Published in final edited form as:
Psychiatry Res. 2013 February 28; 205(3): 253–261. doi:10.1016/j.psychres.2012.09.019.

Psychometric properties of the treatment-emergent activation and suicidality assessment profile (TEASAP) in youth with OCD

Regina Bussing[*],
Department of Psychiatry, Pediatrics and Clinical & Health Psychology, University of Florida, Gainesville, Fl

Tanya K. Murphy,
Departments of Pediatrics and Psychiatry & Behavioral Neurosciences, University of South Florida, St. Petersburg, FL

Eric A. Storch,
Departments of Pediatrics and Psychiatry & Behavioral Neurosciences, University of South Florida, St. Petersburg, FL

Joseph P.H. McNamara,
Department of Psychiatry, University of Florida, Gainesville, Fl

Adam M. Reid,
Department of Psychiatry, University of Florida, Gainesville, Fl

Cynthia W. Garvan, and
School of Human Development and Organizational Studies in Education, University of Florida, Gainesville, Fl

Wayne K. Goodman
Department of Psychiatry, Mount Sinai School of Medicine, New York, NY

Abstract

This study evaluated the psychometric properties of the treatment-emergent activation and suicidality assessment profile (TEASAP) in a clinical sample of 56 youth ages 7 to 17 with obsessive compulsive disorder (OCD) who participated in a double-blind randomized controlled trial. The 38-item TEASAP demonstrated good internal consistency for its total score ($\alpha = 0.93$) and adequate to good performance for its five subscale scores ($\alpha = 0.65$ to 0.92). One week test-retest stability (N = 18) was adequate (Intraclass correlation coefficient [ICC] = 0.68 to 0.80) except for Self-Injury (ICC = 0.46). Construct validity was supported by total and subscale TEASAP score relationships with related constructs, including irritability, hyperactivity, externalizing behaviors, manic symptoms, and suicidal ideation, and the absence of relationships with unrelated constructs. Predictive validity was established for the Disinhibition subscale through significant associations with subsequent activation events. Furthermore, TEASAP sensitivity to change in activation scores over time was supported by longitudinal associations of TEASAP scores with clinician ratings of activation over the course of treatment. Findings indicate

Corresponding author: Regina Bussing, Box 100234, UFHSC, Gainesville, FL 32610-0234, Phone: (352) 273-7550, Fax: (352) 846-1455, rbussing@ufl.edu.

that the TEASAP has acceptable psychometric properties in a clinical sample of youth with OCD and merits further study in larger samples for additional refinement of its measurement approaches.

Keywords

SSRI; activation syndrome; reliability; validity; multilevel longitudinal modeling

1. Introduction

Undesirable behavioral side effects associated with selective serotonin reuptake inhibitor (SSRI) use among children and adolescents have received considerable recent clinical and empirical attention (Bridge et al., 2007; Goodman et al., 2007). The phenomenon of transient worsening of agitation and anxiety during antidepressant treatment is not restricted to SSRIs nor to children; though first described decades ago, descriptions of the construct still vary, as summarized in a recent comprehensive review that refers to “antidepressant-induced jitteriness/anxiety syndrome” (Sinclair et al., 2009). Consistency in defining the syndrome is still lacking, even though restlessness, irritability and increased energy appear in most descriptions (Sinclair et al., 2009). Concerns over agitation, hypomania or “behavioral activation” in response to SSRI treatment in children emerged shortly after fluoxetine became available in the US (Emslie et al., 1999; King et al., 1991; Riddle et al., 1990). More recently behavioral activation in children has, been thought to manifest as any individual or combination of the following symptoms: irritability, agitation, anxiety/panic, depression, suicidality, restlessness, hostility, aggressiveness, insomnia, disinhibition, emotional lability, impulsivity, social withdrawal, restlessness, hypomania/mania, paranoia and other psychotic symptoms, or other unusual changes in behavior or mood (Murphy et al., 2008). Some evidence suggests SSRI-induced activation is linked to suicidality (Hammad, 2004), and it has been proposed that activation phenomena may represent an intermediary state change that fosters increased risk of suicidality (Murphy et al., 2008). Through a developmental process involving conceptual review, empirical work and expert consultation described in Reid et al. SSRI-related behavioral activation has been conceptualized as a worsening of a patient’s clinical presentation across one or more of five symptom domains, irritability, akathisia, disinhibition, mania and self-harm (Reid et al., 2010),

Although a number of well-validated measures exist to assess specific symptom domains associated with behavioral activation, e.g., the Columbia-Suicide Severity Rating Scale (Posner et al., 2011), to date, no specific measure assesses the range of symptoms thought to be associated with behavioral activation in youth. The development of such a measure is of critical significance for applied clinical work and research, because early identification of youth who develop activation in response to antidepressant treatment is required for rapid, targeted intervention. Availability of a validated assessment tool could allay concerns about the use of SSRIs in situations where treatment appears indicated, but apprehension about efficacious side effect monitoring prevails. Indeed, immediately after the introduction of the United States Food and Drug Administration Black Box Warning in 2004, a downward shift was witnessed in the rates of SSRI prescriptions; for instance, SSRI prescriptions decreased by approximately 22% across the Netherlands and United States between 2003 and 2005 (Gibbons et al., 2007). Yet, dramatic increases in youth suicide occurred during these periods as the Netherlands experienced an increase of 49% between 2003 and 2005, while the United States had an increase of 14% between 2003 and 2004 (Gibbons et al., 2007). This downward trend in antidepressant prescription despite generally robust efficacy and safety data is likely due, in part, to concerns about monitoring pediatric patients on SSRI

medications. However, the presence of an instrument that quickly and accurately assesses a child's changing clinical picture associated with SSRI use would have important implications for safety monitoring.

With the goal of creating a parent-report instrument sufficiently brief yet comprehensive enough to assess the range of clinical symptoms associated with behavioral activation to antidepressant use in children and adolescents, we developed the Treatment-Emergent Activation and Suicidality Assessment Profile (TEASAP). The TEASAP assesses the presence of symptoms across five domains of behavioral activation (Irritability; Akathisia, Hyperkinesis, and Somatic Anxiety; Disinhibition and Impulsivity; Mania; and Self-injury, Suicidality, and Harm to Others). The domains were initially conceptually derived, subjected to an extensive development process, and empirically supported through a pilot study of the psychometric properties of a preliminary clinician-administered version of the measure which yielded positive reliability and validity data, as well as directions for refinement (Reid et al., 2010). Most importantly, we revised the TEASAP for completion by parents instead of clinicians to ease administration and future dissemination. In pilot testing of the parent-report version, we found that (relative to the clinician-rated version) information was obtained in a more rapid manner; and quality of the information provided appeared to be comparable. Furthermore we combined frequency and intensity ratings of activation symptoms into the response format, due to their strong association and refined item content for better understandability. A copy of the parent version of the TEASAP and the instructions provided to parents is provided in Appendix 1.

This paper describes the psychometric performance of the parent-rated TEASAP as a measure of SSRI-related activation in a clinical sample of patients with obsessive compulsive disorder (OCD), prior to exposure to SSRI and in response to such exposure. OCD was selected as treatment indication for SSRI rather than depression, to minimize confounding of such activation symptoms as depression and suicidal ideation, which can represent core symptoms of depression, but not of OCD. It is important, however, to consider that OCD symptoms themselves can present as activation-like behaviors, and that OCD may be accompanied by other disorders. Indeed, none of the symptoms of the proposed activation syndrome are unique or specific to SSRI activation. Psychometric validation of the TEASAP offers unique challenges because TEASAP scores prior to SSRI exposure are quite variable. Furthermore, activation is a time-varying response and does not yield static groupings that could be subjected to traditional discriminant validity testing. We therefore conducted psychometric assessments to document traditional parameters of reliability and validity, but also utilized longitudinal modeling of TEASAP scores and activation ratings to examine the time-varying construct of activation.

Our research aims are as follows:

1. Describe variability of TEASAP Total and subscale scores prior to SSRI exposure and associations with psychiatric diagnoses.
2. Document traditional psychometric characteristics of the TEASP and its five subscales, including internal consistency, test-retest stability, concurrent (convergent and divergent) and predictive validity.
3. Establish TEASAP and subscale sensitivity to change in activation scores by examining longitudinal relationships with clinician ratings of activation.

2. Methods

2.1. Participants and procedures

Participants consisted of 56 youth enrolled for a double-blind randomized controlled 17-week OCT treatment study conducted at two sites in Florida, the University of Florida and the University of South Florida. Participants were randomized to one of three medication arms (regular sertraline titration, slow sertraline titration or placebo) and all received cognitive behavioral therapy starting at follow-up week 4 for a total of 14 CBT sessions. The sample included 22 (39%) females, 54 Caucasians (96%), five Hispanics (9%) and the average participant was 11.7 years old (S.D. = 3.3 years; range 7 to 17 years). The study was approved by the respective University Institutional Review Boards and prior to study procedures, informed consent and assent were obtained from parents and youth. Once consent/assent were obtained, a screening visit determined study eligibility, including presence of obsessive compulsive disorder for at least 6 months duration and absence of exclusionary conditions, most notably pervasive developmental disorders, mental retardation, psychosis, bipolar disorder, substance abuse or dependence within the past 6 months, seizure disorder or degenerative neurological disease. The study design allowed presence of certain comorbid disorders (depression, anxiety disorder, attention deficit hyperactivity disorder (ADHD), tic disorder) as long as OCD was the primary diagnosis. Other relevant exclusion criteria were presence of suicidal intent or a history of suicide attempt within the past 12 months. Patients on stable stimulant regimens for ADHD could continue their medication use upon study enrollment, no other concurrent psychotropic regimens were allowed. Diagnostic assessments were conducted using the Kiddie Schedule for Affective Disorders and Schizophrenia, Present and Lifetime Version (K-SADS-PL) (Kaufman et al., 1997) which yields diagnoses consistent with the Diagnostic and Statistical Manual, Fourth Edition, Text Revision (American Psychiatric Association, 2000). Of the 56 participants, 10 (18%) met diagnostic criteria for ADHD, 31 (54%) for any non-OCD anxiety disorder, 12 (21%) for tic disorder, 3 (5%) for depressive disorder, and 8 (14%) for other (elimination disorders, oppositional defiant disorder and phonological disorder). Altogether 11 (20%) of participants only met DSM-IV criteria for an OCD diagnosis, 22 (39%) for one and 23 (41%) for two or more additional comorbidities. The baseline visit was followed by 17 follow-up visits during which participants were closely assessed by blinded examiners for the occurrence of activation symptoms. All measurement scales used in this study were obtained at the baseline visit; in addition we used TEASAPs and clinician ratings of activation obtained at subsequent weekly visits for study aims 2 (predictive validity) and 3 (sensitivity to change in activation scores). Mild activation events as determined by blinded clinician rating (CGI-SA score above 2) were common and occurred in 33 (59%) of patients, including 25 (66%) of sertraline-exposed and 8 (44%) of placebo-exposed participants.

2.2. TEASAP and its development

The TEASAP (Reid et al., 2010) was developed to assess common symptoms of activation syndrome in youth due to SSRI usage. Initial scale development was based on extant literature and clinical experiences, and was followed by pilot testing with informal expert feedback as well as formal feedback through a Delphi procedure, resulting in expert consensus on the measurement constructs to be included. Originally piloted as a clinician report measure, for this study the TEASAP was refined to be a 38-item, parent-rated instrument. The TEASAP consists of five subscales representing the presumed dimensions of activation syndrome: (1) Irritability (9 items), (2) Akathisia/Hyperkinesis/Somatic Anxiety, referred to as “Akathisia” (6 items), (3) Disinhibition/Impulsivity, referred to as “Disinhibition” (7 items), (4) Mania (10 items), and (5) Self-injury/Suicidality/Harm to others, referred to as “Self-Injury” (6 items). During the baseline visit parents rate the

child's behavior in the past week, considering the frequency and impairment associated with the behavior (see Appendix 1 for instructions), on a 4-point Likert-style scale (*0 = none; 1 = mild; 2 = moderate; and 3 = severe*). Subscale scores are constructed by summing up subscale items and the TEASAP Total score equals the sum of all items. The internal consistency, temporal stability, inter-rater reliability, convergent and divergent validity of the TEASAP have been evaluated and supported in a pilot study of 30 subjects undergoing open-label treatment with SSRI for anxiety or depressive disorders (Reid et al., 2010).

2.3. Additional measures

2.3.1 Aberrant Behavior Checklist (ABC)—The ABC (Aman and Singh, 1986) is a 58-item instrument originally developed for patients with developmental disabilities. For this study, only the 15-item irritability subscale was used due to item content consistency with symptoms of activation syndrome. Parents rate their child's behavior on a 0–3 point scale ranging from 0 (*not at all a problem*) to 3 (*the problem is severe in degree*), and answers are summed to yield a subscale score. The ABC has good test-retest reliability and validity by convergent and divergent measures (Aman and Singh, 1986), as well as good inter-rater reliability (Rojahn and Helsel, 1991). The ABC Irritability Subscale was used to test convergent validity of the TEASAP Irritability subscale.

2.3.2 Barnes Akathisia Rating Scale (BARS)—The BARS is a 4-item measure of akathisia commonly administered to individuals who are receiving antipsychotic medication (Barnes, 1989, 2003). It assesses objective (observable movements) and subjective (restlessness awareness and distress) symptoms of akathisia on a scale of 0 (*lack of symptoms*) to 3 (*severe symptoms*). It also includes an overall global clinical assessment scored from 0 (*absent*) to 5 (*severe akathisia*) and yields an overall global severity score consisting of the sum of the four items (Barnes, 2003). Responses to all four items are combined to provide a total score. The reliability, validity, and clinical utility of the BARS have been examined and supported (Barnes, 1989, 2003). The BARS was used to test the convergent validity of theTEASAP akathisia subscale.

2.3.3 Behavior Rating Inventory of Executive Function (BRIEF) Parent Form—The BRIEF is an 86-item questionnaire designed to measure executive functioning in children aged 5–18. Parents rate the frequency of certain child behaviors (*never, sometimes, and often*). The BRIEF yields 8 subscales (Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor), 2 indices (Behavioral Regulation Index; Metacognition Index), and a Global Executive Composite score, with higher scores indicating higher problems. The BRIEF has high internal consistency (α = 0.80–0.98) and test-retest reliability (r = 0.81) (Malloy and Grace, 2005), as well as good clinical utility (McCandless and O'Laughlin, 2007) and validity (Reddy et al., 2011). The BRIEF Global Executive Composite score was used to test the convergent validity of the TEASAP Disinhibition subscale.

2.3.4 Child Behavior Checklist (CBCL)—The CBCL for Ages 6–18 is a 119-item checklist that assesses specific child behaviors from the parent's perspective. Parents rate their child's behaviors on a scale from 0 (*not true*) to 2 (*very true or often true*). The CBCL yields a total problem score, two broadband scores (internalizing and externalizing) and eight syndrome scales (Achenbach, 2001). The CBCL has good psychometric properties (Aschenbrand et al., 2005; Dedrick et al., 2008) and clinical utility (Eimecke et al., 2011). The CBCL Externalizing Subscale was used to test the convergent validity of the total TEASAP score.

2.3.5 Children's Depression Rating Scale-Revised (CDRS-R)—The CDRS-R is a clinician-administered instrument to diagnose youth depression and assess its severity (Poznanski and Mokros, 1996). Through semi-structured interviews with adult caregiver and child the interviewer rates 17 symptom areas of depression, including those that comprise DSM-IV criteria for a diagnosis of depression. Specific symptom areas include suicidal ideation and morbid thinking. Reported interrater reliability (r=0.92), test-retest reliability (r=0.80), and criterion-related validity of the CDRS-R are strong (Poznanski and Mokros, 1996), as is construct validity (Mayes et al., 2010). The CDRS-R was used to test the divergent validity of the total TEASAP score and the CDRS-R suicide question to test the convergent validity of the Self-Injury subscale.

2.3.6 Children's Yale-Brown Obsessive-Compulsive Scale (CY-BOCS)—The CYBOCS (Scahill et al., 1997) is a semi-structured, clinician rated instrument to measure OCD symptom severity in youth. The CY-BOCS contains a symptom checklist and a severity scale. Through the symptom checklist the clinician assesses current and past experiences of over 60 potential obsessions and compulsions. The severity scale includes 5-item subscales for obsessions and compulsions where the clinician rates time spent, interference, distress, resistance and control on a range from 0 (*none*) to 4 (*extreme*). Summing of the subscales produces an overall CY-BOCS score. Studies have documented good psychometric properties of the CY-BOCS (Gallant et al., 2008; Scahill et al., 1997; Storch et al., 2004). The CYBOCS was used to test whether baseline TEASAP scores vary with child diagnostic status, specifically severity of OCD.

2.3.7 Clinical Global Impression – Severity of Activation (CGI-SA)—The CGI-SA was adapted from the Clinical Global Impressions – Severity of Illness (CGI-SI) rating (Guy, 1976). The CGI-SI is commonly used in clinical studies of children and adults and has been extensively validated (Zaider et al., 2003). On the CGI-SA clinicians rate the severity of activation symptoms on a range from 0 (*no activation*) to 7 (*extremely severe symptoms, functionally highly impaired and/or extreme distress*). CGI-SA scores of 1 indicated that activation was doubtful, transient and without functional impairment or distress; however, such scores prompted collection of further information. Activation was considered present with CGI-SA scores of 2 (*mild symptoms, little functional impairment and/or minimal distress*) or greater. The CGI-SA was used to test the predictive validity of the TEASAP Total and subscale scores as well as sensitivity to changes over time.

2.3.8 Multidimensional Anxiety Scale for Children (MASC)—The MASC (March, 1997) is a 39-item self-report questionnaire to assess anxiety symptoms of children aged 8–19 years. Each item is rated on a 4-point Likert scale from 0 (*Never true about me*) to 3 (*Often true about me*). The MASC yields a total score and 4 subscales, Physical Symptoms (tense/somatic), Harm Avoidance (perfectionism/anxious coping), Social Anxiety (humiliation/performance fears), and Separation Anxiety/Panic (March et al., 1997). Scores are converted into T-scores in order to compare the child's score to a normative group (March, 1997). Psychometric properties including reliability (March et al., 1997; March et al., 1999) and validity (March et al., 1997; Rynn et al., 2006) have been established. Because general anxiety was not considered to be a domain of activation syndrom the MASC was used to test the divergent validity of the total TEASAP score.

2.3.9 Swanson Nolan and Pelham, Version IV (SNAP-IV)—The SNAP-IV is a rating scale consisting of operationalized DSM-IV criteria for ADHD and assesses symptoms of inattention (9 items) and hyperactivity/impulsivity (9 items). Each item is rated by parents on a 4-point Likert scale based on intensity (0 = not at all, 1 = just a little, 2 = pretty much, and 3 = very much); items are summed to produce subscale and total scores.

Internal consistency of the original SNAP-IV was reportedly high (> 0.9 for all symptom clusters), and 2-week test-retest reliability was 0.7 for inattention items, 0.8 for impulsivity items, and 0.9 for hyperactivity items (Swanson, 1992). Norms and diagnostic utility have also been established (Bussing et al., 2008). We used the SNAP-IV inattention scores to test the convergent validity of Disinhibition subscale and the SNAP-IV hyperactivity/impulsivity scores for the akathisia and Disinhibition subscale scores.

2.3.10 Suicidal Ideation Questionnaire- Junior High School Version (SIQ-JR)— The SIQ-JR is a 15-item version of the original SIQ (30 items), designed to measure a wide range of thoughts related to death and dying, passive and active suicidal ideation, and suicidal intent in younger adolescents (Reynolds, 1988). All 15 items are rated on a 7-point Likert scale ranging from 0 (“I never had this thought”) to 6 (“This thought was in my mind almost every day”). The SIQ-JR has good internal consistency as well as adequate test-retest reliability and construct validity (Reynolds, 1988). The SIQ-JR was used to test the convergent validity of the Self-Injury subscale.

2.3.11 Young Mania Rating Scale- Parent Version (YMRS-P)—The YMRS-P is an 11-item parent-rated measure of mania symptoms (Gracious et al., 2002) with five explicitly defined severity grades. The 11 items are summed to yield a total score ranging from 0–60, with higher scores indicating greater severity. Good to excellent psychometric properties, including inter-rater reliability, internal consistency and discriminative validity have been supported (Gracious et al., 2002; Marchand et al., 2005; Youngstrom et al., 2002) (Pavuluri, 2006). The YMRS-P was used to test the convergent validity of the mania subscale.

2.4 Analysis

For Aim 1, Wilcoxon tests were conducted to compare TEASAP total and subscale scores by comorbidity status and Spearman correlates were conducted to test associations between TEASAP and OCD severity.

For Aim 2, to assess internal consistency reliability, coefficient alphas were calculated for the 38- item total TEASAP and for the five subscales. By convention, a lenient cut-off of 0.60 is common in exploratory research; alpha should be at least 0.70 or higher to retain an item in an "adequate" scale; and a cut-off of 0.80 is required for a "good scale" (Litwin, 2002). One week test-retest stability was examined using Intraclass Correlations (ICCs) in a subsample of 18 adolescents. It is suggested that any measure should have an ICC of at least 0.6 to be useful (Chinn, 1991). Convergent construct validity was assessed by examining associations between TEASAP Total and subscale scores and established measures of emotional and behavioral constructs related to activation phenomena as outlined above for each measure. Conversely, divergent validity was assessed through the absence of such associations with constructs hypothesized to be unrelated to activation syndrome, including anxiety and depression (assessed through MASC, CDI and CDRS-R, respectively). The TEASAP’s ability to predict activation events was examined testing associations of its baseline total and subscale scores with subsequent first-time activation events as determined by clinician CGI-SA ratings of 2 or above, using the Wilcoxon t approximation, two-sided.

For Aim 3 multilevel modeling (Singer and Willett, 2003) was utilized to test the ability of the TEASAP to capture activation severity throughout the 17 week study. This technique allows for the analysis not only of group level effects (such as in regression or factor analysis), but additionally individual level effects, and thus has recently emerged as an innovative psychometric tool (Wilhelm and Schoebi, 2007). When used to establish validity, multilevel modeling provides a unique ability to investigate a scale’s sensitivity to change over time; our analysis aimed to not only capture the ability of the TEASAP to explain

average changes in activation throughout treatment, but also sensitivity to week-by-week changes in activation scores within each subject.

3. Results

3.1 Variations of TEASAP Total and subscale scores prior to SSRI exposure and associations with psychiatric diagnoses

TEASAP Total and subscale scores varied considerably among study participants at baseline (see Table 1), but did not vary by presence of individual psychiatric comorbid diagnosis of ADHD, any anxiety disorder, depressive disorder, or tic disorder. However, TEASAP Total, Irritability, Disinhibition, and Self-Injury scores were significantly lower for youth with OCD only (n=11) than for those with OCD plus any comorbidity (N=45) (12.2 versus 26.8, p=0.003; 5.2 versus 11.9, p=0.003; 0.3 versus 1.6; p=0.029; 1.8 versus 5.7, p=0.021; respectively). Furthermore, OCD severity as measured by CYBOCS total score correlated with the TEASAP total, Irritability and Self-Injury scores (r=0.29, p=0.031; r=0.38, p=0.004; r=0.29, p=0.031, respectively).

3.2 Internal consistency reliability

Coefficient alphas for the total 38 items and four of the five subscales suggested good internal consistency reliability, and internal consistency estimates did not increase with removal of any items (see Table 2). However, the Self-Injury subscale only achieved an alpha coefficient of 0.67, with one item (preparatory action such as giving away possessions or farewell letters) occurring infrequently and showing a low correlation of 0.175 with total subscale score.

3.3 Test-retest stability

Test-retest correlations generally supported the temporal stability for four of five subscales, except Self-Injury, and for the overall measure. One-week test-retest ICC for the overall score was 0.80; ICCs for the subscales ranged from 0.46 to 0.80 (see Table 2).

3.4 Concurrent validity: convergent and divergent

3.4.1 Convergent validity—Convergent validity was supported by finding hypothesized specific significant correlations outlined above in our measure descriptions for each of the TEASAP subscales and the total score (see Table 3, which displays originally hypothesized relationships that are confirmed in **BOLD** and identifies unconfirmed relationships by underlining). Strong correlations, for example, were found between the TEASAP Irritability subscale and the ABC-Irritability subscale (r=0.80, p<0.000); the TEASAP akathisia subscale and the SNAP Hyperactivity-Impulsivity scale (r=0.48, p<0.0002); the TEASAP Disinhibition subscale and the SNAP Hyperactivity-Impulsivity scale (r=0.70, p<0.000) and the BRIEF Global Executive Composite Score (r=0.59, p<0.000); the TEASAP Mania subscale and the YMRS-P (r=0.58, p<0.000); and the TEASAP Self-Injury subscale and the CDRS-R suicide question (r=0.27, p=0.043). Of note, the hypothesized correlation between the Barnes CGA and the TEASAP akathisia subscale was not confirmed, nor was the correlation between the SIQ-JR and the TEASAP Self-Injury subscale. Furthermore, total TEASAP scores correlated as expected with CBCL externalizing scores, and additionally with all other measures except the Barnes CGA, SIQ-JR and the CDRS-R suicide question.

3.4.2 Divergent validity—Divergent validity was supported by the absence of correlations between TEASAP Total and subscale scores and the unrelated constructs of anxiety and depression. Only one small correlation between CDRS-R and TEASAP Irritability emerged (see Table 3).

3.5 Predictive validity

Of the five subscale scores, only one was predictive of activation; specifically, higher baseline Disinhibition scores were associated with subsequent activation events (see Table 4). The total TEASAP score was not associated with activation events.

3.6 Multilevel longitudinal analysis to test sensitivity to activation change over time

Multilevel analysis included 5 nested models: 1) an unconditional means model (UMM), 2) an unconditional growth model (Model A), 3) an age model (Model B), 4) a medication arm model (Model C) and 5) the TEASAP model (Model D). The TEASAP model first included the TEASAP total score (Model D1) and then was exactly replicated using the individual TEASAP subscales in place of the total score variables (Model D2). A baseline-centered (rather than mean-centered) version of each TEASAP scale was calculated in order to capture variability in activation from the intake session (not impacted by any treatment effects) throughout the pharmacological and psychological treatment regimen. At convergence, all estimates were *z*-transformed to allow for direct comparison of *all* predictors and to gauge relative effect sizes. The details surrounding the implementation of multi-level modeling in this analysis, as well as a description of the unconditional growth model (Model A) and the covariate models (Model B and C), can be found in Appendix 2. It should be noted that multilevel modeling procedures utilized parallel the conventions outlined by Singer and Willett (Singer and Willett, 2003). A detailed description of the results for all models can be found in Table 5. In the description of the results below, *Fixed Effects* refers to between-subject variability in activation and *Random Effects* refers to within-subject variability in activation over the course of treatment.

3.6.1 TEASAP Total Score Model (Model D1)—To provide validational evidence of the TEASAP's ability to capture average activation scores over the course of treatment, as well as week-week sensitivity to change in activation scores, the TEASAP total score and a baseline-centered version of the total score were entered into the model simultaneously. The TEASAP total score model resulted in a significant decrease in the -2LL from 2310.361 to 2170.982 (χ^2 (3, N =53) =80.24, p <0.000). The overall R^2 increased by 7.9% from the linear growth model (Model B and C dropped). Both these fit indices suggest the TEASAP Total Score composite has longitudinal sensitivity to activation.

3.6.1.1 TEASAP Total Score Model Fixed Effects: Two variables were entered as *Fixed Effect* predictors in this model: The TEASAP Total Score mean and a baseline-centered TEASAP Total Score. These *Fixed Effect* predictors explained 50.66% of the between-subject variability in activation. A one standard deviation *increase in the TEASAP total score* significantly predicted a 0.273 (roughly one-fourth standard deviation) increase in activation (p <0.000), on average. Likewise, a one standard deviation *increase from an individual's baseline TEASAP total score* significantly predicted a 0.271 increase in activation symptomology (p =0.011), on average. Both of these predictors can be considered to have small effect sizes. After the addition of the TEASAP total score, no significant variance in the intercept of activation remained, supporting the utility of the TEASAP in capturing average activation during a multi-week pharmacological and psychotherapy treatment trial.

3.6.1.2 TEASAP Total Score Model Random Effects: One *Random Effects* predictor was entered into the TEASAP Total Score model: a baseline-centered TEASAP Total Score. This predictor explained 11.13% of the within-subject variability in activation. A one standard deviation increase in week-to-week variability from an individual's baseline TEASAP total score significantly predicted a 0.272 (roughly one-forth standard deviation)

increase in week-to-week variability from an individual's baseline activation. This predictor can be considered to have a small effect size.

3.6.2 TEASAP Subscale Model (Model D2)—In order to validate the TEASAP subscales ability to capture activation scores over the course of treatment, as well as week-week sensitivity to change in activation scores, the TEASAP score for each subscale was entered into the model (*Fixed Effect* predictors), along with its respective baseline-centered form (*Random Effect* predictors). The TEASAP subscales resulted in a significant decrease in the -2LL from 2310.361 to 2170.982 (χ^2 (10, N =46) = 139.379, p <0.000) and caused the overall R^2 to increase by 23.7% from the linear growth model. Taken together, these fit indices provide empirical support for the longitudinal sensitivity of activation of the TEASAP subscales.

3.6.2.1 TEASAP Subscale Model Fixed Effects: The subscales of the TEASAP were entered as orthagonalized *Fixed Effect* predictors and explained 44.5% of the *Fixed Effects* variance. Self-Injury proved to be the most sensitive to changes in average activation (p <0.000), where a one standard deviation increase in Self-Injury predicted a 0.261 (roughly one-fourth a standard deviation) increase in activation, on average. Disinhibition (0.240, *p* <0.000) and Akathisia (0.128, p <0.05) were also significant predictors of change in average activation, and all three subscales could be said to have small effect sizes. The Irritability subscale trended toward significance (p =0.077), while Mania was not a significant predictor (p =0.727). After the addition of the TEASAP subscale scores, no significant variance in the intercept of activation remained, supporting the utility of the TEASAP subscales in capturing average activation during a multi-week pharmacological and psychotherapy treatment trial.

3.6.2.2 TEASAP Subscale Model Random Effects: All the baseline-centered versions of the TEASAP subscales were entered as *Fixed Effect* predictors and explained 36.09% of the *Random Effect* variance. The most sensitive scale to week-to-week fluctuations in activation was the Disinhibition subscale (p <0.000), where a one standard deviation increase in week-to-week variability of Disinhibition predicted a 0.733 (roughly three-fourths a standard deviation) increase in activation variability. This can be considered a large effect size. Mania (p <0.01) and Self-Injury (p <0.05) both captured weekly fluctuations in activation with medium effect sizes. A one standard deviation increase in week-to-week variability of Mania predicted a 0.354 (roughly a one-third standard deviation) increase in activation variability. Self Injury captured a similar level of week-to-week fluctuation in activation (p =0.401). The Irritability (p =0.071) and Akathisia (p =0.051) subscales were both trending toward significance.

4. Discussion

Our study investigated the psychometric properties of the TEASAP, a measure designed to capture relevant domains of SSRI activation syndrome, in the context of SSRI treatment for OCD in a pediatric sample. Study results provide further strong support for this measure; internal consistency, item selection, test-retest reliability and construct validity (i.e., both convergent and divergent validity) of the TEASAP and most of its subscales were found acceptable. Even though the Self-Injury scale exhibited several shortcomings in traditional psychometric analyses, it proved significant in the longitudinal analysis of sensitivity to change and is discussed in more detail below. Results from the multilevel modeling indicate that the both the TEASAP Total score and subscales are sensitive to average activation severity and week-to-week fluctuations in treatment-emergent activation symptomology that occurs over the course of an SSRI pharmacological trial.

TEASAP scores varied considerably between patients prior to SSRI exposure, indicating that the TEASAP captures behaviors present before activation phenomena occur. Our study findings indicate that patients with more severe OCD and with higher levels of comorbidities were more likely to have increased baseline TEASAP scores. Yet, increased baseline TEASAP Total scores did not increase the risk for subsequent activation phenomena - only baseline Disinhibition scores predicted risk of activation syndrome. The Disinhibition subscale also exhibited significant sensitivity to week-to-week fluctuations in treatment-emergent activation, discussed in more detail below.

Our traditional psychometric analyses supported existing item selection and subscale construction of the TEASAP for the most part. An important exception was the Self-Injury scale. Results of the test of internal consistency, test-retest reliability, and convergent validity for the Self-Injury subscale provoke questions of this subscale's performance and suggest the need for further subscale refinement, informed by current research on best assessment practices for this complex phenomenon (Mundt et al., 2010; Plener et al., 2012; Posner et al., 2011). However, Self-Injury displayed the highest sensitivity to average activation scores across the study and the third highest sensitivity to week-to-week fluctuations behind Disinhibition and Mania in the MLM analyses. Thus, while subscale refinement is still warranted, the utility of the Self-Injury subscale has considerable empirical support. Considering this Self-Injury subscale's importance, with suicidal ideation or suicide attempts representing the worst-case outcomes of activation syndrome (Goodman et al., 2007), we suggest that further measurement development work must focus on improving the assessment of behaviors signaling injurious or suicidal risks.

The Disinhibition subscale's sensitivity to week-to-week fluctuations in treatment-emergent activation suggests that disinhibition is the best short-term indicator of a child becoming more activated while on an SSRI. One possible explanation for this finding is that disinhibition may be an initial symptom of activation that is prevalent during low-moderate activation severity and thus is more predictive of weekly changes in activation. Additionally, disinhibition has a consistently high correlation with other symptoms of activation on the TEASAP (e.g., Mania) and thus could represent a foundational symptom of activation which increases the development of other activation symptom domains (Reid et al., 2010). This hypothesis would align with research that has found that disinhibition is the most widely observed activation symptom during the use of multiple SSRI's (Gualtieri and Johnson, 2006).

Study findings also raise important questions about the assessment of akathisia/hyperkinesis/somatic anxiety. The TEASAP Akathisia subscale had adequate internal consistency and test-retest reliability, but hypothesized relationships with the "gold standard" assessment for akathisia in form of the BARS were not confirmed. However, the subscale scores strongly correlated with measures of hyperactivity or externalizing behaviors, like the SNAP and CBCL-ext. The BARS reportedly performs well for akathisia assessment in youth (Zalsman et al., 2011); however, more typically in the context of studies assessing antipsychotic mediation effects (e.g., (Gebhardt et al., 2006). Zalsman et al. suggested that the BARS needs to distinguish restlessness associated with ADHD, which is intentional, goal-directed and not associated with subjective distress, from akathisia (Zalsman et al., 2011). We question whether SSRI activation syndrome presents with hyperactivity symptoms more akin to ADHD symptomology than with the experience of akathisia. Of note, in an instrument that seeks to track SSRI responses in adults, akathisia was not included as a construct and rating items such as feelings of restlessness or inability to sit still were attributed to an anxiety factor (Trivedi et al., 2011). Anxiety is also presumed to have a more significant role in activation syndrome in a recent systematic review referring to "antidepressant-induced jitteriness/anxiety syndrome" (Sinclair et al., 2009). Further

refinement of the construct of akathisia/hyperkinesis/somatic anxiety as side-effect domain in SSRI treatment and its measurement appear indicated.

4.1 Limitations and future directions

Due to sample size limitations we could not conduct a confirmatory factor analysis to further substantiate the TEASAP subscale constructs. Furthermore, whereas the concurrent validity analyses were conducted using well-established instruments, predictive and discriminative analyses were conducted using a newly developed adaptation of the CGI, the CGI-SA that assesses activation events based on clinicians' best judgment. Lastly, validation of the TEASAP occurred in the context of an OCD treatment study that also served the purpose of further elucidating the construct of SSRI-related activation phenomena. In other words, measurement and construct development occurred simultaneously, which limits the immediate application and generalization of findings.

Future work should include normative studies in nonclinical and clinical populations to document expected normal variability of TEASAP scores. Further content refinement is indicated to enhance TEASAP utilization in clinical settings, in particular to allow prospective identification of those at increased activation risk during SSRI treatment. Such risk status determination will allow identification of youth that need closest monitoring for early signs of adverse events.

4.2 Clinical implications

Although our findings should be interpreted with these limitations in mind, our study demonstrates that the TEASAP captures relevant domains of activation syndrome. In combination with the predictive validity findings, the results of this study suggest that youth exhibiting symptom constellations of high disinhibition prior to *or* during pharmacological treatment should be closely monitored and dose increases be made particularly carefully, also considering the data linking higher behavioral disinhibition and attempted suicide in youth (Dougherty et al., 2009). The FDA Black Box warning notes that youth started on SSRI treatments should be closely monitored for unusual behaviors, worsening of depression and suicidal ideation. Parent-reported TEASAP scores corresponded with relevant "gold standard" measures and were sensitive to physician-rated activation scores completed by child and adolescent psychiatrists, suggesting that this instrument could make relevant clinical contributions in the SSRI monitoring process in youth.

Acknowledgments

This research was supported by grant 5UO1 MH078594 from the NIMH. The authors thank study coordinators Dana Mason and Jeannette Reid, all staff members who contributed to data collection, the families for their participation, and Drs. Gary R. Geffken, Ayesha Lall, Jane Mutch and Omar Raman for their contribution to the study interventions.

References

Achenbach, TM. Child Behavior Checklist for Ages 6–18. Vermont: Achenbach System of Empirically Based Assessment; 2001.

Aman, MG.; Singh, NN. Aberrant Behavior Checklist Manual. New York: Slosson Educational Publications, Inc.; 1986.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. District of Columbia: American Psychiatric Association; 2000.

Aschenbrand SG, Angelosante AG, Kendall PC. Discriminant validity and clinical utility of the CBCL with anxiety-disordered youth. Journal of Clinical Child and Adolescent Psychology. 2005; 34:735–746. [PubMed: 16232070]

Barnes TR. A rating scale for drug-induced akathisia. British Journal of Psychiatry. 1989; 154:672–676. [PubMed: 2574607]

Barnes TR. The Barnes Akathisia Rating Scale--revisited. Journal of Psychopharmacology. 2003; 17:365–370. [PubMed: 14870947]

Bridge JA, Iyengar S, Salary CB, Barbe RP, Birmaher B, Pincus HA, Ren L, Brent DA. Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment: a meta-analysis of randomized controlled trials. Journal of the American Medical Association. 2007; 297:1683–1696. [PubMed: 17440145]

Bussing R, Fernandez M, Harwood M, Hou W, Garvan CW, Eyberg SM, Swanson JM. Parent and teacher SNAP-IV ratings of attention deficit hyperactivity disorder symptoms: Psychometric properties and normative ratings from a school district sample. Assessment. 2008; 15:317–328. [PubMed: 18310593]

Chinn S. Statistics in respiratory medicine. 2. Repeatability and method comparison. Thorax. 1991; 46:454–456. [PubMed: 1858087]

Dedrick RF, Tan TX, Marfo K. Factor structure of the Child Behavior Checklist/6–18 in a sample of girls adopted from China. Psychological Assessment. 2008; 20:70–75. [PubMed: 18315401]

Dougherty DM, Mathias CW, Marsh-Richard DM, Prevette KN, Dawes MA, Hatzis ES, Palmes G, Nouvion SO. Impulsivity and clinical symptoms among adolescents with non-suicidal self-injury with or without attempted suicide. Psychiatry Research. 2009; 169:22–27. [PubMed: 19631392]

Eimecke SD, Remschmidt H, Mattejat F. Utility of the Child Behavior Checklist in screening depressive disorders within clinical samples. Journal of Affective Disorders. 2011; 129:191–197. [PubMed: 20825996]

Emslie GJ, Walkup JT, Pliszka SR, Ernst M. Nontricyclic antidepressants: current trends in children and adolescents. Journal of the American Academy of Child and Adolescent Psychiatry. 1999; 38:517–528. [PubMed: 10230183]

Gallant J, Storch EA, Merlo LJ, Ricketts ED, Geffken GR, Goodman WK, Murphy TK. Convergent and discriminant validity of the Children's Yale-Brown Obsessive Compulsive Scale-Symptom Checklist. Journal of Anxiety Disorders. 2008; 22:1369–1376. [PubMed: 18329843]

Gebhardt S, Hartling F, Hanke M, Mittendorf M, Theisen FM, Wolf-Ostermann K, Grant P, Martin M, Fleischhaker C, Schulz E, Remschmidt H. Prevalence of movement disorders in adolescent patients with schizophrenia and in relationship to predominantly atypical antipsychotic treatment. European Child and Adolescent Psychiatry. 2006; 15:371–382. [PubMed: 16648965]

Gibbons RD, Brown CH, Hur K, Marcus SM, Bhaumik DK, Erkens JA, Herings RM, Mann JJ. Early evidence on the effects of regulators' suicidality warnings on SSRI prescriptions and suicide in children and adolescents. The American Journal of Psychiatry. 2007; 164:1356–1363. [PubMed: 17728420]

Goodman WK, Murphy TK, Storch EA. Risk of adverse behavioral effects with pediatric use of antidepressants. Psychopharmacology. 2007; 191:87–96. [PubMed: 17180618]

Gracious BL, Youngstrom EA, Findling RL, Calabrese JR. Discriminative validity of a parent version of the Young Mania Rating Scale. Journal of the American Academy of Child and Adolescent Psychiatry. 2002; 41:1350–1359. [PubMed: 12410078]

Gualtieri CT, Johnson LG. Antidepressant side effects in children and adolescents. Journal of Child and Adolescent Psychopharmacology. 2006; 16:147–157. [PubMed: 16553535]

Guy, W. Guy, W. ECDEU Assessment Manual for Psychopharmacology – Revised. Maryland: U.S. Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, NIMH Psychopharmacology Research Branch, Division of Extramural Research Programs; 1976. Clinical Global Impression Improvement Scale; p. 217-221.

Hammad, TA. Results of the analysis of suicidality in pediatric trials of newer antidepressants; Presented at: FDA joint Psychopharmacologic Drugs Advisory Committee and the Pediatric Advisory Committee meeting; 2004.

Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. Journal of the American Academy of Child and Adolescent Psychiatry. 1997; 36:980–988. [PubMed: 9204677]

King RA, Riddle MA, Chappell PB, Hardin MT, Anderson GM, Lombroso P, Scahill L. Emergence of self-destructive phenomena in children and adolescents during fluoxetine treatment. Journal of the American Academy of Child and Adolescent Psychiatry. 1991; 30:179–186. [PubMed: 2016219]

Litwin, MS. How to assess and interpret survey psychometrics. California: Sage Publications; 2002.

Malloy P, Grace J. A Review of Rating Scales for Measuring Behavior Change Due to Frontal Systems Damage. Cognitive and Behavioral Neurology. 2005; 18:18–27. [PubMed: 15761273]

March, JS. Multidimensional Anxiety Scale for Children: Technical Manual. New York: Multi-Health Systems Inc.; 1997.

March JS, Parker JDA, Sullivan K, Stallings P, Conners CK. The Multidimensional Anxiety Scale for Children (MASC): Factor Structure, Reliability, and Validity. Journal of the American Academy of Child & Adolescent Psychiatry. 1997; 36:554–565. [PubMed: 9100431]

March JS, Sullivan K, Parker J. Test-Retest Reliability of the Multidimensional Anxiety Scale for Children. Journal of Anxiety Disorders. 1999; 13:349–358. [PubMed: 10504106]

Marchand WR, Clark SC, Wirth L, Simon C. Validity of the parent young mania rating scale in a community mental health setting. Psychiatry. 2005; 2:31–35. [PubMed: 21179627]

Mayes TL, Bernstein IH, Haley CL, Kennard BD, Emslie GJ. Psychometric properties of the Children's Depression Rating Scale–Revised in adolescents. Journal of Child and Adolescent Psychopharmacology. 2010; 20:513–516. [PubMed: 21186970]

McCandless S, O'Laughlin L. The clinical utility of the Behavior Rating Inventory of Executive Function (BRIEF) in the diagnosis of ADHD. Journal of Attention Disorders. 2007; 10:381–389. [PubMed: 17449837]

Mundt JC, Greist JH, Gelenberg AJ, Katzelnick DJ, Jefferson JW, Modell JG. Feasibility and validation of a computer-automated Columbia-Suicide Severity Rating Scale using interactive voice response technology. Journal of Psychiatry Research. 2010; 44:1224–1228.

Murphy TK, Segarra A, Storch EA, Goodman WK. SSRI adverse events: how to monitor and manage. International Review of Psychiatry. 2008; 20:203–208. [PubMed: 18386213]

Plener PL, Straub J, Kapusta ND, Fegert JM, Sprober N. Assessment of suicidal ideation in adolescents: comparison of two instruments. Praxis der Kinderpsychologie und Kinderpsychiatrie. 2012; 61:4–15. [PubMed: 22400375]

Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, Currier GW, Melvin GA, Greenhill L, Shen S, Mann JJ. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. The American Journal of Psychiatry. 2011; 168:1266–1277. [PubMed: 22193671]

Poznanski, EO.; Mokros, HB. Children's Depression Rating Scale-Revised (CDRS-R) Manual. California: Western Psychological Services; 1996.

Reddy LA, Hale JB, Brodzinsky LK. Discriminant validity of the Behavior Rating Inventory of Executive Function Parent Form for children with attention-deficit/hyperactivity disorder. School Psychology Quarterly. 2011; 26:45–55.

Reid J, Storch E, Murphy T, Bodzin D, Mutch P, Lehmkuhl H, Aman M, Goodman W. Development and Psychometric Evaluation of the Treatment-Emergent Activation and Suicidality Assessment Profile. Child and Youth Care Forum. 2010; 39:113–124. [PubMed: 20473344]

Reynolds, WM. Suicidal Ideation Questionnaire (SIQ): Professional Manual. Florida: Psycholgogical Assessment Resources; 1988.

Riddle MA, Hardin MT, King R, Scahill L, Woolston JL. Fluoxetine treatment of children and adolescents with Tourette's and obsessive compulsive disorders: preliminary clinical experience. Journal of the American Academy of Child and Adolescent Psychiatry. 1990; 29:45–48. [PubMed: 2295577]

Rojahn J, Helsel W. The aberrant behavior checklist with children and adolescents with dual diagnosis. Journal of Autism and Developmental Disorders. 1991; 21:17–28. [PubMed: 2037547]

Rynn MA, Barber JP, Khalid-Khan S, Siqueland L, Dembiski M, McCarthy KS, Gallop R. The psychometric properties of the MASC in a pediatric psychiatric sample. Journal of Anxiety Disorders. 2006; 20:139–157. [PubMed: 16464701]

Scahill L, Riddle MA, McSwiggin-Hardin M, Ort SI, King RA, Goodman WK, Cicchetti D, 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]

Sinclair LI, Christmas DM, Hood SD, Potokar JP, Robertson A, Isaac A, Srivastava S, Nutt DJ, Davies SJ. Antidepressant-induced jitteriness/anxiety syndrome: systematic review. British Journal of Psychiatry. 2009; 194:483–490. [PubMed: 19478285]

Singer, JD.; Willett, JB. Applied longitudinal analysis: Modeling change and event occurrence. New York: Oxford University Press; 2003.

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

Swanson, JM. School-based assessments and interventions for ADD students. California: K.C. Publishing; 1992.

Trivedi MH, Wisniewski SR, Morris DW, Fava M, Kurian BT, Gollan JK, Nierenberg AA, Warden D, Gaynes BN, Luther JF, Rush AJ. Concise Associated Symptoms Tracking scale: a brief self-report and clinician rating of symptoms associated with suicidality. Journal of Clinical Psychiatry. 2011; 72:765–774. [PubMed: 21733477]

Wilhelm P, Schoebi D. Assessing mood in daily life - Structural validity, sensitivity to change, and reliability of a short-scale to measure three basic dimensions of mood. European Journal of Psychological Assessment. 2007; 23:258–267.

Youngstrom EA, Danielson CK, Findling RL, Gracious BL, Calabrese JR. Factor structure of the Young Mania Rating Scale for use with youths ages 5 to 17 years. Journal of Clinical Child and Adolescent Psychology. 2002; 31:567–572. [PubMed: 12402575]

Zaider TI, Heimberg RG, Fresco 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]

Zalsman G, Opler M, Steinberg T, Apter A, Weizman A. Evaluation of side effects using the BARS, SAS and AIMS in pediatric psychopharmacologic studies. European Psychiatry. 2011; 26:128–129. [PubMed: 21036550]

Table 1

Mean scores (and standard deviations) on the TEASAP subscales at baseline, by comorbidity status

Subscale Name	Total (*n* = 56)	OCD only (*n* = 11)	Comorbid OCD (*n* = 45)	Wilcoxon test
Irritability	10.6 (6.7)	5.2 (4.7)	11.9 (6.4)	*WR* = 164.5; *p* = .003
Akathisia	3.7 (3.1)	2.6 (2.1)	4.0 (3.3)	*WR* = 260.0; *p* = .275
Disinhibition	4.9 (4.7)	1.8 (1.4)	5.7 (4.9)	*WR* = 198.0; *p* = .021
Mania	3.4 (3.8)	2.3 (2.0)	3.6 (4.1)	*WR* = 290.0; *p* = .632
Self-harm	1.3 (2.1)	0.3 (0.6)	1.6 (2.3)	*WR* = 213.5; *p* = .029
Total TEASAP	23.9 (15.9)	12.2 (8.8)	26.8 (16.0)	*WR* = 164.0; *p* = .003

Table 2

Internal consistency, test-retest reliability, and subscale correlations

Subscale Name	Number of items	Coefficient Alpha (range of Alpha if items deleted)	One week test-retest ICC (*n* = 18)	(1)	(2)	(3)	(4)	(5)
Irritability (1)	9	0.92 (0.91–0.92)	0.68	1.00	0.31	0.54	0.31	0.48
Akathisia (2)	6	0.72 (0.56–0.75)	0.80	0.31	1.00	0.58	0.52	0.32
Disinhibition (3)	7	0.84 (0.79–0.86)	0.77	0.54	0.58	1.00	0.55	0.47
Mania (4)	10	0.75 (0.69–0.78)	0.68	0.31	0.52	0.55	1.00	0.45
Self-harm (5)	6	0.65 (0.56–0.67)	0.46	0.48	0.32	0.47	0.45	1.00
Total TEASAP	38	0.93 (0.92–0.93)	0.80					

Table 3

Total TEASAP and subscale concurrent (convergent and divergent) validity

Instrument Name	Irritability		Akathisia		Disinhibition		Mania		Self-Injury		Total TEASAP	
	r	*p*	*R*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Convergent: Hypothesized related												
ABC [a]	**.80**	**<.0001**	.26	.0525	.54	<.0001	.40	.0027	.51	<.0001	.73	<.0001
Barnes CGA [b]	−.19	.1620	−.09	.5034	−.26	.0596	−.14	.3189	−.33	.0139	−.23	.0879
SNAP/Hyp-Imp [c]	.45	.0006	**.48**	**.0002**	**.70**	**<.0001**	.54	<.0001	.31	.0235	.68	<.0001
SNAP/Inatt [d]	.25	.0647	.28	.0404	**.39**	**.0035**	.44	.0007	.33	.0129	.43	.0011
BRIEF [e]	.47	.0003	.37	.0052	**.59**	**<.0001**	.59	<.0001	.48	.0002	.66	<.0001
YMRS-P [f]	.57	<.0001	.33	.0173	.58	<.0001	**.58**	**<.0001**	.40	.0030	.69	<.0001
CBCL-Ext [g]	.71	<.0001	.32	.0146	.52	<.0001	.43	.0009	.49	.0001	**.70**	**<.0001**
SIQ-JR [h]	−.03	.8108	.18	.2087	−.00	.9925	.03	.8388	.03	.8127	.05	.7202
CDRS-R Suicide [i]	−.02	.8733	.13	.3356	.05	.7171	.01	.9281	**.27**	**.0425**	.04	.7525
Divergent: Hypothesized unrelated												
MASC [j]	.02	.8960	.10	.4711	−.19	.1628	−.08	.5502	−.10	.4858	−.04	.7792
CDRS-R [k]	.27	.0491	−.04	.7518	−.00	.9852	.04	.7713	.23	.0889	.14	.2995

Notes:

[a] Aberrant Behavior Checklist – Irritability Subscale;

[b] Barnes Akathisia Rating Scale Clinical Global Assessment;

[c] Swanson Nolan and Pelham- Hyperactivity and Impulsivity,

[d] Swanson Nolan and Pelham- Inattention,

[e] Behavior Rating Inventory of Executive Function, Global Executive Composite Score,

[f] Young Mania Rating Scale – Parent Version,

[g] Child Behavior Checklist- Externalizing Problems,

[h] Suicidal Ideation Questionnaire - Junior High School Version,

[i] Children's Depression Rating Scale- Revised Suicide Subscale,

[j] Multidimensional Anxiety Scale for Children,

[k] Childrens' Depression Rating Scale-Revised. Shown in **BOLD** are originally hypothesized relationships supporting concurrent convergent validity, whereas underlined correlation results show failure to confirm hypothesized relationships.

Table 4

Total TEASAP and subscale predictive validity

Subscale Name	Subsequent Activation		Wilcoxon *T* Approximation, two-sided
	No (n=24) *M (SD)*	Yes (n=32) *M (SD)*	
Irritability	10.2 (7.2)	10.8 (6.3)	0.5420
Akathisia	3.3 (2.9)	4.0 (3.3)	0.4552
Disinhibition	3.3 (3.9)	6.2 (5.0)	0.0149
Mania	2.6 (3.2)	3.9 (4.2)	0.2073
Self-Injury	1.3 (2.6)	1.4 (1.7)	0.3474
Total TEASAP	20.7 (16.5)	26.3 (15.2)	0.1119

Psychiatry Res. Author manuscript; available in PMC 2014 February 28.

Table 5

Results of multi level model testing of TEASAP longitudinal validity

	UMM	Model A	Model B	Model C	Model D1	Model D2
Fixed Effects						
Intercept	.279(.086)**	.230(.083)**	.221(.079)**	.215(.079)**	.114(.060)	.128(.127)
Age			−.115(.072)	Dropped	Dropped	Dropped
Drug				.259(.170)[a]	Dropped	Dropped
Randomization				.193(.180)[b]		
Total Score-M					.273(.057)**	
Total Score-BC					.271(.101)**	.087(.049)+
Irritability-M						.128(.055)**
Akathisia-M						.240(.060)**
Disinhibition-M						.016(.044)
Mania-M						.261(.058)**
Self-Injury-M						
Random Effects						
Residual	.831(.043)**	.800(.042)**	.800(.042)**	.801(.042)**	.711(.039)**	.511(.035)**
Time (linear)		.048(.026)+	.048(.025)+	.049(.026)+	.042(.024)+	.012(.016)
Total Score-BC					.272(.111)**	.212(.118)+
Irritability-						.418(.214)+
						.733(.269)**

	UMM	Model A	Model B	Model C	Model D1	Model D2
BC						**
Akathisia-						.354(.178) *
BC						*
						.401(.171) *
Disinhibition-BC						
Mania-BC						
Self Injury-BC						
Fit Statistics						
-2LL	2322.419	2314.350	2311.840	2311.947	2230.117	2170.982
AIC	2328.419	2322.350	2321.840	2323.947	2244.117	2198.982
BIC	2342.616	2341.279	2345.501	2352.34	2272.192	2265.133
Overall R^2	47.80%	53.90%	53.80%	53.71%	61.80%	77.60%
ΔFixed		17.36%	3.8%	6.5%	50.66%	44.55%
Pseudo R^2		3.71%	0.05%	0%	11.13%	36.09%
ΔRandom						
Pseudo R^2						

Note

$^{+}$p<.08

*p<.05,

**p<.01,

***p<.001.

^{a}Slow titration arm,

^{b}Regular titration arm; compared to placebo arm. M= Mean, BC=Baseline Centered