Independent Evaluator (IE) Manual

RFP-NIH-NIMH 98-DS-0008

Treatment for Adolescents with Depression Study (TADS)

Final Version (4.1)

Principal Investigator: John S. March, MD, MPH, Duke University Medical Center with Duke Clinical Research Institute (DUMC/DCRI), P.O. Box 3001, Durham, NC 27710

Contact: John S. March, M.D.
Department of Child and Adolescent Psychiatry
Department of Psychology – Soc & HLTH
718 Rutherford St Durham, NC 27705
Box 3527 Med Ctr Durham, NC 27710
(p) 1-919-416-2404
(f) 1-919-416-2420
jsmarch@acpub.duke.edu

Adapted with permission from materials developed for ADHD clinical trials by C. Keith Conners, Ph.D., 1998, and from the IE Manual from the Treatment of Pediatric OCD Study (John March, Duke University, and Edna Foa, University of Pennsylvania, Co-Principal Investigators).

TABLE of CONTENTS

I. Overview of IE Role

In a multi-site trial it is especially important to have common standards for the use of the CGI and other scales, else widely divergent levels of severity and improvement could be represented at different sites due to different clinician standards for rating severity and change with treatment. Common standards for rating the main dependent measures are also necessary to insure that the same rater rating the same patient over time captures the initial status of the patient and subsequent change with treatment accurately. By reducing unreliability between and within raters, the validity of the trial is protected from method variance in the application of rating scales with a consequent loss of study power.

To this end TADS will (1) follow five important principles and (2) implement specific guidelines for generating scores on the IE-administered dependent measures. These principles are as follows:

First, when patients enter the study they must be of sufficient severity as indicated by a K-SADS diagnosis of MDD that is pervasive and stable and by an IE-administered CDRS-R Score $\geq$ 45 so that there is *room for improvement*. That is, if a patient is not impaired, then the treatment could not demonstrate any improvement that is clinically meaningful. **Hence, the initial CGI-S score should be at least moderately ill (CGI-S = 4) at the start of the study.**

Second, it is important to use interview information provided by the patient and parent, and to reconcile these sources, to generate a valid and reliable summary judgment of the patient's clinical status with respect to severity of symptoms. Specifically, it is assumed that the IE will have the benefit of (a) clinical interview information from the patient and the primary caretaker; and (b) direct observations during the interview with the patient and primary caretaker. As in the case of severity judgments, the IE must ultimately integrate information from patient and parent direct interviews into a final judgment of the degree of change noted in the patient. The IE should not use the TADS teen self-report measures (e.g., RADS, MASC, CNCE, etc.), however, to make this determination

Third, while it is possible to define severity solely in terms of a statistical criterion on standardized rating scales, this is not entirely satisfactory because one can have severe *symptoms*, without having severe *impairment*. For example, a patient can be quite depressed, but have such a good environment and personal assets that depression causes relatively little impairment in functioning at home or school. (TADS requires that to enter the study a patient must show impaired function due to MDD in at least two of three domains: school, home, social relationships.) Hence, the CGI-S and CGI-I Scales include guidelines for rating the patient on specific impairment criteria.

Fourth, with both severity and improvement ratings, it is important to have some guidelines so that Independent Evaluator's at the same or different sites will use the scales in a similar fashion. This is accomplished by following a set of guidelines or algorithms as specified in this manual.

Fifth, while IE ratings will not directly impact clinical care of the patient in TADS, the guidelines for IE and, similarly, clinician ratings must predict the dose of treatment within a clinically relevant stages of treatment framework, viz. increasing dose in Stage I and, if appropriate, Stage II and maintenance treatment during Stage III as a function of clinical status. Dose of pharmacotherapy means number of milligrams of FLX and frequency of medication visits; dose for CBT means number and frequency of sessions of CBT. This goal is met by using the **clinician (not the IE) CGI-S** to drive the pharmacotherapy dosing schedule in Stage I and II and to link the medication and CBT interventions in combination treatment.

II. IE Qualifications

Independent Evaluators (IEs) in TADS must meet one of the following professional qualifications:

- MD pediatric psychiatrist with experience administering research-related structured clinical interviews
- PhD or PsyD clinical psychologist with experience administering research-related structured clinical interviews
- DSW or MSW social worker with experience administering research-related structured clinical interviews
- Psychiatric nurse practitioner with at least 6 months of recent experience administering research-related structured clinical interviews
- Masters-level psychologist with at least 6 months of recent experience administering research-related structured clinical interviews

In addition, all IE's must have experience in making determinations related to differential diagnosis. Experience with research-related structured clinical interviews should preferably be with instruments that assess multiple diagnoses, not just one diagnosis of interest in one study. For the purpose of TADS, training and certification on the K-SADS-AD or CDRS-R will meet this requirement.

All IEs must be experienced with depression or with adolescent psychiatric patients or, preferably, both.

All IEs must agree to be available for at least 21 (full length of trial) months when assigned a new patient.

As part of the IE certification/supervision procedures outlined in the TADS QA Manual, it is expected that the IE Supervisor will review the IE Manual and the CDRS-R Manual with the site IE(s) initially and at least yearly thereafter.

III. Procedures

At all major and minor assessment points, an Independent Evaluator (IE) will obtain the following measures:

- The Children's Depression Rating Scale-Revised, which provides a scalar view of symptoms of pediatric MDD.
- The CGI-S and CGI-I assess MDD-specific severity of illness and degree of improvement since baseline.
- Used in British studies of pediatric depression, the HoNOSCA assesses both domain-specific and global functioning (administered at five major assessment timepoints).

- The CGAS targets overall functioning, including impairments stemming from MDD as well as all other causes, e.g. comorbidity or family problems.

Table 1: IE Dependent Measures

Measure	Respondent	When	Assessment Domain
CDRS-R	Child and Parent	All	Continuous measure of MDD**
CGI-I	Child and Parent	All	MDD-specific improvement*
CGI-S	Child and Parent	All	MDD-specific severity of illness*
HoNOSCA	Child and Parent	Baseline, Weeks 12, 36 and 6 months and 1 year in Stage IV	Multi-domain Global Outcome**
CGAS	Child and Parent	All	Global functioning*

*Record single summary measure by rater using data from both respondents
**Record child, parent and best judgment (combined/summary) score

During the entry procedures, the IE will have conducted the K-SADS-PL (Gate B), the K-SADS-AD (Gate C1) and, following the K-SADS-AD, the baseline CDRS-R to insure study eligibility for randomization. Thus, in most cases the IE will know the patient and parent and will have begun to inculcate a common language and culture for administering structured rating scales.

Prior to administering the K-SADS affective disorders module, the IE will also administer the Psychiatric Treatment History form at Gate B1.

Before each IE session, the IE will review the CDRS-R (total and item) and the CGI-S (but not the CGI-I) from (1) Gate C1 and (2) the immediately previous assessment to be clear about the trajectory of change in target symptoms with respect to change from baseline (T0).

Although convenience may dictate that major and minor assessment and treatment visits occur on the same day, the assessment schedule is independent of the treatment schedule. For example, the week 12 assessment visit occurs some time during week 12 irrespective of whether the patient has completed all of the scheduled Stage I treatment visits.

If the assessment and treatment visits are scheduled back-to-back, the IE visit (and preferably all other assessments too) should come before the treatment visit in order not to bias the subject's responses because of proximity (recency) effects from the treatment session.

Instructions to the IE, Clinician and Patient/Parent Regarding IE Blinding

To insure that the IE remains blind, we will put the following safeguards in place:

- At the Gate C2 visit, the site team member who reveals randomization will inform the patient/parent that the IE will not know the patient's treatment assignment and that they (patient and parent) should not tell them. Furthermore, the patient/parent should be told that this is to minimize the chance that an IE will have a favorite treatment and, even without intending to do so, will give that treatment better scores.
- Immediately before each IE visit, the patient and parent will be reminded by the study coordinator not to telegraph or to declare their treatment assignment.
- Additionally, the IE Supervisor will remind the IE to be careful not to ask questions regarding the effect of treatment.

- To remove any sense of "coaching" the parent/patient regarding their responses to the IE, TADS clinicians (CBT and pharmacotherapist) will be instructed not to bring up or to discuss the IE battery.
- Ideally, the week 12 IE visit should come before the week 12 clinical visit where a determination is made regarding whether TADS Stage I treatment continues. If not, the clinician/SC will warn the patient not to reveal continuation status at this or subsequent IE visits. An analogous caution exists for all patients undergoing premature termination.
- QA procedures for the IE and Clinicians will include reviewing procedures for insuring that the IE remains blind.
- IE blindness will be assessed at weeks 6, 12, 18, 24, 30 and 36.

IV. Guidelines for Administering the CDRS-R

The CDRS-R will be administered according to procedures outlined in the CDRS-R Manual, which for TADS purposes will be considered to be part of the IE Manual. With particular attention to the interview prompts found in Appendix A of the CDRS-R Manual, each IE is expected to have read and to be familiar with the CDRS-R Manual, including committing the CDRS items and anchors to memory.

To insure reliable and valid administration of the CDRS-R, TADS Independent Evaluators will be expected to follow the following guidelines:

- Pre- and post-randomization, the CDRS-R will be given first to the patient, then to the parent and then, if necessary, the IE will interview both together to reconcile any differences in reliability (in contrast to true differences) of reporting before assigning CGI scores.
- Based on previous (e.g. previous administration of the CDRS-R) and current experience, the IE must make a judgment about the reliability and validity of the informant. For the purpose of TADS, **reliability** is defined as the extent to which the respondent displays a consistent interpretation bias across all the CDRS-R items. **Validity** is defined as the extent to which the respondent displays an interpretation bias that reflects the actual status of the patient in contrast to amplifying (common in parents) or minimizing (common in children) symptoms. Note that reliability and validity are not the same as true differences in reporting that depend on the observing position of the respondent. For example, a child may report mid-nocturnal or terminal insomnia about which the parent may be unaware. Hence, it is also important to weight each item with respect to **who is in the best position to know**, which in most cases will also represent the higher score when the responses are reliable and valid. In this regard, it is important to review Table 3 from the CDRS-R manual, which notes that on average parents usually report higher, or less commonly, identical symptoms than children.
- The K-SADS will always be given prior to the CDRS-R. Even so, the IE still must administer the CDRS-R in its entirety to insure that the CDRS-R anchors are followed in generating a CDRS-R score. Although it is appropriate to refer back to information gained when administrating the K-SADS when administering the CDRS-R, do not assume that the K-SADS maps directly on the CDRS-R as the items are phrased differently and the response metrics are not identical.
- For all probes, the time frame for reporting for CDRS is the previous week. It is important to clarify time frame as the patient and/or parent may report responses that

reflect non-representative recency, e.g. fight with Mom immediately before the visit, or remote effects, e.g., a child who reports participating in sports or going camping as a pleasurable activity but has not have done this activity for several months.

- It is expected that the CDRS-R will be given to both parent and child in the clinic setting. Phone administration is not acceptable unless all other options have been exhausted. Similarly, it is expected that both parent and child will be present for the CDRS-R even if parent and patient show very high levels of agreement.
- In general, the CDRS-R items should be administered in the order in which they are presented in the CDRS-R scoring sheet. On the other hand, when a particular item is deemed likely to lead to an invalid response bias for the subsequent items, it is reasonable and clinically correct to alter the order in which the items are administered to place the offending item later in the sequence. For example, with a teenager who has gained considerable weight that he minimizes because it triggers intense sadness and irritability, this item may be moved to the end of the interview.
- When asking each item, first define terms that may be confusing (e.g. morbid thinking or suicidal thinking) and then ask each anchor question attempting to bracket the true response by eliminating lower and higher anchor points. Pay particular attention to elements of an anchor that may indicate the appropriateness of an even numbered score, e.g. a score that falls between the defined odd numbered anchor points. For example, a score of 6 would be appropriate on the fatigue item (number six) if the patient endorses being tired all the time, takes naps and feels at least somewhat refreshed on awakening; all other things being equal, the lack of feeling refreshed would indicate a score of 7 on this item. **This level of detail, which is essential to correct administration of the CDRS-R, requires that the IE commit the CDRS-R items and their anchor points to memory.**
- To avoid biasing the results because of social desirability factors, the IE should adopt an affectively warm but otherwise neutral pose that avoids coaching or leading the patient to what the IE might think should be the expected response. Similarly, though some degree of open dialogue is often important, in general it is best to stick to the probes and to limit otherwise empathic comments as these may bias the respondents answers.
- Given reliable and valid responses from the child and parent, the IE should choose the **most severe score** as the true score for purposes of establishing a summary judgment. When one or both respondents appear unreliable or display a consistent response bias that raises concerns about validity of responding, then the IE must make a clinical judgment regarding the best rating for the item in question. In most but not all cases, this rating should be within one point (either higher or lower) of the most severe score.
- The quality of data obtained when interviewing depressed youth is heavily dependent on the sensitivity and clinical acumen of the interviewer. In this regard, the IE is expected to be familiar with and use the CDRS-R probe questions (Appendix A in the CDRS-R Manual) in a fashion that is developmentally sensitive.
- Because many patients will report “no problem” at first, it is critical not to accept the first response, but to persist in probing across anchor points until it is clear that the subjects responses will not change with further questioning.
- Most CDRS-R items use a 7-point Likert scale. For these items, a CDRS item score of 1 or 2 is consistent with subclinical or no symptoms; a score of 3-4 with clinical

symptoms; a score of 5, 6 or 7 with severe symptoms. For the three items using a 5 point scale, a score of 4 or 5 is indicative of severe symptoms.

- Issues that are of importance in assigning specific ratings at the item level.
    - Item 6: Do not score hypersomnia as sleep disturbance.
    - Define guilt (item 9), morbid thought (item 12) and suicidal thinking (item 13) and make sure that the parent /child understand the concept before ascertaining the anchors.
    - Morbid thinking can include thoughts about one's own death, but without any desire to die, e.g. passive or active suicidal ideation.
    - Item 8: On the patient interview, observable irritability automatically generates a score of 5 or worse.
    - In contrast, do not rely on non-verbal report, e.g. "looking guilty," to rate guilt.
    - To interpret self-esteem, rely on probes to communicate the definition using a "to and fro" interview style, e.g. what do you, don't you like about yourself, your looks? Teasing and name calling (often reported by the parent, not the child) is often helpful to anchor a score of 7.

V. Guidelines for CGI Scores

A. Overview

The following guidelines link the CDRS-R score and the concepts of impairment and improvement to be ascertained on the CGI-S and CGI-I scores across three domains of functioning (home, school and peers). The CDRS-R score provides an estimation of deviance from normal that includes (implicitly and explicitly) the concept of associated impairment due to MDD symptoms. However, the concept of impairment allows for the possibility of adequate functioning, c.f. "holding it together at school," despite the presence of significant MDD symptoms. Based on the CDRS-R score and a clinically sensitive interview to ascertain the extent to which the subject experiences impairment due to MDD, the IE should generate a best-estimate CGI-S and CGI-I score, which should reflect the subject's current (over the past week) clinical status in light of the overall trajectory of change since baseline (T0), including considering the previous assessment's CGI-S and CDRS-R ratings.

Because TADS is a study of treatments for MDD and these are the primary outcome measures for MDD in the study, the IE should be especially careful to focus on only MDD when composing CGI scores. Other factors unrelated to MDD, such as other externalizing or internalizing disorders, family functioning, neighborhood problems, difficulties at school, should not be considered. For example, when rating the CGI-S for a patient with an IQ of 80 who struggles in school, the IE should consider impact of struggling in school on MDD and, in turn, the impact of MDD on school performance, but the impact of a low IQ per se should not be considered relative to age and gender matched peers who are a reference group for normal mood only.

Finally, it is important to remember that the CGI Score is a clinician rated measure. While rules for anchoring the CGI scores are provided, it is incumbent on the IE to assign a CGI score that he/she believes provides the best estimate of the patient's status relative to MDD and not to slavishly follow numerical anchors even if they diverge at the margins from the IE's judgment.

B. Guidelines for making CGI-Severity Ratings

These guidelines for scoring the CGI-S use CDRS-R ranges that were derived from the Emslie study of FLX versus PBO in pediatric MDD, from the TADS entry criteria with respect to CDRS-R and MDD inclusion criteria, and from the response metric inherent in the CDRS-R itself. **Note that the suggested CDRS-R anchors are not mandatory, but rather guidelines for the expected relationship between the CDRS and the CGI-S.** However, CGI-S scores usually should not be more than one point lower or higher than the suggested CDRS-R range.

Discounting the not assessed point, the CGI-S is a 7-point scale generally anchored as follows:

GLOBAL SEVERITY

0= Not assessed	4=Moderately ill
1= Normal, not mentally ill	5=Markedly ill
2= Borderline mentally ill	6= Severely ill
3= Mildly ill	7= Extremely ill

Specific Anchors:

1. Normal, not mentally ill.

    - CDRS-R Score generally below 25
    - The patient shows no impairment from MDD symptoms
    - Does not meet criteria for MDD

2. Borderline mentally ill.

    - CDRS-R Score around 30 (range 25-35)
    - There is little or no impairment from MDD symptoms
    - Does not meet criteria for MDD

3. Mildly ill

    - CDRS Score below around 40 (range 35-44)
    - There is impairment in one setting only (home, school, social)
    - Does not meet criteria for MDD

4. Moderately ill

    - CDRS-R Score around 50 (range $\geq$ 45 by entry criterion up to 55)
    - There is impairment in two settings (home, school, social)
    - Meets criteria for MDD

5. Markedly ill

    - CDRS-R Score around 60 (range 55-64)
    - There is impairment in two or more settings (home, school or social)
    - Unequivocally meets criteria for MDD

6. Severely ill

    - CDRS-R Score around 70 (range 65-74)
    - There is impairment in 3 settings (home, school or social)
    - Unequivocally meets criteria for MDD

7. Extremely ill

    - CDRS-R Score is greater than 75
    - There is impairment in 3 settings (home, school or social)
    - Unequivocally meets criteria for MDD

C. Guidelines for making CGI-Improvement Ratings.

GLOBAL IMPROVEMENT

0= Not assessed	4=No change
1= Very much improved	5=Slightly Worse
2= Much Improved	6= Much worse
3= Minimally improved	7= Very much worse

The following definitions of the CGI-I require a definition of what constitutes clinically significant improvement. **Unlike the CGI-S, which is ranked against a prototypic normal youth of the same age and gender, the CGI-I is always a judgment made in comparison to the patient's state at the start of the trial (baseline).** Thus**,** *clinically significant* improvement is said to occur when the patient's depressive symptoms move closer to normal functioning than to the patient's level at baseline.

Note that the IE must consider that a CGI-I score of 1 or 2 defines a responder; a CGI-Score of 3 defines a partial albeit minimal responder. CGI scores of 4 or worse define non-response, so it is critical for the IE to pay close attention to the guidelines when rating the CGI scores. When doing so, the IE should ask the following question: Compared to his/her condition for MDD only at randomization (T0) to the project, how much has he/she changed? Rate total improvement whether or not, in your judgment, it is due entirely to treatment.

1. Very Much Improved

    - All or most of the target symptoms are reported by the patient as improved to a significant degree; that is, the changes have a major and discernible impact on functioning, both subjectively and as observed by others
    - Symptomatic and functional status are such that the clinician would want to treat the patient with a maintenance treatment regimen
    - There is little or no room for further improvement, e.g. the patient has normalized relative to age and gender matched peers.
    - Does not meet criteria for MDD

2. Much Improved

- Several symptoms have improved to a degree that the patient feels an important change has occurred. The change is large enough that it is distinctly noticeable by other observers; some information may not fully confirm the changes.
- There is still some room for improvement, but if necessary, this dosage of treatment would still be considered very positive and perhaps sufficient.
- Does not meet criteria for MDD

3. Minimally Improved

- Only 1 or 2 symptoms (at most) show improvement, and the degree of change is not sufficient that one would consider not increasing the dose of treatment. Improvement may not be consistent across informants or domains of function.
- There is considerable room for improvement, clearly requiring further therapeutic adjustments.
- Probably meets criteria for MDD, but may be a "near miss."

4. No Change

- There are no symptoms that show improvement compared to the baseline state, or changes are so minimal as to be completely inconsequential
- The patient's condition by self-report as well as by independent observations clearly indicates lack of desired benefits
- Meets criteria for MDD

5. Slightly Worse

- There is some worsening over the baseline state in the target symptoms, though not to a degree that one would wish to completely stop treatment
- Although there may be positive changes, they are insufficient in comparison to the negative changes to consider staying at this dosage or with this treatment
- Meets criteria for MDD

6. Much Worse

- The patient's condition is worse enough that stopping or changing treatment should be considered, even though not all target symptoms are worse
- Worsened condition is clearly discernible by the patient and other observers
- Meets criteria for MDD

7. Very Much Worse

- Most of the target symptoms are worse; stopping, changing or adding new treatments is clearly indicated
- The patient may spontaneously stop or be unable to continue treatment due to worsened symptoms or more impaired functioning than at baseline
- Meets criteria for MDD

VI. HoNOSCA

The HoNOSCA is included in TADS to provide comparability to Richard Harrington's large UK treatment outcome study of adolescent depression. The HoNOSCA is a 15 item interview measure given jointly to the patient and parent. It covers all the major internalizing and externalizing symptoms and sets the stage for a CGAS rating. Some elements of the HoNOSCA, e.g. ADHD, overlap the K-SADS; others, e.g. school performance, do not. When the K-SADS interview is sufficient to score the HoNOSCA, these items do not need to be repeated.

VII. CGAS

The Clinical Global Assessment Scale (CGAS) is a measure of global functioning, not just functioning related to MDD. Hence, unlike the CGI and CDRS scores, which are focused on major depression, the IE should rate the CGAS relative to all areas of functioning factoring in impairment from all sources, including all mental disorders, cognitive capacity, family and peer problems, SES and stressors insofar as he or she is aware of these factors. In addition, the IE should rate the CGAS using the defined CGAS anchor points relative to age, gender matched peers, considering the baseline CGAS rating and change trajectory over the course of treatment.

VIII. Summary

This manual presents guidelines for the IE to use when rating the primary dependent measures. IE supervisor should be familiar with the IE manual and use it to supervise IE ratings according to quality assurance procedures specified in the QA manual.