and providing a valid indication of clinically meaningful treatment response (Boswell et al., 2015; Lambert, 2013). These factors are especially important in the treatment of youth with CTDs, as symptoms are chronic and infrequently remit following treatment. Thus, establishing the PTQ as a valid and efficient measure of treatment response would address a crucial gap in the field.

This study conducted a comprehensive psychometric examination of the PTQ in a large sample and investigated the optimal percent and raw reduction in PTQ Total scores associated with positive treatment response. First, the internal consistency and test-retest reliability of the PTQ was examined. Second, the convergent and discriminant validity of the

PTQ was investigated. Finally, the optimal percent and raw reduction in PTQ Total Tic score was explored.

Method

PARTICIPANTS

Participants were 126 children and adolescents ages 9 to 17 (M = 11.73, SD = 2.32) with TD, Chronic Motor Tic Disorder or Chronic Vocal Tic Disorder, who participated in a NIH-funded randomized controlled-comparison of Comprehensive Behavioral Intervention for Tics (CBIT) and Psychoeducation and Supportive Psychotherapy (PST) between 2004 and 2007. Participants were enrolled at three sites, including University of California, Los Angeles (n = 45), Johns Hopkins University (n = 41), and University of Wisconsin-Milwaukee (n = 40), with support for data coding, therapist supervision, and data analysis and management from the University of Texas Health Science Center, Massachusetts General Hospital/Harvard Medical School, and Yale University, respectively. The gender distribution of the sample was 78.6% male, and the racial background was 84.9% Caucasian, 7.1% Hispanic, 3.2% African-American, 3.2% Asian/Pacific Islander, and 1.6% other (see Table 1). Piacentini et al. (2010) and Specht et al. (2011) provide additional information regarding sample characteristics and other aspects of the original trial methodology.

MEASURES

Parent Tic Questionnaire (PTQ)—As previously described, the PTQ (Chang et al., 2009) is a parent-rated tic severity scale assessing tic frequency and intensity for individual tics, which sum to Motor Tic, Vocal Tic, and Total scores. Motor Tic and Vocal Tic scores range from 0 to 112, and Total scores may range from 0 to 224.[1] Within the present sample, the actual range of baseline PTQ Total scores was 5 to 107 (M = 36.11; SD = 20.55), with PTQ Total scores at week 10 ranging from 2 to 92 (M = 23.09; SD = 17.29). The PTQ has demonstrated initial reliability and validity in a small sample, as described in the introduction (Chang et al., 2009).

Yale Global Tic Severity Scale (YGTSS)—The YGTSS (Leckman et al., 1989) is a semistructured clinician-rated instrument assessing motor and vocal tic severity in the past week. Motor and vocal tics are rated separately (ranging from 0 to 5) across five domains: number, frequency, intensity, complexity, and interference. The YGTSS produces a Motor Tic score and Vocal Tic score each ranging from 0 to 25. The Motor and Vocal tic scores are summed to yield a Total Tic score ranging from 0 to 50. The YGTSS has demonstrated excellent interrater reliability (Walkup, Rosenberg, Brown, & Singer, 1992), fair to excellent stability (Storch et al., 2005), and good convergent and discriminant validity (Leckman et al., 1989; Storch et al., 2005).

---

[1]The PTQ lists 14 motor tics and 14 vocal tics with one "other" item listed under the motor tic scale and an "other" and "other vocal" tic items listed under the vocal tic scale. However, in the original trial (Piacentini et al., 2010) the "other" and "other vocal" tic items were collapsed for data analyses to ensure equal weighting of this item between motor and vocal scales (Abramovitch et al., 2015). This yielded 13 remaining items for the vocal tic scale.

Hopkins Motor/Vocal Tic Scale (HM/VTS)—The HMVTS (Walkup, Rosenberg, Brown, & Singer, 1992) is a measure of severity of motor and vocal tics over the past week. A modified version was used, wherein participants listed up to 5 motor tics and 5 vocal tics considered most bothersome. A clinician then rated each tic on a 0 (*none*) to 4 (*severe*) scale, informally factoring in the patient and/or parent's verbal report of tic frequency, intensity, interference, and emotional distress. Tic ratings were summed to create composites for motor tic, vocal tic, and total tic severity. The HM/VTS has good interrater reliability, good concurrent validity, good divergence from ADHD, fair divergence from OCD (Walkup et al., 1992) and treatment sensitivity (McGuire et al., 2015).

ADHD Rating Scale-IV (ADHD RS-IV)—The ADHD RS-IV (DuPaul et al., 1998) is an 18-item measure with parent and teacher versions, used to assess ADHD symptom severity in the past week. Each item is categorized as either Hyperactive/Impulsive or Inattentive and is rated from 0 (*no symptoms*) to 3 (*severe symptoms*) based on clinician interview with the parent and child. The sum of these items ranges from 0 (*no symptoms*) to 54 (*the most severe symptoms*), reflecting the patient's overall ADHD symptom severity. The ADHD RS-IV possesses satisfactory interrater reliability, good internal consistency, excellent test-retest reliability, acceptable convergent and discriminant validity (Zhang et al., 2005), and strong predictive validity of specific ADHD diagnostic status (Power et al., 1998).

Children's Yale-Brown Obsessive Compulsive Scale (CY-BOCS)—The CY-BOCS (Scahill et al., 1997) is a semistructured, clinician-administered scale measuring the presence and severity of obsessions and/or compulsions over the past week. The CYBOCS consists of a 5-item obsession rating scale and a 5-item compulsion rating scale. The sum of all 10 items provides a CY-BOCS Total score, with higher scores indicative of greater symptom severity. The CY-BOCS demonstrates good to excellent interrater reliability (Scahill et al., 1997; Storch et al., 2004), high internal consistency (Storch et al., 2004), strong convergent validity, and adequate to good discriminant validity (Scahill et al., 1997; Storch et al., 2004).

Child Behavior Checklist/6-18 (CBCL/6-18)—The CBCL/6-18 (Achenbach & Rescorla, 2001) is a 118-item parent/caregiver-report questionnaire that assesses a wide variety of emotional and behavioral problems experienced by youths. Behavior is rated on a 3-point Likert scale, with item responses corresponding to "Not True" (0), "Somewhat or Sometimes True" (1), and "Very True or Often True" (2). The CBCL/6-18 includes three overarching scales, including Internalizing Problems, Externalizing Problems, and Total Problems. The measure also includes the following syndrome subscales: Anxious/depressed, Withdrawn/depressed, Somatic Complaints, Rule Breaking, Aggressive Behavior, Social Problems, Thought Problems, and Attention Problems. Of interest in the present analysis was the Externalizing Problems subscale (comprised of Rule Breaking and Aggressive Behavior subscales). Raw scores are converted to age- and gender-normed *T* scores, with scores of 60 or greater on the overarching scales and 70 or greater on the syndrome subscales indicative of clinically significant symptoms. The CBCL/6-18 has evidenced strong psychometric properties, including high test-retest reliability (ICC = .95 for specific

problem items), high internal consistency (α = .78 to.97), good convergent and discriminant validity, and a factor structure supporting its subscales (Achenbach & Rescorla, 2001).

Clinical Global Impression – Improvement Scale (CGI-I)—The CGI-I (Guy, 1976) is a clinician-rated instrument designed to assess global improvement in functioning following illness. A version of this scale, modified to assess global tic-related impairment and commonly used as a primary outcome measure in trials involving patients with TD (e.g., Piacentini et al., 2010; Scahill et al., 2001; Wilhelm et al., 2012), was used. Clinicians rate the perceived patient global improvement in tic-related impairment according to the following 7-point scale: Very Much Improved (1), Much Improved (2), Improved (3), Minimally Improved (4), No change (5), Minimally worse (6), and Very Much Worse (7). A score of 1 or 2 was used to classify positive treatment response in the original trial (Piacentini et al., 2010).

TREATMENTS

CBIT and PST were administered by trained clinicians with a master's degree or higher during the acute treatment period. CBIT is a multicomponent behavioral treatment protocol designed to reduce tic severity (Woods et al., 2008). The primary component of CBIT is Habit Reversal Training (HRT), which consists of several techniques including, most prominently, awareness training, competing response training, and social support. The goal of HRT is to enhance awareness of premonitory urges and tic occurrence, train the use of a behavior that is physically incompatible with tic occurrence, and encourage use of these techniques with parental praise and prompting (Woods et al., 2008). A second core component of CBIT is function-based assessment and intervention, with the goal of identifying settings, events, affective states, and social reactions exacerbating symptoms, and reducing the impact of these stimuli on tic symptoms. CBIT also includes relaxation techniques (diaphragmatic breathing, progressive muscle relaxation), a behavioral reward system, and relapse prevention strategies. PST included psychoeducation and discussion of issues relevant to tics, with no direct tic intervention. Both interventions included two initial 90-minute sessions followed by six 60-minute sessions, with the first six occurring weekly and the final two each occurring in 2-week intervals (Piacentini et al., 2010).

INDEPENDENT EVALUATOR TRAINING

Assessments were administered by master's-level or higher independent evaluators (IE), who were trained to criterion, certified, and monitored according to procedures outlined in the trial (Piacentini et al., 2010). Training on clinician-rated assessment was directed by experts, LS and JW, and involved providing co-ratings of three videotaped YGTSS and CYBOCS assessments and CGI-I ratings of three written case vignettes. IEs were required to score within 15% of the expert's rating on the YGTSS and CYBOCS and within 15% of the group mean for all IEs undergoing training. IEs were also required to score within 1 point of the expert's rating on the CGI-I, with 100% cross-IE agreement required on no fewer than two of the four ratings. Cross-site reliability was maintained through IE supervision at each study site and during bi-weekly cross-site conference calls. Additionally, 13% of assessment video recordings were randomly selected for co-rating by the quality assurance site, yielding good reliability and no cross-site variability.

PROCEDURE

Following completion of IRB-approved consent and assent procedures, IEs blinded to treatment condition screened youth for study eligibility. Participants were included in the trial based on a DSM-IV-TR (APA, 2000) diagnosis of TD or CTD (i.e., chronic motor tic disorder or chronic vocal tic disorder), established through administration of the Anxiety Disorders Interview Schedule – Research Lifetime Version (Silverman & Albano, 2002), a modified version of the Anxiety Disorders Interview Schedule (Silverman & Albano, 1996), which included added modules on chronic tic disorders, and several other psychiatric disorders; moderate tic severity as evidenced by a YGTSS score ≥ 14 for TD and ≥ 10 for CTD; fluency in English; and intellectual functioning in the low average range or higher (IQ score N80), determined through administration of the Wechsler Abbreviated Scale of Intelligence.

Psychotropic medications were allowed provided participants had been on a stable dosage at least 6 weeks prior to study entry and no changes in dosage were planned during the course of study participation. Individuals were excluded based on a lifetime diagnosis of psychosis, mania, or pervasive developmental disorder, current diagnosis of substance abuse or dependence, any medical condition interfering in study participation, and 4 or more prior sessions of behavior therapy for tics. Eligible participants received a baseline assessment 7 to 10 days following screening, and were randomized to receive either CBIT or PST. Participants received mid- (5 week) and post-treatment (10 week) assessments during a 10-week acute treatment period (Piacentini et al., 2010). IE-administered interview measures (YGTSS, CY-BOCS, HM/VTS) were completed with the child and parent concurrently. The CGI-I was completed by an IE based on clinical judgment of improvement in tic-related impairment. The parent completed the PTQ, CBCL, and ADHD RS-IV. With respect to the timing of administration of instruments used in the present analysis, the YGTSS, ADHD RS-IV, CBCL, and PTQ were administered at screening; the YGTSS, CY-BOCS, HM/VTS, and PTQ were administered at baseline and week 5 assessments, with the CGI-I and ADHD RS-IV also completed at week 5. All measures of interest were re-administered at week 10 (posttreatment).

ANALYTIC PLAN

In the present study, Cronbach's alpha and ICC calculated internal consistency and test-retest reliability of the PTQ Total tic score, respectively. For internal consistency, α values ≥ .90 were considered excellent, .80 to .89 were considered good, .70 to .79 were considered fair, and b.70 poor (Cicchetti, 1994). Meanwhile, ICC of .75 to 1.00 was indicative of excellent test-retest reliability, values of .60 to .74 signified good reliability, coefficients ranging from .40 to .59 indicated poor agreement, and <.40 was considered poor agreement (Cicchetti, 1994). Second, Pearson correlations examined the convergent and discriminant validity of the PTQ Total tic score. A correlation value of >.50 between the PTQ and other measures of tic severity indicated good convergent validity. Correlations of .30 to .49 and .10 to .29 represented fair and poor convergent validity, respectively. Good discriminant validity was represented by correlations of .10 to .29 between the PTQ and measures that did not assess tic severity. Correlation values that exceeded this range were considered fair (.30 to .49) and poor (>.50) discriminant validity (Cicchetti, 1994). Finally, a receiver operating

characteristic (ROC) assessed a range of percentage and raw reduction PTQ Total tic score cutoffs in relation to treatment responder status using the CGI-I scale. The PTQ Total tic scores were divided into raw reduction cutoff scores in 2-point increments and raw percent reduction cutoffs (set by 5-point increments). We performed ROC curves, plotting sensitivity, or true positive rate, referring to the proportion of treatment responders obtaining scores above various percent and raw reduction cutoffs on the y-axis, and false positive rate (1-specificity), referring to the proportion of treatment responders who failed to obtain scores above cutoffs on the x-axis (Swets, 1996). Sensitivity and false positive rates for each cutoff were used in conjunction with formulas by Kraemer and colleagues (Kraemer, Periyakoil, & Noda, 2002) to establish specificity (the rate of nontreatment responders who did not score above various cutoffs), positive predictive value (proportion of participants with raw or percent reductions above various PTQ cutoffs who were classified as treatment responders), negative predictive value (rate of participants not exceeding PTQ cutoffs who were identified as nontreatment responders), and efficiency (the concordance rate between cutoffs and treatment responder status; Glaros & Kline, 1988; Lalken & McClusky, 2008). Youden Index J (sensitivity – specificity – 1), a common ROC curve summary statistic, was also calculated, as it provides an optimal cutoff point for establishing a given test's ability to discriminate between diagnostic groups when sensitivity and specificity are equally considered (Youden, 1950).

Although ROC analysis is highly useful for diagnostic decision making, it has several limitations. First, sensitivity and specificity assess the proportion of patients correctly categorized but do not assess the ability of a test to differentiate between diagnostic groups; and second, due to their properties, sensitivity and specificity values lack an interpretive statistical scale of reference (Gilchrist, 1992). To address these limitations and provide additional ROC interpretive measures, Quality Receiver Operating Characteristic (QROC) analysis was performed. QROC analysis rescales sensitivity and specificity values to weighted kappa coefficients or quality values, which provide a standardized measure of ROC values (Kraemer et al., 2002; Moore, Andlauer, Simon, & Mignot, 2014). Specific kappa coefficients calculated in the present analysis include rescaled measures of sensitivity (k_1), specificity (k_0), and efficiency ($k_{0.5}$; Gilchrist, 1992).

Results

INTERNAL CONSISTENCY

The internal consistency for the PTQ Motor tic score ($\alpha = .82$; $\alpha = .81$), PTQ Vocal tic score ($\alpha = .80$, $\alpha = .83$), and PTQ Total tic score ($\alpha = .86$; $\alpha = .86$) were good at the screening and baseline visits, respectively.

TEST-RETEST RELIABILITY

Test-retest reliability between screening and baseline administrations of the PTQ Motor tic score (ICC = .84; 95% CI = .76 to .89), PTQ Vocal tic score (ICC = .85; 95% CI = .77 to .90), and PTQ Total tic score (ICC = .89; 95% CI = .84 to .92) were excellent.

CONVERGENT VALIDITY

Across screening and baseline assessments correlations between the PTQ and YGTSS for PTQ Motor tic and YGTSS Motor tic scores ($r = .62$, $p < .001$; $r = .66$, $p < .001$), PTQ Vocal tic and YGTSS Vocal tic scores ($r = .53$, $p < .001$; $r = .58$, $p < .001$), and PTQ Total tic and YGTSS Total tic scores ($r = .68$, $p < .001$; $r = .64$, $p < .001$) were indicative of good convergent validity. Convergence between baseline PTQ and HM/VTS ratings was good for the PTQ Motor tic score and HM/VTS Motor tic severity composite ($r = .50$, $p < .001$); PTQ Vocal tic score and HM/VTS Vocal tic severity composite ($r = .64$, $p < .001$); and PTQ Total tic score and HM/VTS Total tic severity composite ($r = .61$, $p < .001$).

DISCRIMINANT VALIDITY

At the screening assessment the PTQ Total tic score did not significantly correlate with the ADHD RS-IV Hyperactivity score ($r = .14$, $p = .15$) or ADHD RS-IV Total score ($r = .14$ $p = .14$). Correlations were also not significant for the ADHD RS-IV scores and PTQ Motor tic ($r = .11$-0.12, $ps = .19$-$.24$) and Vocal tic scores ($r = .06$-$.08$, $ps = .41$-$.57$), indicating good discriminance between scales. Additionally, there was good discriminance between baseline PTQ Total tic scores, CY-BOCS Total scores ($r = .16$, $p = .09$), and CY-BOCS Compulsion scores ($r = .10$, $p = .30$). At the screening assessment, the PTQ Motor Tic score ($r = .09$, $p = .38$), PTQ Vocal Tic score ($r = .12$, $p = .24$) and PTQ Total Tic score ($r = .12$, $p = .19$) exhibited good discriminance from the CBCL/6-18 Externalizing T score.

INTER-SCALE CORRELATIONS

Inter-scale correlations between the screening administration of the PTQ Total Tic score and PTQ Motor Tic ($r = .90$, $p < .001$) and Vocal Tic ($r = .79$ $p < .001$) scores were strong. The correlation between PTQ Motor Tic and Vocal Tic scores at screening was fair ($r = .47$, $p < .001$). At baseline, inter-scale correlations between the PTQ Total Tic score and PTQ Motor Tic ($r = .86$, $p < .001$) and Vocal Tic ($r = .81$, $p < .001$) scores were also strong. The correlation between baseline PTQ Motor and Vocal Tic scores was ($r = .45$, $p < .001$) fair.

PREDICTING POSITIVE TREATMENT RESPONSE USING PTQ TOTAL PERCENTAGE REDUCTION

Forty-four participants (34.9%) within the sample were classified as treatment responders on the CGI-I at the 10-week posttreatment assessment. A 55% reduction in PTQ Total score yielded the highest rescaled efficiency score ($k_{0.5} = .52$), indicative of a 52% likelihood of concordance between the 55% reduction cutoff and CGI-I rating. The cutoff of 52% yielded the highest Youden Index J score (.50), indicating this cutoff provides the most optimal balance between sensitivity (.61) and specificity (.89). The positive predictive and negative predictive values at this cutoff were both acceptable at .78 each. However, a decrease in PTQ Total of 45% yielded similar rescaled efficiency ($k_{0.5} = .48$) and Youden Index J (.49) scores. See Table 2 and Figure 1 for QROC analyses of PTQ percent reduction cutoffs predicting treatment responder status.

PREDICTING POSITIVE TREATMENT RESPONSE USING PTQ TOTAL RAW SCORE REDUCTION

A raw PTQ Total reduction of 10 points yielded optimal rescaled efficiency ($k_{0.5}$ = .41), indicating a 41% chance of agreement between the 10-point reduction cutoff and CGI-I rating. This cutoff yielded a Youden Index *J* score of .43, representing maximum equivalence between sensitivity (.78) and specificity (.65). A positive predictive value of .58 and negative predictive value of .82 were found at this cutoff. See Table 3 and Figure 2 for QROC analyses of PTQ raw reduction cutoffs defining positive treatment response.

Discussion

Given the time burden and limited utilization of clinician-rated measures to monitor treatment response in clinical practice, this study examined the psychometric properties of the PTQ and conducted a signal detection analysis to assess the efficiency of tic severity reduction cutoffs associated with positive treatment response. The PTQ was found to have good internal consistency and excellent test-retest reliability. Additionally, the PTQ exhibited strong convergent validity with clinician-rated measures of tic severity, and strong discriminant validity from constructs that frequently co-occur with tics but are distinct (e.g., ADHD symptom severity, OCD symptom severity, severity of externalizing problems, including defiant behavior and aggression). Additionally, the PTQ showed strong inter-scale correlations between the total score and motor and vocal subtotals. Understandably, correlations between motor and vocal subscales of the PTQ were weaker, as the two independent scales are combined to yield a total score. These findings are consistent with the initial psychometric investigation of the PTQ, which showed high internal consistency, good to excellent test-retest reliability, excellent convergent validity with the YGTSS, and preliminary evidence of discriminant validity from symptoms of inattention and OCD (Chang et al., 2009). However, this report extends this initial psychometric investigation by utilizing a larger treatment-seeking clinical sample, and investigating discriminant validity using a broad array of comorbid constructs (e.g., internalizing and externalizing symptoms). Findings were also consistent with the initial psychometric investigation of the adult version of this scale, the Adult Tic Questionnaire, which shows strong psychometric properties (Abramovitch et al., 2015).

In addition, this report examined the efficiency of tic severity reduction cutoffs associated with positive treatment response. The signal detection analyses identified a 55% reduction in the PTQ Total score as optimal for defining positive treatment response, with a range from 45% to 55% being most representative of positive treatment response. When examining raw-score reductions, a 10-point reduction in the PTQ total tic score was maximally indicative of positive clinical response. Notably, these differences are larger than prior studies, which have found a 25% to 35% (or 6-to-7 point) reduction in total tic severity on the YGTSS to correspond with a positive treatment response (Jeon et al., 2013; Storch et al., 2011). The difference in percent reduction between scales may be related to the high range of PTQ scores relative to the YGTSS. Alternatively, differences may be related to discrepancies between parent and clinician perspectives. Furthermore, differences may be attributed to distinctions in the structure of YGTSS and PTQ scales. Specifically, the YGTSS assesses tic

severity across five domains (i.e., number, frequency, intensity, complexity, interference), whereas the PTQ takes into account tic number, frequency, and intensity. While the YGTSS is still considered to be the gold-standard measure to evaluate tic severity in clinical research, it requires considerable training and time to administer. Comparatively, the PTQ offers advantages over the YGTSS in clinical practice due to its strong psychometric properties and ease of administration. Moreover, the present findings provide optimal benchmarks to help clinicians establish positive response to treatment for tics, thereby increasing the utility of the scale.

Findings should be considered within the context of study limitations. First, the PTQ is a heterogeneous measure with ratings varying considerably across patients depending on tic classification (i.e., motor and/or vocal) and number endorsed within the past week. Therefore, ratings are more meaningful when used to track symptoms over time within cases rather than used as a comparative benchmark of overall tic severity across youth with CTDs. Second, generalizability of our findings to the broader population of youth with tics may be limited by the demographics of our sample (i.e., predominantly Caucasian), and the context (i.e., research setting) within which the study was conducted. However, the demographics, clinical characteristics, and settings are largely consistent with other treatment studies of youth with CTDs. Additionally, the brevity of the 7-to-10 day test-retest reliability window —used in the context of a treatment trial—may have influenced participant ratings. Moreover, as tic listings differed between participants, many tics were rated as absent in the past week, resulting in frequency and intensity ratings of 0. Therefore, internal consistency outcomes should be interpreted cautiously. Finally, due to the heterogeneity in tic symptom presentation across participants, multiple factors may have influenced IE ratings of treatment response. For example, in some cases those rated as positive treatment responders may have had significant reductions in one or two tics, while in other cases individuals may have experienced general reductions across all tics.

Within the past decade, researchers have worked to disseminate empirically supported behavioral treatment for tics into standard clinical practice. As utilization of behavior therapy for tics in community practice expands, so will the need for reliable, standardized measures to assess tic symptom change over time. Although the YGTSS continues to serve as the gold-standard measure to assess tic severity, there are several pragmatic limitations that constrain its use across clinical settings. Comparatively, our findings establish the PTQ as an efficient and psychometrically sound instrument for use with parents of children with TD symptoms of a moderate or worse nature (i.e., those with mild symptoms were not included in the clinical trial). Moreover, we have outlined benchmarks for assessing clinically meaningful symptom change in youth undergoing treatment for CTDs using the PTQ. This is particularly relevant for treatment of tics as remission is rare; thus, guidelines will allow clinicians to empirically assess therapeutic response even when tic symptoms persist. Thus, the combination of strong psychometric properties, ease of administration, and clinical benchmarks for treatment improvement facilitate the utility of the PTQ for monitoring treatment response in clinical practice.

Acknowledgments

References

Abramovitch A, Reese H, Woods DW, Peterson A, Deckersbach T, Piacentini J, Wilhelm S. Psychometric properties of a self-report instrument for the assessment of tic severity in adults with tic disorders. Behavior Therapy. 2015; 46:786–796. [PubMed: 26520221]

Achenbach, TM., Rescorla, L. The manual for the ASEBA school-age forms & profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families; 2001.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th. Washington, DC: Author; 2000. text revision

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th. Washington, DC: Author; 2013.

Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. Psychotherapy Research. 2015; 25:6–19. [PubMed: 23885809]

Cath DC, Hedderly T, Ludolph AG, Stern JS, Murphy T, Hartmann A, the ESSTS Guidelines Group. European clinical guidelines for Tourette Syndrome and other tic disorders. European Child & Adolescent Psychiatry. 2011; 20:155–171. [PubMed: 21445723]

Cavanna AE, David K, Bandera V, Termine C, Balottin U, Schrag A, Selai C. Health-related quality of life in Gilles de la Tourette Syndrome: A decade of research. Behavioural Neurology. 2013; 27:83–93. [PubMed: 23187148]

Cavanna AE, Rickards H. The psychopathological spectrum of Gilles de la Tourette Syndrome. Neuroscience & Biobehavioral Reviews. 2013; 37:1008–1015. [PubMed: 23131314]

Chang S, Himle MB, Tucker BTP, Woods DW, Piacentini J. Initial psychometric properties of a brief parent-reported instrument for assessing tic severity in children with chronic tic disorders. Child & Family Behavior Therapy. 2009; 31:181–191.

Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment. 1994; 6:284–290.

DuPaul GJ, Anastopoulos AD, Power TJ, Reid R, Ikeda MJ, McGoey KE. Parent ratings of Attention-Deficit/Hyperactivity Disorder Symptoms: Factor structure and normative data. Journal of Psychopathology and Behavioral Assessment. 1998; 20:83–102.

Garland AF, Kruse M, Aarons GA. Clinicians and outcome measurement: What's the use? Journal of Behavioral Health Services and Research. 2003; 30:393–405. [PubMed: 14593663]

Gilchrist J. QROC curves and kappa functions: New methods for evaluating the quality of clinical decisions. Ophthalmic and Physiological Optics. 1992; 12:350–360. [PubMed: 1454373]

Glaros AG, Kline RB. Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive model. Journal of Clinical Psychology. 1988; 44:1013–1023. [PubMed: 3216006]

Guy W. Clinical global impression scale. The ECDEU Assessment Manual for Psychopharmacology-Revised Volume DHEW Publ No ADM. 1976; 76(338):218–222.

Hallett M. Tourette Syndrome: Update. Brain and Development. 2015; 37:651–655. [PubMed: 25604739]

Hatfield DR, Ogles BM. Why some clinicians use outcome measures and others do not. Administration and Policy in Mental Health and Mental Health Services Research. 2007; 35:283–291.

Jeon S, Walkup JT, Woods DW, Peterson A, Piacentini J, Wilhelm S, Scahill L. Detecting a clinically meaningful change in tic severity in Tourette Syndrome: A comparison of three methods. Contemporary Clinical Trials. 2013; 36:414–420. [PubMed: 24001701]

Knight T, Steeves T, Day L, Lowerison M, Jette N, Pringsheim T. Prevalence of tic disorders: A systematic review and meta-analysis. Pediatric Neurology. 2012; 47:77–90. [PubMed: 22759682]

Kraemer HC, Periyakoil VS, Noda AK. Kappa coefficients in medical research. Statistics in Medicine. 2002; 21:2109–2129. [PubMed: 12111890]

Lalken AG, McClusky A. Clinical tests: Sensitivity and specificity. Continuing Education in Anaesthesia, Critical Care and Pain. 2008; 8:221–223.

Lambert MJ. Outcome in psychotherapy: The past and important advances. Psychotherapy. 2013; 50:42–51. [PubMed: 23505980]

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. Journal of the American Academy of Child and Adolescent Psychiatry. 1989; 566:566–573.

McGuire JF, Kugler BB, Park JM, Horng B, Lewin AB, Murphy TK, Storch EA. Evidence-based assessment of compulsive skin picking, chronic tic disorders and trichotillomania in children. Child Psychiatry and Human Development. 2012; 43:855–883. [PubMed: 22488574]

McGuire JF, Piacentini J, Scahill L, Woods DW, Villarreal R, Wilhelm S, Walkup JT, Peterson AL. Bothersome tics in patients with chronic tic disorders: Characteristics and individualized treatment response to behavior therapy. Behaviour Research and Therapy. 2015; 70:56–63. [PubMed: 25988365]

Moore HE Jr, Andlauer O, Simon N, Mignot E. Exploring medical diagnostic performance using interactive, multi-parameter sourced receiver operating characteristic scatter plots. Computers in Biology and Medicine. 2014; 47:120–129. [PubMed: 24561350]

Murphy TK, Lewin AB, Storch EA, Stock S. Practice parameter for the assessment and treatment of children and adolescents with tic disorders. Journal of the American Academy of Child and Adolescent Psychiatry. 2013; 52:1341–1359. [PubMed: 24290467]

Piacentini J, Woods DW, Scahill L, Wilhelm S, Peterson AL, Chang S, Walkup JT. Behavior therapy for children with Tourette Disorder: A randomized controlled trial. Journal of the American Medical Association. 2010; 303:1929–1937. [PubMed: 20483969]

Power TJ, Doherty BJ, Panichelli-Mindel SM, Karustis JL, Eiraldi RB, Anastopoulos AD, DuPaul GJ. The predictive validity of parent and teacher reports of ADHD symptoms. Journal of Psychopathology and Behavioral Assessment. 1998; 20:57–81.

Robertson MM. The Gilles de la Tourette Syndrome: The current status. Archives of Disease in Childhood Education and Practice Edition. 2012; 97:166–175. [PubMed: 22440810]

Scahill L, Chappell PB, Kim YS, Schultz RT, Katsovich L, Shepherd E, Leckman JF. A placebo-controlled study of guanfacine in the treatment of children with tic disorders and attention deficit hyperactivity disorder. American Journal of Psychiatry. 2001; 158:1067–1074. [PubMed: 11431228]

Scahill L, Riddle MA, McSwiggin-Hardin M, Ort SI, King RA, Goodman WK, Leckman JF. Children's Yale-Brown obsessive compulsive scale: reliability and validity. Journal of the American Academy of Child and Adolescent Psychiatry. 1997; 36:844–852. [PubMed: 9183141]

Scahill L, Specht M, Page C. The prevalence of tic disorders and clinical characteristics in children. Journal of Obsessive-Compulsive and Related Disorders. 2013; 3:394–400.

Schapiro NA. "Dude, you don't have Tourette's:" Tourette's syndrome, beyond the tics. Pediatric Nursing. 2002; 28:243–246. [PubMed: 12087644]

Silverman, WK., Albano, AM. The anxiety disorders interview schedule for children for DSM-IV [Child and Parent Versions]. San Antonio: Psychological Corporation; 1996.

Silverman, WK., Albano, AM. The anxiety disorders interview schedule for DSM-IV, research and lifetime version for children and parents (ADIS-RLV). New York: Columbia University; 2002. Unpublished manuscript

Specht MW, Woods DW, Piacentini J, Scahill L, Wilhelm S, Peterson AL, Walkup JT. Clinical characteristics of children and adolescents with a primary tic disorder. Journal of Developmental and Physical Disabilities. 2011; 23:15–31. [PubMed: 24999300]

Storch EA, De Nadai AS, Lewin AB, McGuire JF, Jones AM, Mutch PJ, Murphy TK. Defining treatment response in pediatric tic disorders: A signal detection analysis of the Yale Global Tic Severity Scale. Journal of Child and Adolescent Psychopharmacology. 2011; 21:621–627. [PubMed: 22070181]

Storch EA, Murphy TK, Geffken GR, Soto O, Sajid M, Allen P, Goodman WK. Psychometric evaluation of the Children's Yale–Brown Obsessive-Compulsive Scale. Psychiatry Research. 2004; 129:91–98. [PubMed: 15572188]

Storch EA, Murphy TK, Geffken GR, Sajid M, Allen P, Roberti JW, Goodman WK. Reliability and validity of the Yale Global Tic Severity Scale. Psychological Assessment. 2005; 17:486. [PubMed: 16393016]

Swets, JA. Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers. Mahwah, NJ: Lawrence Erlbaum Associates; 1996.

Walkup JT, Rosenberg LA, Brown J, Singer HS. The validity of instruments measuring tic severity in Tourette's syndrome. Journal of the American Academy of Child and Adolescent Psychiatry. 1992; 31:472–477. [PubMed: 1592779]

Wilhelm S, Peterson AL, Piacentini J, Woods DW, Deckersbach T, Sukhodolsky DG, Scahill L. Randomized trial of behavior therapy for adults with Tourette syndrome. Archives of General Psychiatry. 2012; 69:795–803. [PubMed: 22868933]

Woods, DW., Piacentini, J., Chang, SW., Deckersbach, T., Ginsburg, G., Peterson, AL., Wilhelm, S. Managing Tourette Syndrome: A Behavioral Intervention for Children and Adults. New York, NY: Oxford University Press; 2008.

Youden WJ. Index for rating diagnostic tests. Cancer. 1950; 3:32–35. [PubMed: 15405679]

Zhang S, Faries DE, Vowles M, Michelson D. ADHD Rating Scale IV: Psychometric properties from a multinational study as a clinician-administered instrument. International Journal of Methods in Psychiatric Research. 2005; 14:186–201. [PubMed: 16395872]
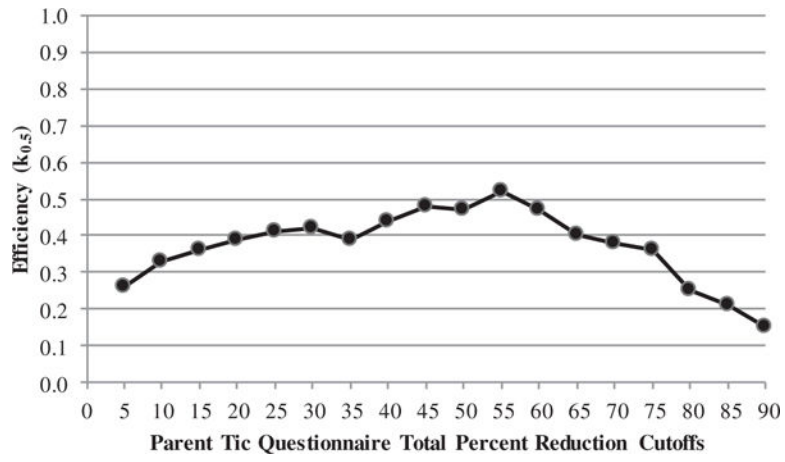
FIGURE 1.
Efficiency ($k_{0.5}$) for Parent Tic Questionnaire Total tic score percent reduction cutoffs predicting positive treatment response.
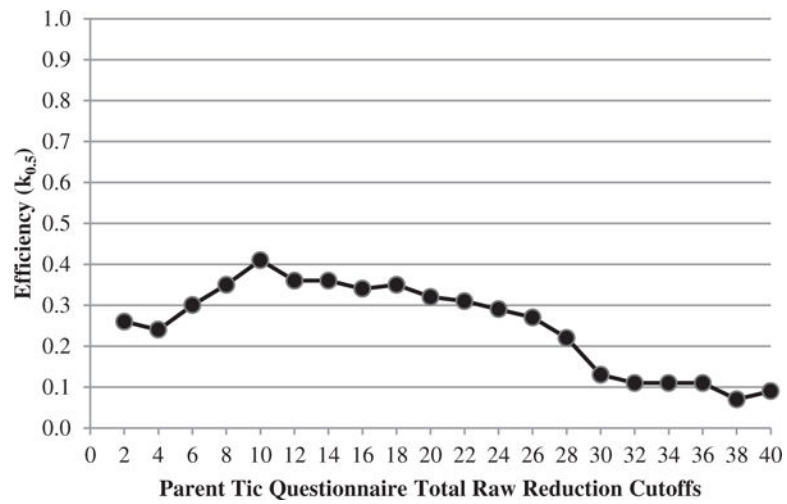
FIGURE 2.

Efficiency ($k_{0.5}$) for Parent Tic Questionnaire Total tic score raw reduction cutoffs predicting positive treatment response.

Table 1

Demographic and Clinical Characteristics of the Full Sample

	N = 126
Demographics	
Age M(SD)	11.7 (2.3)
Male N(%)	99 (78.5)
Racial/Ethnic Minority N(%)	19 (15.1)
On Tic Meds N(%)	46 (36.6)
Tic Disorder Diagnosis N(%)	
Tourette Syndrome	118 (93.7)
Chronic Motor Tic Disorder	7 (5.6)
Chronic Vocal Tic Disorder	1 (0.8)
Lifetime Comorbid Diagnosis N(%)	
Attention Deficit Hyperactivity Disorder	15 (11.9)
Obsessive-Compulsive Disorder	24 (19.0)
Generalized Anxiety Disorder	25 (19.8)
Social Phobia	27 (21.4)
Separation Anxiety Disorder	11 (8.7)
Baseline Tic Severity M(SD)	
PTQ Total	36.11 (20.54)
PTQ Motor	32.78 (12.86)
PTQ Vocal	14.38 (10.99)
YGTSS Total	24.66 (6.06)
YGTSS Motor	14.64 (3.78)
YGTSS Vocal	10.02 (4.55)
Screening CBCL/6-18 M(SD)	
Externalizing Problems	48.96 (10.38)
Baseline CY-BOCS M(SD)	
Total	6.51 (7.79)
Compulsions	4.22 (4.91)
Screening ADHD RS-IV M(SD)	
Total	14.86 (12.24)
Hyperactivity Subscale	5.95 (5.92)

Note. CBCL/6-18 = Child Behavior Checklist for Ages 6-18; CY-BOCS = CBCL 6-18 = Children's Yale-Brown Obsessive-Compulsive Scale; ADHD RS-IV = ADHD Rating Scale-IV; The CBCL 6/18 and ADHD RS-IV were not administered at baseline.

Table 2

Predicting Positive Treatment Response From Parent Tic Questionnaire Total Score Percent Reduction Cutoffs

PTQ Total Tic Score Percent Reduction Cutoff	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Efficiency	k_1	k_0	$k_{0.5}$	Youden Index J
5	.95	.35	.48	.92	.58	.79	.16	.26	.3
10	.95	.43	.51	.93	.63	.83	.21	.33	.38
15	.90	.49	.53	.89	.65	.72	.24	.36	.39
20	.90	.52	.54	.89	.67	.74	.26	.39	.43
25	.85	.59	.57	.86	.69	.66	.30	.41	.44
30	.81	.64	.58	.84	.70	.59	.32	.42	.44
35	.76	.65	.58	.81	.69	.51	.32	.39	.41
40	.76	.70	.61	.82	.72	.54	.37	.44	.45
45	.71	.78	.67	.81	.75	.51	.46	.48	.49
50	.63	.83	.69	.78	.75	.44	.51	.47	.46
55	**.61**	**.89**	**.78**	**.78**	**.78**	**.45**	**.64**	**.52**	**.50**
60	.54	.91	.78	.76	.76	.37	.64	.47	.44
65	.46	.91	.75	.73	.73	.30	.60	.40	.37
70	.44	.91	.74	.72	.72	.28	.59	.38	.34
75	.39	.94	.80	.71	.73	.25	.67	.36	.33
80	.24	.97	.83	.67	.69	.15	.73	.25	.21
85	.20	.98	.88	.66	.68	.12	.81	.21	.18
90	.12	1.00	1.00	.64	.66	.08	.99	.15	.12
95	.00	1.00							

Table 3

Predicting Positive Treatment Response From Parent Tic Questionnaire Total Tic Score Raw Reduction Cutoffs

PTQ Total Tic Score Raw Reduction Cutoff	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Efficiency	k_1	k_0	$k_{0.5}$	Youden Index J
2	.95	.35	.48	.92	.58	.12	.23	.26	.3
4	.85	.41	.48	.82	.58	.11	.24	.24	.27
6	.81	.52	.52	.81	.63	.19	.32	.30	.33
8	.81	.57	.54	.82	.66	.25	.36	.35	.38
10	**.78**	**.65**	**.58**	**.82**	**.70**	**.34**	**.42**	**.41**	**.43**
12	.68	.68	.58	.77	.68	.26	.37	.36	.37
14	.63	.73	.60	.76	.69	.27	.38	.36	.36
16	.59	.75	.59	.74	.68	.23	.35	.34	.33
18	.59	.76	.61	.74	.69	.25	.36	.35	.35
20	.51	.79	.61	.72	.69	.20	.32	.32	.31
22	.46	.83	.62	.71	.69	.18	.30	.31	.29
24	.42	.86	.65	.70	.69	.16	.27	.29	.27
26	.34	.91	.69	.69	.69	.13	.23	.27	.25
28	.29	.91	.70	.67	.67	.09	.19	.22	.20
30	.20	.92	.61	.65	.64	.03	.10	.13	.12
32	.17	.92	.58	.64	.63	.02	.08	.11	.09
34	.17	.92	.58	.64	.63	.02	.08	.11	.09
36	.17	.92	.58	.64	.63	.02	.08	.11	.09
38	.12	.94	.55	.63	.62	.01	.05	.07	.06
40	.12	.95	.62	.63	.63	.01	.05	.09	.07