Published in final edited form as:

J Am Acad Child Adolesc Psychiatry. 2007 September ; 46(9): 1204–1212.

A Psychometric Evaluation of the CDRS and MADRS in Assessing Depressive Symptoms in Children

Shailesh Jain, M.D., M.P.H.¹, Thomas Carmody, Ph.D.^{1,a}, Madhukar H. Trivedi, M.D.¹, Carroll Hughes, Ph.D.¹, Ira Bernstein, Ph.D.^{2,a}, David W. Morris, Ph.D.¹, Graham J. Emslie, M.D.¹, and A. John Rush, M.D.¹

¹ University of Texas Southwestern Medical Center, Dallas

² University of Texas at Arlington

^a Bio-statisticians

Abstract

Objective— This study compared the psychometric properties of the Children's Depression Rating Scale-Revised (CDRS-R) and the Montgomery-Asberg Depression Rating Scale (MADRS) in children with major depressive disorder (MDD).

Method— Children (n=96) (8–11 years inclusive) with nonpsychotic MDD were enrolled. Participants were part of a multisite, outpatient, randomized, placebo-controlled, 9-week trial of fluoxetine (10 mg/day for the first week and 20 mg/day thereafter). The CDRS-R and MADRS were completed based on clinician interviews with both parents and children. Classical Test Theory and Item Response Theory analyses were conducted.

Results— The MADRS and CDRS-R total scores were correlated at baseline ($r=0.51$) and at study exit (0.85). The Cronbach's alpha was 0.86 (CDRS-R) and 0.82 (MADRS) at exit. The effect sizes for change from baseline to exit between the fluoxetine and placebo groups were 0.78 (CDRS-R) and 0.61 (MADRS). Agreement between the CDRS-R and MADRS in the declaration of treatment response (50% improvement from baseline to exit) occurred in 84.2% of children. Test information function favored the CDRS-R.

Conclusion— The CDRS-R showed greater effect size for differentiating drug and placebo and better test information than the MADRS in this study of depressed children.

Keywords

Major depression; depressive symptoms; Child Depression Rating Scale-Revised; Montgomery-Asberg Depression Rating Scale

Introduction

Major depressive disorder in children has a point prevalence of about 2% (Weisz et al. 2006). Untreated and undiagnosed depression undermines social and school functioning, generates severe family stress, prompts significant use of mental health services (Clarke et al. 2003), and is linked to increased risk of other psychiatric disorders such as drug use and suicide (Gould et al. 1998).

The careful measurement of depressive symptoms can assist in establishing the severity of the disorder and response to treatment. The CDRS-R (Children's Depression Rating Scale – Revised) (Poznanski and Mokros 1996) is the most widely used instrument to assess depressive symptoms in child clinical trials. However, psychometric data on this scale are largely from non-clinical trial samples, with only limited psychometric data from clinical trials. Yet, the structure and complexity of the CDRS-R have restricted its use to clinical research. For example, the CDRS-R is multidimensional, with up to 5 factors within the scale (Guo et al., *in press*).

The MADRS (Montgomery Asberg Depression Rating Scale) (Montgomery and Åsberg 1979), an observer-rated depression scale, is widely used in depressed adults given its ease of administration and good psychometric properties. However, the MADRS is not commonly used in pediatric patients, and there are no psychometric data on the MADRS in pediatric patients. Furthermore, the MADRS has not been compared to the CDRS-R. As the MADRS is shorter, unidimensional, and may be easier to utilize in youth, comparing the psychometric properties of the two scales could allow future child clinical research to replace the CDRS-R with the MADRS. Such studies are lacking to date.

This report compares the CDRS-R and the MADRS psychometrically to address the following questions: 1. Are the CDRS-R and MADRS total scores correlated at baseline, at exit, and in terms of change from baseline to exit? 2. What are the item total correlations for each item on the CDRS-R and MADRS? 3. What are the within- and between-group effect sizes for CDRS-R and MADRS totals and items for a trial of fluoxetine versus placebo? 4. What are the conversion values between total scores on the CDRS-R and MADRS?

Methods

This report is based on a dataset generated in a previously reported 9-week acute trial (Emslie et al. 2002) of children and adolescents (n=219) with nonpsychotic major depressive disorder (MDD) (single or recurrent) who were randomized to fluoxetine or placebo. For this report, we focus on children (ages 8 to 11 inclusive) (n=96). The following summarizes the key features of the study.

Participants

Fifteen academic and private clinic investigators throughout the United States recruited pediatric outpatients with nonpsychotic MDD (single or recurrent CDRS-R>40, CGI≥4) (American Psychiatric Association 2000) from site patient populations, as well as by newspaper and radio advertising. The participants had normal cardiograms, hematological and blood chemistry studies, and urinalyses at baseline with no significant medical comorbidities. Participants with first-degree relatives with history of bipolar disorder were excluded, as were those who had used neuroleptic or monoamine oxidase inhibitors within two weeks prior to study entry (Emslie et al. 2002). Parents or legal guardians of all participants provided written informed consent prior to study entry. Institutional Review Board (IRB) approvals were acquired for each site.

Materials

The CDRS-R (Poznanski and Mokros 1996) is an observer-rated, 17-item instrument for pediatric patients (above 6 years). The scale encompasses cognitive, somatic, affective, and psychomotor symptoms. Items are rated for severity on a 7-point scale (1 to 7) for 14 items and on a 5-point scale (1 to 5) for three items. Anchors are provided for every other scale point for each item. A trained interviewer may take 30 to 40 minutes to complete the rating (raw score range 17–113).

The reliability, validity, and sensitivity to change of the CDRS-R are well documented (Brooks and Kutcher 2001). The scale has shown high internal consistency ($\alpha = 0.85$) in cross-sectional samples (Poznanski et al. 1985), and good interrater reliability ($r = 0.92$) (Poznanski et al. 1984), but poor inter-informant reliability (child-parent total score correlation was 0.38) (Mokros et al. 1987). Test-retest reliability indicated very good stability over a 2-week interval. Correlations between the CDRS-R scores and Global rating scale (0.87) or the 17-item Hamilton Rating Scale for Depression (HRSD₁₇) (Hamilton 1960; Hamilton 1967) (0.48) indicated moderate to good convergence. Item-total correlations ranged from 0.28 (impaired schoolwork) to 0.78 (depressed feelings) in the clinical sample (Poznanski and Mokros 2005). Discriminate validity of the CDRS-R comparing depressed and nondepressed children indicated that children diagnosed with depressive disorders (mean CDRS-R = 53.7; SD = 15.7) scored higher on the CDRS-R than the nondepressed group (mean CDRS-R = 34.12; SD = 8.4). While the CDRS-R has been used to assess change in depressive symptom severity (Emslie et al. 1997), some reports question its sensitivity to change in preadolescents (Stark et al. 1987). Additionally, the CDRS-R is somewhat long, generally taking 30–45 minutes to administer to both the child and parent (approximately 15–35 minutes per informant). Information from all informants is then synthesized to create a final clinician score for each item.

The MADRS is a 10-item clinician rating of depressive symptoms. Each item is scored on a 7-point scale (0 to 6) (range 0–60). Anchors are provided for even numbered scale points. Higher scores represent higher levels of depression. The administration and scoring of the MADRS takes a trained rater approximately 10 to 20 minutes, depending upon the severity of the symptoms and difficulty of the interview. Its psychometric properties have been studied extensively in adults (Mulder et al. 2003).

MADRS in adults has high interrater reliability, with estimates as high as 0.97 between a psychiatrist and general practitioner (Montgomery and Åsberg 1979). Correlation between each item and total score ranges between 0.12 (reduced appetite) and 0.84 (apparent sadness). The MADRS correlates with global ratings (0.70) and the HRSD₁₇ (0.59). The MADRS has established scores for varying levels of symptom severity. Scores of 44, 32, 23, 15, and 7 have been suggested to designate very severe, severe, moderate, mild, and recovered depressions, respectively (Kearns et al. 1982). The scale is unidimensional and not confounded by somatic or psychomotor symptoms (Carmody et al. 2006; Rocca et al. 2002). While the MADRS does not specify a time frame for rating the symptoms, it is usually used to assess the prior week. The MADRS generally takes about 10–15 minutes to administer to each informant. Psychometric properties of the MADRS in child or adolescent depression are lacking.

Procedures

Diagnoses were rendered after three diagnostic interviews (Diagnostic Interview for Children and Adolescents, DICA) (Herjanic and Reich 1982) were performed by three different qualified and experienced pediatric healthcare providers, including one psychiatrist. One diagnostic interview was conducted each week during a 2-week diagnostic screening period prior to starting study medications. The child was interviewed separately from the parent.

Treatment began with a one-week, single blind, placebo run-in. Participants with at least a 30% decrease in the CDRS-R total score in the run-in period were excluded from the study. Following medication initiation, participants returned for efficacy and adverse event assessments for six visits (at weeks 1, 2, 3, 5, 7, and 9). CDRS-R and MADRS (using both parent and child reports) were used at each visit to capture efficacy. The evaluators were trained in both the scales though vignette training was only done with the CDRS-R. The order of administration was not randomized. The same interviewer completed both ratings. The CDRS-R was the predefined primary outcome.

Statistical Methods

Classical test theory (CTT) measures of consistency — Cronbach's alpha (α) (Cronbach 1951) and item-total correlations (not corrected for item/total overlap) — were computed for the CDRS-R and MADRS at study exit. Effect sizes were computed for the total scale score and for each item for each measure. The within-group effect sizes were computed as the exit item (or total) score minus the baseline item (or total) score divided by the standard deviation of the change in item (or total) score. Let $D = X_{Base} - X_{Exit}$, where X_{Base} and X_{Exit} are baseline and exit item (or total) scores. The within-group effect size (ES) is then simply D/S_d , where D is the mean of D and S_d is its standard deviation. The between-group ES for an item or total score is $D_{Fluox} - D_{Plac}$ divided by its standard deviation.

The CDRS-R and MADRS were compared on their ability to identify response at exit. Response for this manuscript was defined a priori as a $\geq 50\%$ reduction in baseline total score for each measure. Kappa (κ) statistics were used to measure the strength of agreement between CDRS-R and MADRS classifications of response and nonresponse. McNemar's test was used to determine whether the MADRS had a greater likelihood of misclassifying patients as responders or nonresponders relative to the CDRS-R.

Item Response Theory (IRT) methods were used to equate CDRS-R and MADRS total scores using exit data. Samejima's graded IRT model (Samejima 1997) item parameters were estimated for each item on each measure. These parameters were used according to the procedure of Orlando et al. (Orlando et al. 2000) (and associated software) to generate an IRT score for each possible total score on the CDRS-R and MADRS. The IRT score, usually called theta, is a unitless measure of depression estimated from the IRT procedure commonly scaled to a mean of 0 and a standard deviation of 1. The total scores were equated by matching the IRT scores for the two corresponding scales (Orlando et al. 2000). When an IRT score did not match exactly, best judgment was used to equate the scales taking into account the matching of total scores immediately above and below the total score in question.

The graded IRT model was also used to compute the test information function (TIF) (Birnbbaum 1968) for each scale. The "test information" provided by a scale is defined to be the inverse of the standard error of the total score of the scale. Thus, a total score that provides a precise estimate of symptom severity contains more "information" than a total score that provides a less precise estimate. The TIF allows one to see at which levels of symptom severity the total score provides the most precise estimates and allows comparison of the precision of two or more measures across all levels of symptom severity.

An assumption of the IRT approach is that the measures assess only depression (i.e., are unidimensional, see the above IRT references). Therefore, a principal components factor analysis was conducted to assess the dimensionality of each measure at exit. Parallel analysis (Horn 1965; Humphreys and Ilgen 1969; Humphreys and Montanelli 1975; Montanelli and Humphreys 1976) was used to infer how many "real" factors (dimensions) were present in the data. This approach avoids some of the limitations of the traditional Kaiser-Guttman eigenvalue-greater-than-1 rule. Parallel analysis involves comparing the eigenvalues from a principal components analysis of the real data to eigenvalues that might be expected to arise by chance alone. To determine how large the latter eigenvalues might be, we generated a series of simulated datasets consisting of random numbers (where correlations between all variables are expected to be zero) using the same number of observations and items as the real data. Eigenvalues of the principal components for each simulated dataset are computed and averaged over replications. The number of principal components for which the real eigenvalues exceed the simulated eigenvalues defines the dimensionality.

Results

Table 1 shows the demographic features of the sample.

Total scores were correlated between MADRS and CDRS-R at both baseline ($r=0.51$) and at exit ($r=0.85$). The Cronbach's alpha at exit for CDRS-R was 0.86 and for MADRS was 0.82. The disattenuated correlation (correlation between hypothetical error-free versions of the test) at exit is thus 0.61 (Nunnally and Bernstein 1994). This is a relatively low correlation for two measures putatively measuring the same trait and implies that they differ by more than simple random error.

Table 2 shows the exit CDRS-R and MADRS item total correlations. The CDRS-R items that showed the highest correlations with the total score were 'depressed feeling' (0.79), 'capacity to have fun' (0.76), 'irritability' (0.75), 'self esteem' (0.71), and 'social withdrawal' (0.69). For the MADRS, those items that showed the highest correlations with the total scores were 'reported sadness' (0.80), 'apparent sadness' (0.76), 'pessimistic thoughts' (0.74), 'inability to feel' (0.72), and 'inner tension' (0.72).

Table 3 presents effect sizes for total scores and each item on CDRS-R and MADRS for patients treated with fluoxetine and placebo separately (within-group effect size for change). It also shows between-group effect sizes (drug vs. placebo effect size) for each measure and item.

The drug vs. placebo total score effect size was greater for the CDRS-R than for the MADRS (0.78 for CDRS-R versus 0.61 for MADRS). Among CDRS-R items, between group effect sizes were greatest for 'capacity to have fun' (0.58) and least for 'hypoactivity' (0.11). Between-group effect sizes on MADRS were greatest for 'reduced sleep' (0.63) and least for 'concentration' (0.10).

The CDRS-R response rate for all subjects was 49.5%, and the MADRS response rate 52.6%. The CDRS-R and MADRS agreed on the classification of response or nonresponse in 84.2% of participants. When there was disagreement between the CDRS-R and MADRS, the MADRS was neither more nor less likely than the CDRS-R to classify children as responders (6.3% were CDRS-R responders/MADRS nonresponder, and 9.5% were MADRS responders/CDRS-R nonresponders) (McNemar's test: $p=0.4386$). Kappa for response was 0.68, which indicates a "good" level of agreement (Landis and Koch 1977).

The CDRS-R was found to be multidimensional. Principal components factor analyses with parallel analysis identified three factors. The average of the first four eigenvalues from the simulated datasets were 1.81, 1.64, 1.50, and 1.38, which were compared to the first four real eigenvalues of 6.66, 1.88, 1.58, and 1.25. Three factors were identified because the first three real data eigenvalues were larger than the first three simulated data eigenvalues.

Table 4 shows the IRT parameters for the CDRS-R. Items with numerically higher slopes ("a" values) contribute more to the measurement of depression because these items are better able to discriminate among different levels of depression. The three most discriminating items were depressed feelings (2.82), capacity to have fun (2.55), and social withdrawal (2.29). Higher values of the "b" parameters indicate responses that occur less frequently (i.e., a higher level of depression severity is needed to respond). Table 4 shows that morbid ideation, suicidal ideation, and speech tempo responses were relatively less frequent.

Table 5 shows the MADRS IRT parameters. The three items relating most highly to depression (i.e., large "a" parameters) were reported sadness (2.87), pessimistic thoughts (2.62), and apparent sadness (2.59). Reduced appetite and suicidal thoughts were infrequently reported symptoms for the MADRS (i.e., large "b" parameters).

Figure 1 shows the test information functions (TIFs) for the CDRS-R and MADRS. In these figures, theta represents a unitless measure of depression estimated from the IRT procedure scaled to a mean of 0 and a standard deviation of 1. The TIF for the CDRS-R was almost twice that of the MADRS for thetas from -1 to +2 indicating that for average levels of depression, the CDRS-R was about twice as precise as the MADRS. For very high or low levels of depression, the differences between scales were not as great.

Table 6 shows the conversion between equivalent levels of symptom severity for the CDRS-R and MADRS. A CDRS-R of 28 corresponds to a MADRS of 4–5, while a CDRS-R of 40 corresponds to a MADRS of 11–13.

Discussion

We compared the performance and psychometric properties of the CDRS-R and the MADRS in a group of children (n=96) enrolled in a 9-week, randomized, double-blind, placebo-controlled trial (Emslie et al. 2002) for which CDRS-R was the primary outcome.

The MADRS and the CDRS-R showed acceptable internal consistency. High Cronbach's alpha for CDRS-R (0.86) and MADRS (0.82) show both the scales have individual response items constructed to measure severity of depression. While total score correlations were reasonable (0.51–0.85), the disattenuated correlation (0.61) indicates that the two scales are not measuring the same construct.

Item total correlations at exit depict some items on both the scales show stronger correlations to total score than others. Item total correlations exceeded 0.5 for all MADRS items save for appetite (9 out of 10 items). However, for the CDRS-R, 6 of the 17 items did not reach this criterion. Notably, suicidal thoughts were poorly correlated with total depression with CDRS-R but not with MADRS.

Effect sizes were high for CDRS-R total scores (1.53 for fluoxetine; 1.19 for placebo) and for MADRS total scores (1.07 for fluoxetine; 1.00 for placebo). Nearly all items on both scales had effect sizes ≥ 0.30 . As for effect size to differentiate drug and placebo, total scale score effect size was greater for the CDRS-R (0.56) than for the MADRS (0.35). Effect sizes for items that differentiated drug and placebo were low (< 0.3) for five CDRS-R items (appetite, guilt, hyperactivity, speech tempo, weeping) and three MADRS items (inner tension, reduced appetite, concentration). On the other hand, only 2 of 10 items on the MADRS had an effect size (drug vs. placebo) of ≥ 0.5 , while 7 of 17 items on the CDRS-R performed similarly.

Based on the small "b₀" parameters, depressed children are more likely to endorse CDRS-R items low self esteem, irritability, fatigue, schoolwork, social withdrawal, not able to experience fun in previously enjoyed activities, depressed affect, and depressed feelings. For MADRS depressed children are more likely to report lassitude, concentration, feelings of sadness, reported sadness, inner tension, pessimistic thoughts, and not able to experience strong feelings. Thus, clinically probing these symptoms rather than appetite or guilt is most useful in assessing the effects of treatment.

Additional evidence favors the CDRS-R. Despite its multifactorial nature, the CDRS-R had far better test information function than the MADRS; that is, the CDRS-R item obtaining more detailed (elaborate) information about depression than the MADRS.

Based on this sample, the generally accepted remission cut-offs for the CDRS-R and MADRS do not correspond. In pediatric trials, CDRS-R total scores ≥ 40 are generally considered commensurate with a depressive disorder. Scores ≤ 28 are considered in remission, albeit with some residual symptoms. In adult trials, a MADRS total score ≥ 32 is suggestive of a depressive

disorder, while ≤ 10 is considered remission criteria. In this sample of children, the CDRS-R remission score of 28 corresponded to a MADRS score of 4–5, substantially lower than the generally accepted score of 10 for remission. Furthermore, a CDRS-R score of 40, which is often an entry criteria in pediatric trials, corresponded with a MADRS score of 11–13, which is only slightly higher than adult remission criteria. Further comparison of CDRS-R and MADRS scores are needed to determine the best cut-off scores for these scales in children.

Limitations

Study limitations include the limited sample size. Results from procedures such as factor analysis and IRT, which involve estimating a large number of parameters, should be verified in larger samples. The generalizability of the results is limited since all subjects have to qualify for the randomized trial. In addition, the order of administration of the two instruments was not randomized. It is likely that the CDRS-R was often obtained first since it was the primary outcome. Greater attention and interview time was likely given to the CDRS-R than to the MADRS.

Clinical Implications

This is the first study to our knowledge to compare the psychometric properties of the CDRS-R and MADRS scales in children. The CDRS-R appears to be preferred in assessing symptom change in children. The MADRS, on the other hand, is more easily administered (less time and no specialized training) in clinical settings. Additional evaluation of the MADRS is needed in adolescents.

Acknowledgements

We wish to thank Eli Lilly Pharmaceuticals for providing the database for these analyses. These analyses were supported in part by the National Institute of Mental Health (NIMH), National Institutes of Health (MH-68851 to the University of Texas Southwestern Medical Center at Dallas, A. John Rush, M.D., PI, and by MH-68852 to the University of Texas at Arlington, Ira H. Bernstein, Ph.D., PI), and by T32-MH067543 for Dr. Jain.

References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4. Washington DC: American Psychiatric Press; 2000. Text Revision
- Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In: Lord, FM.; Novick, MR., editors. Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley; 1968.
- Brooks SJ, Kutcher S. Diagnosis and measurement of adolescent depression: a review of commonly utilized instruments. *J Child Adolesc Psychopharmacol* 2001;11(4):341–376. [PubMed: 11838819]
- Carmody TJ, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D, Woo A, Trivedi MH. The Montgomery Asberg and the Hamilton ratings of depression: A comparison of measures. *Eur Neuropsychopharmacol*. 2006;10.1016/j.euroneuro.2006.04.008
- Clarke, GN.; DeBar, LL.; Lewinsohn, PM. Cognitive-behavioral group treatment for adolescent depression. In: Kazdin, AE.; Weisz, JR., editors. Evidence-based Psychotherapies for Children and Adolescents. New York: Guilford Press; 2003. p. 120-134.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;6:297–334.
- Emslie GJ, Heiligenstein JH, Wagner KD, Hoog SL, Ernest DE, Brown E, Nilsson M, Jacobson JG. Fluoxetine for acute treatment of depression in children and adolescents: a placebo-controlled, randomized clinical trial. *J Am Acad Child Adolesc Psychiatry* 2002;41(10):1205–1215. [PubMed: 12364842]
- Emslie GJ, Rush AJ, Weinberg WA, Kowatch RA, Hughes CW, Carmody T, Rintelmann J. A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression. *Arch Gen Psychiatry* 1997;54(11):1031–1037. [PubMed: 9366660]

- Gould MS, King R, Greenwald S, Fisher P, Schwab-Stone M, Kramer R, Flisher AJ, Goodman S, Canino G, Shaffer D. Psychopathology associated with suicidal ideation and attempts among children and adolescents. *J Am Acad Child Adolesc Psychiatry* 1998;37(9):915–923. [PubMed: 9735611]
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56–62. [PubMed: 14399272]
- Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967;6(4):278–296. [PubMed: 6080235]
- Herjanic B, Reich W. Development of a structured psychiatric interview for children: agreement between child and parent on individual symptoms. *J Abnorm Child Psychol* 1982;10(3):307–324. [PubMed: 7175040]
- Horn JL. An empirical comparison of various methods for estimating common factor scores. *Educ Psychol Meas* 1965;25:313–322.
- Humphreys LG, Ilgen D. Note on a criterion for the number of common factors. *Educ Psychol Meas* 1969;29:571–578.
- Humphreys LG, Montanelli RG Jr. An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behav Res* 1975;10:193–206.
- Kearns NP, Cruickshank CA, McGuigan KJ, Riley SA, Shaw SP, Snaith RP. A comparison of depression rating scales. *Br J Psychiatry* 1982;141:45–49. [PubMed: 7116071]
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159–174. [PubMed: 843571]
- Mokros HB, Poznanski E, Grossman JA, Freeman LN. A comparison of child and parent ratings of depression for normal and clinically referred children. *J Child Psychol Psychiatry* 1987;28(4):613–624. [PubMed: 3654810]
- Montanelli RG Jr, Humphreys LG. Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: a Monte Carlo study. *Psychometrika* 1976;41:341–348.
- Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;134:382–389. [PubMed: 444788]
- Mulder RT, Joyce PR, Frampton C. Relationships among measures of treatment outcome in depressed patients. *J Affect Disord* 2003;76(1–3):127–135. [PubMed: 12943942]
- Nunnally, JC.; Bernstein, IH. *Psychometric Theory*. New York: McGraw-Hill; 1994.
- Orlando M, Sherbourne CD, Thissen D. Summed-score linking using item response theory: application to depression measurement. *Psychol Assess* 2000;12(3):354–359. [PubMed: 11021160]
- Poznanski E, Mokros HB, Grossman J, Freeman LN. Diagnostic criteria in childhood depression. *Am J Psychiatry* 1985;142(10):1168–1173. [PubMed: 4037128]
- Poznanski EO, Grossman JA, Buchsbaum Y, Banegas M, Freeman L, Gibbons R. Preliminary studies of the reliability and validity of the children's depression rating scale. *J Am Acad Child Psychiatry* 1984;23(2):191–197. [PubMed: 6715741]
- Poznanski, EO.; Mokros, HB. *Children's Depression Rating Scale, Revised (CDRS-R) Manual*. Los Angeles, CA: Western Psychological Services; 1996.
- Poznanski, EO.; Mokros, HB. Psychometric properties of the CDRS-R. In: Mokros, HB.; Poznanski, EO., editors. *Children's Depression Rating Scale, Revised (CDRS-R)*. Los Angeles, CA: Western Psychological Services; 2005. p. 52-53.
- Rocca P, Fonzo V, Ravizza L, Rocca G, Scotta M, Zanalda E, Bogetto F. A comparison of paroxetine and amisulpride in the treatment of dysthymic disorder. *J Affect Disord* 2002;70(3):313–317. [PubMed: 12128243]
- Samejima, F. Graded response model. In: van Linden, W.; Hambleton, RK., editors. *Handbook of Modern Item Response Theory*. New York: Springer-Verlag; 1997. p. 85-100.
- Stark KD, Reynolds WM, Kaslow NJ. A comparison of the relative efficacy of self-control therapy and a behavioral problem-solving therapy for depression in children. *J Abnorm Child Psychol* 1987;15:91–113. [PubMed: 3571741]
- Weisz JR, McCarty CA, Valeri SM. Effects of psychotherapy for depression in children and adolescents: a meta-analysis. *Psychol Bull* 2006;132(1):132–149. [PubMed: 16435960]

Figure 1. Test Information Functions for CDRS-R and MADRS for Children at Exit
CDRS-R= Child Depression Rating Scale-Revised; MADRS= Montgomery-Asberg Depression Rating Scale

Table 1

Demographic Characteristics (n=96)

Male (%)	57.3 (55/96)
Ethnicity (%)	
Black	7.3
Hispanic	3.1
Other	6.3
White	83.3
	Mean (standard deviation)
Age (years)	10.2 (1.0)
Age range (years)	8–11
Duration of current episode (weeks)	52.7 (69.8)
Age of onset (years)	8.1 (1.8)
CDRS-R (baseline)	52.9 (9.9)
CDRS-R (exit)	36.0 (12.1)
MADRS (baseline)	18.9 (7.6)
MADRS (exit)	10.5 (7.4)

CDRS-R = Child Depression Rating Scale-Revised; MADRS = Montgomery-Asberg Depression Rating Scale

Table 2

CDRS-R and MADRS Item Total Correlations (n=96) at Study Exit

CDRS-R Item	Item-total Correlation	p-value	MADRS Item	Item-total Correlation	p-value
Schoolwork	0.50	<.0001	Apparent sadness	0.76	<.0001
Capacity to have fun	0.76	<.0001	Reported sadness	0.80	<.0001
Social withdrawal	0.69	<.0001	Inner tension	0.72	<.0001
Sleep	0.39	<.0001	Reduced sleep	0.57	<.0001
Appetite or eating	0.37	0.0002	Reduced appetite	0.06	0.5845
Excessive fatigue	0.49	<.0001	Concentration diff	0.52	<.0001
Physical complaints	0.53	<.0001	Lassitude	0.58	<.0001
Irritability	0.75	<.0001	Inability to feel	0.72	<.0001
Guilt	0.56	<.0001	Pessimistic thoughts	0.74	<.0001
Self-esteem	0.71	<.0001	Suicidal thoughts	0.67	<.0001
Depressed feelings	0.79	<.0001			
Morbid ideation	0.30	0.0025			
Suicidal ideation	0.40	<.0001			
Weeping	0.64	<.0001			
Depressed affect	0.63	<.0001			
Tempo of speech	0.34	0.0006			
Hypoactivity	0.47	<.0001			

CDRS-R= Child Depression Rating Scale-Revised; MADRS= Montgomery-Asberg Depression Rating Scale

Table 3

Effect Size for Change from Baseline to Exit within the Fluoxetine and Placebo Groups and Between Fluoxetine and Placebo Effect Size for Change from Baseline to Exit

CDRS Item	Fluox	PBO	Effect Size	MADRS Item	Fluox	PBO	Effect Size
Total Score	1.51	0.94	0.78	Total Score	1.16	0.72	0.61
Schoolwork	0.82	0.42	0.51	Apparent Sadness	0.71	0.49	0.34
Capacity to have fun	1.24	0.72	0.58	Reported Sadness	0.94	0.67	0.46
Social withdrawal	1.11	0.41	0.56	Inner Tension	0.50	0.49	0.12
Sleep	1.03	0.63	0.50	Reduced Sleep	0.85	0.27	0.63
Appetite	0.51	0.39	0.14	Reduced Appetite	0.25	0.14	0.16
Fatigue	0.86	0.44	0.42	Concentration	0.47	0.42	0.10
Physical complaints	0.77	0.24	0.54	Lassitude	0.88	0.35	0.42
Irritability	1.11	0.43	0.54	Inability to Feel	0.90	0.39	0.53
Guilt	0.61	0.45	0.23	Pessimistic Thought	0.86	0.56	0.43
Self-esteem	1.34	0.79	0.49	Suicidal Thought	0.78	0.56	0.42
Depressed feelings	0.97	0.81	0.36				
Morbid ideation	0.67	0.29	0.55				
Suicidal ideation	0.54	0.19	0.47				
Weeping	0.64	0.33	0.27				
Depressed affect	0.90	0.54	0.45				
Speech tempo	0.64	0.36	0.20				
Hypoaactivity	0.60	0.49	0.11				

CDRS-R = Child Depression Rating Scale-Revised; fluox = fluoxetine; MADRS = Montgomery-Asberg Depression Rating Scale; PBO = placebo

Cohen defines a small effect size as 0.2–0.4, a median effect size as 0.4–0.6, and a large effect size as >0.6.

Table 4

Item Response Analyses for the CDRS-R

CDRS-R Item	a	b ₀	b ₁	b ₂	b ₃	b ₄	b ₅
1. Schoolwork	0.92	-0.85	0.35	1.76	2.59	4.08	5.31
2. Capacity for fun	2.55	-0.79	-0.04	0.63	1.07	2.11	2.59
3. Social withdrawal	2.29	-0.83	0.22	1.04	1.51	1.80	2.73
4. Sleep	0.55	-0.10	1.68	3.45	5.90		
5. Appetite	0.42	0.42	1.84	4.72	7.66		
6. Fatigue	0.89	-1.02	0.58	1.56	2.65	4.15	5.44
7. Physical complaint	0.92	0.11	1.13	2.16	2.90	4.10	11.31
8. Irritability	1.68	-1.34	-0.28	0.51	1.16	1.79	2.18
9. Guilt	1.56	0.45	1.16	2.74	3.49	7.80	*
10. Self-esteem	1.98	-1.47	-0.20	0.62	1.51	2.00	2.40
11. Depressed feelings	2.82	-0.59	0.41	1.13	1.46	2.06	4.56
12. Morbid ideation	0.64	1.72	3.91	6.20	7.29	14.47	*
13. Suicidal ideation	1.26	1.69	2.33	3.52	4.08	8.63	*
14. Weeping	1.39	0.08	0.76	1.19	1.57	1.90	2.44
15. Depressed affect	1.94	-0.64	0.71	1.55	2.21	2.99	6.14
16. Speech tempo	1.23	1.14	2.61	19.92	*	*	*
17. Hypoactivity	1.37	0.87	1.76	3.28	8.84		

* Insufficient data to estimate this parameter; CDRS-R= Child Depression Rating Scale-Revised

Table 5

Item Response Analyses for the MADRS

MADRS Item	a	b ₀	b ₁	b ₂	b ₃	b ₄	b ₅
1. Apparent sadness	2.59	-0.69	0.16	0.81	1.28	2.77	5.24
2. Reported sadness	2.87	-0.40	0.28	0.97	1.54	2.33	4.87
3. Inner tension	1.84	-0.28	0.55	1.51	2.36	3.27	7.17
4. Reduced sleep	0.93	0.07	0.88	1.97	2.92	5.34	11.54
5. Reduced appetite	0.29	4.63	6.10	10.61	19.87		*
6. Concentration	0.75	-0.93	-0.02	2.14	3.10	5.42	11.90
7. Lassitude	1.06	-1.04	0.73	2.17	3.40	10.45	*
8. Inability to feel	2.40	-0.02	0.63	1.12	1.98	2.82	5.21
9. pessimistic thoughts	2.62	-0.24	0.61	1.26	2.44	2.72	4.92
10. Suicidal thoughts	2.43	0.95	1.77	2.75	6.41	*	*

* Insufficient data to estimate this parameter; MADRS= Montgomery Asberg Depression Rating Scale

Table 6Conversion between CDRS-R and MADRS at Exit for Children^a

CDRS-R Total Score	MADRS Total Score (Child)	CDRS-R Total Score	MADRS Total Score (Child)
17–20	0–0	65–68	29–31
21–24	1–3	69–72	31–33
25–28	4–5	73–76	33–35
29–32	6–8	77–80	35–37
33–36	9–11	81–84	38–40
37–40	11–13	85–88	40–42
41–44	14–16	89–92	43–44
45–48	16–18	93–96	45–46
49–52	19–21	97–100	46–48
53–56	22–23	101–104	49–51
57–60	24–25	105–113	52–60
61–64	26–28		

^aDue to the length of the table, 4-point intervals have been used. The complete table is available from the authors.

CDRS-R= Child Depression Rating Scale-Revised; MADRS= Montgomery-Asberg Depression Rating Scale