

Childhood Depression: Prevalence Using DSM-III Criteria and Validity of Parent and Child Depression Scales

Daniel A. Lobovits and Paul J. Handal¹

Saint Louis University

Received August 4, 1983; revised November 23, 1984

DSM-III criteria for major affective disorder were used to establish a reliable and valid prevalence estimate for depression among latency-aged children. The prevalence data, based on separate mother/child interviews, were then compared to two additional paper-and-pencil depression scales: the depression scale of the Personality Inventory for Children (PIC-D); and the Children's Depression Inventory (CDI). Results indicated that DSM-III criteria yield a conservative yet valid estimate of childhood depression. Further evidence for the validity of both the PIC-D and CDI scales was obtained, as well as results which supported their use as gross-screening measures of depression.

KEY WORDS: childhood depression; personality inventory for children; children's depression inventory; DSM-III; depression.

Considerable controversy has arisen concerning the prevalence of childhood depression, with reported rates ranging from a low of 14% (Kovacs, Betof, Celebre, Mansheim, Petty, & Raynak, 1976) to a high of 62% (Brumback, Jackoway, & Weinberg, 1980). It appears that several factors may account for such widely varying prevalence rates. Such factors include the use of different diagnostic criteria; the absence of interrater reliability data for di-

¹All correspondence should be addressed to Paul J. Handal, Psychology Department, St. Louis University, 221 North Grand Boulevard, St. Louis, Missouri 63103.

agnosis; the inclusion of different age categories (adolescent and/or latency) and referral populations (inpatient and/or outpatient); and the failure to distinguish between information sources (parent, child, teacher) in establishing a diagnosis.

Diagnostic criteria used to assess depression among children have differed considerably across studies. For example, Ling, Oftedal, and Weinberg (1970) employed criteria originally used to diagnose manic-depressive illness in adults which were then modified according to "behaviors more readily observable in children." A positive diagnosis required a recent change in behavior, the absence of any other psychiatric illness, and the presence of at least 4 of 10 associated symptoms, including significant mood changes, social withdrawal, increasingly poor school performance, sleep disturbance, aggressive behavior not present before, self-deprecation and belief in persecution, lack of energy, somatic complaints, school phobia and weight loss/anorexia. A 40% prevalence rate was reported. Both Brumback et al. (1980), and Petti (1978) reported higher prevalence rates by refining Ling et al.'s (1970) criteria to include a primary symptom of dysphoria plus only two of eight associated symptoms. The investigators reported rates of 62% and 59%, respectively, using these more liberal criteria. Contrastingly, Carlson and Cantwell (1980) obtained a prevalence rate of 28% by employing the more stringent DSM-III criteria which require dysphoria plus four of eight associated symptoms. Apparently, prevalence rates vary according to the degree of stringency of different diagnostic criteria.

Past research is also noted for its lack of rigor in establishing diagnostic accuracy through the use of reliability procedures (Ling et al., 1980; Weinberg, Rutman, Sullivan, Penick, & Dietz, 1973; Petti, 1978; Carlson & Cantwell, 1980; Brumback et al., 1980). The failure to establish diagnostic reliability may well contribute to the existing discrepancies in prevalence data for childhood depression.

A third factor which may contribute to the inconsistency in prevalence data is that both age and type of referral population have varied from study to study. Some investigations included adolescents in their sample (Ling et al., 1970; Carlson & Cantwell, 1980) while others excluded postpubertal children (Weinberg et al., 1973; Brumback et al., 1980). Kovacs et al. (1976) and Petti (1978) sampled inpatients; Ling et al. (1970) and Weinberg et al. (1973) sampled outpatients, while Carlson and Cantwell (1980) sampled a mixed group of both in- and outpatients. These differing age and referral populations, coupled with varying diagnostic criteria, have made comparability across samples difficult at best.

A fourth problem is that previous studies have failed to elaborate as to how they have utilized various information sources in arriving at a diagnosis. Apparently, most investigators have combined information from a var-

iety of informants (i.e., child, parent, teacher) when diagnosing the child (Weinberg et al., 1973; Petti, 1978; Brumback et al., 1980; Carlson & Cantwell, 1980). Although discrepancies in reporting inevitably arise, there has been no discussion of how these discrepancies were resolved. As a result it is unclear whether such discrepancies may differentially effect reported prevalence rates, and if so, in what manner.

In summary, the lack of standardized criteria across studies has resulted in a great deal of inconsistency in the literature on childhood depression. Puig-Antich, Blau, Marx, Greenhill, and Chambers (1978) point out that the use of standardized criteria has the advantage of providing a common language to define depressive syndromes in children. In spite of the fact that controversy still exists regarding the etiology and symptomatic manifestations of childhood depression (Schulterbrandt & Raskin, 1977), Kashani, Husain, Shekim, Hodges, Cytryn, and McKnew (1981) have suggested that the most parsimonious approach would be to agree on an initial list of symptoms (i.e., DSM-III), study them, and modify them accordingly.

The purpose of this study was to use DSM-III (American Psychiatric Association, 1980) criteria to establish a reliable and valid prevalence estimate of depression among latency-aged children. A second purpose was to compare prevalence estimates for depression using DSM III as a criterion, with two additional paper-and-pencil depression scales: the depression scale of the Personality Inventory for Children (PIC-D) (Froman, 1971), and the Children's Depression Inventory (CDI) (Kovacs, 1978). The purpose of these comparisons was to evaluate the utility of these alternate measures in terms of their validity and classification accuracy when employed as mass-screening measures.

METHODS

Participants

Participants were fifty 8- to 12-year-old children (mean age = 10 years, 1 month) consecutively referred to two outpatient psychology service centers in St. Louis, Missouri. The sample included 37 boys and 13 girls referred for problems in academic performance and/or behavioral difficulties. The sample consisted predominantly of children from families of classes III and IV socioeconomic status (Hollingshead, 1957). Twenty-six percent were referred by hospital medical clinics, 26% by parents, 22% by school personnel, 12% by private physicians, and 14% by other community social service agencies. In all cases children were accompanied by their mothers who also served as participants.

In order to insure that depression was the primary disorder being diagnosed, the following children were excluded from the study: Children of subnormal intelligence (WISC-R < 80), children with any previous history of neurological impairment, children who demonstrated evidence of psychotic disturbance, and children who recently experienced the death of a significant other ($n = 8$). These exclusion criteria are consistent with the criteria elaborated in DSM III (APA, 1980, pp. 213-214). Thus, from an initial pool of 58 potential participants, 50 children and their parents participated in the study.

Measures

Diagnostic and Statistical Manual of Mental Disorders (DSM-III, APA, 1980) criteria for major depressive episode were used to assess depression. According to DSM-III criteria, a diagnosis of depression requires the primary symptom of dysphoria plus four of the eight associated symptoms listed. These symptoms must be present for at least 2 weeks.

The use of DSM-III criteria in diagnosing depression among adults and children has been established (APA, 1980). The use of these criteria in diagnosing depression in children has gained increasing support (Kashani & Simonds, 1979; Carlson & Cantwell, 1980). However, no inter-interviewer reliability has been reported using these criteria in diagnosing depression in latency-aged children. Therefore, establishing reliability for the diagnosis of depression was deemed necessary. To this end, each of two clinicians audiotaped five child and five parent interviews, yielding a total of 20 interviews (10 for each clinician). The clinicians then made a written record of their diagnoses. The audiotapes were subsequently exchanged for independent diagnosis by the alternate clinician. The diagnoses were then compared to assess the extent of inter-interviewer agreement. Reliability was computed on the basis of agreement/disagreement of the interviewers diagnoses of the presence or absence of depression in the child. Inter-interviewer reliability was computed separately for parent and child interviews.

The Personality Inventory for Children-Depression Scale (PIC-D) (Froman, 1971) is a rationally constructed instrument composed of 46 items. It is completed by the primary care-giver (usually mother) who responds to the items in a true-false manner. Factor-analytic studies of these items yielded two major clusters: brooding, moodiness, isolation which can roughly be compared to the major DSM-III symptom of dysphoria and, other face-valid factors, i.e., crying spells, lack of energy, pessimism, anhedonia, concern with death and separation, sensitivity to criticism, indecisiveness, and poor self-concept. Virtually all these additional factors are directly comparable to DSM-III associated symptoms of depression.

The reliability of the PIC-D was established in three studies that reported test-retest reliability estimates ranging from .80 to .94 (Wirt, Lachar, Klinedinst, & Seat, 1977). Initial support for its validity comes from administration of this scale to a criterion group of nine children who were independently diagnosed as having primary depression. Of the nine children, three obtained a D score 2 standard deviations (70T) above the normative sample mean, and three scored 3 standard deviations (80T) above. Scores for the remaining children were not reported (Wirt et al., 1977).

The Children's Depression Inventory (CDI) (Kovacs, 1978) is a 27-item multiple choice test designed to assess the severity of depression in school-aged children. The items cover an array of overt symptoms such as sadness, anhedonia, suicidal ideation, and sleep disturbance. There are at least two items per symptom that are directly comparable to DSM-III primary and associated symptoms of depression. Its internal consistency is respectable (coefficient $\alpha = .86$), and item-total correlations are all statistically significant (.31-.54). The CDI appears to be a reasonably stable index of symptoms assessed over a 1-month interval ($r = .72$) (Kovacs, 1980). The validity of the inventory is supported by Friedman and Butler's (1979) finding that high CDI scorers tend to have low self-esteem (CDI Piers-Harris correlation = .66) and by a significant correlation between CDI scores and clinicians global ratings of depression ($r = .55$) (Kovacs, 1980). Significant depression is established on the basis of a statistical cutoff score of 19.

Procedure

The experienced child psychology interns, one male, one female, conducted initial interviews to confirm the presenting problem with the parent. The children were then screened on the basis of the exclusion criteria elaborated previously. Parents of those children meeting these criteria were then asked to participate in the study. Consent was obtained and the appropriate forms were completed.

Separate diagnostic interviews were conducted with the parent and child. The parent was first queried regarding their child's general mood and behavior during the past 2 weeks. The interview proceeded until sufficient information was acquired to enable the clinician to rule out or confirm a DSM-III diagnosis of depression for the child based on the parent report. At the conclusion of the interview the parent completed the PIC-D. The child was subsequently interviewed according to the same format and criteria utilized in the parent interview. The CDI was then administered to the child. Each child read the first two items to the examiner to ensure adequate reading and comprehension levels. The test was administered orally to 3 of the

50 children whose ability, in the examiner's estimation, might threaten the validity of the test.

RESULTS

Independent diagnoses by the two clinicians judging the audiotaped interviews yielded an interrater agreement level of 90%. The 90% agreement rate was consistent for both parent and child interviews. Thus, the use of DSM-III criteria resulted in an acceptable level of diagnostic agreement when assessing depression in latency-aged children.

Prevalence rates were compared by gender of child and no significant differences were found. Interviews with boys resulted in a prevalence rate of 35% compared to 31% when girls were interviewed ($\chi^2(1) = 0.08$). Similarly parents were equally likely to rate their sons as depressed (22%) as they were their daughters (23%). This difference was not significant ($\chi^2(1) = 0.011$). Consequently, gender data were combined in all subsequent analyses.

Based on the parent interview, the prevalence estimate for depression in this 8-to 12-year-old outpatient sample referred for school performance and/or behavior problems was 22%. The prevalence rate based on child interview data was 34%. The difference between prevalence rates based on these two information sources approached significance (χ^2 for correlated proportions $(1) = 3.60, p < .06$).

The overall agreement rate between parent and child interviews was 78%. That is, interviews of parents and children yielded agreement on a diagnosis of depressed or nondepressed 78% of the time.

Next, the PIC-D scale was examined for two reasons: first, to determine whether there was statistical significance between PIC-D score means of the DSM-III depressed and nondepressed groups; second, to determine the classification accuracy of the PIC-D scale using DSM-III diagnosis as a criterion measure when suggested PIC-D cutoff scores were employed.

For the total sample, the mean PIC-D T score was 66.50 ($SD = 17.78$). A significant difference was found between the means of the depressed ($n = 11, M = 85.36, SD = 15.25$) and nondepressed ($n = 39, M = 61.18, SD = 14.64$) groups ($t(48) = 4.72, p < .0005$).

Classification accuracy was determined by using cutoff scores of 70T and 80T to identify children as depressed on the PIC-D scale, since preliminary validity data indicated that parents of depressed children rated their children 2 to 3 standard deviations higher than the normative sample mean. A cutoff score of 70T yielded an overall accuracy rate of 72%. That is, the PIC-D accurately classified 72% of the sample as either true depressed or true nondepressed using the DSM-III parent interview as a criterion meas-

ure. Of the 50 subjects who were misclassified, 14 (28%) included one false negative and 13 false positives (i.e., 91% of the depressed group were accurately classified). The large number of false positives primarily accounted for the modest degree of overall classification accuracy. The cutoff score of 70T resulted in the D scale rating of 46% of the total sample as depressed, compared to the prevalence rate of 22% obtained from the parent interview.

A cutoff score of 80T resulted in an overall classification accuracy rate of 90%. That is, 90% of the sample were accurately classified as either true depressed or true nondepressed. Five of the 50 subjects misclassified (10%) included three false negatives and two false positives. Thus, increasing the cutoff score improved the overall classification accuracy; however, the adjustment produced a less sensitive test by yielding more false negatives. A cutoff score of 80T yielded a prevalence rate of 22% which equals the prevalence rate obtained from the DSM-III parent interview.

Next the CDI was examined for significant differences between CDI mean scores for DSM-III depressed and nondepressed groups, as well as for its classification accuracy. The mean CDI score for the total sample was 10.68 ($SD = 7.07$). A significant difference was found between the means of depressed ($n = 17$, $M = 17.59$, $SD = 6.77$) and nondepressed ($n = 33$, $M = 7.12$, $SD = 4.12$) groups ($t(48) = 8.58$, $p < .0005$).

Using the tentative cutoff score of 19 suggested by Kovacs (1980) the CDI accurately classified 84% of the sample as either true depressed or true nondepressed. This cutoff yielded no false positives; however, it did produce nine false negatives, thereby correctly identifying only 47% of the depressed children. A cutoff of 19 yielded a prevalence rate of 16% compared to a rate of 34% obtained from the DSM-III child interviews.

A series of downward adjustments in cutoff scores were then performed to determine which cutoff would yield maximum classification accuracy. Using this stepdown method, maximum classification accuracy was obtained using a cutoff of 12 which correctly classified 88% of the total sample. Six of the 50 subjects misclassified (12%) included four false negatives and two false positives. This adjustment improved the overall accuracy (84% to 88%); it also improved the classification accuracy of the depressed groups from 47% to 76%. A cutoff of 12 also resulted in a prevalence rate of 30% compared to 34% obtained from the DSM-III child interview.

In one further attempt to improve the CDI's discriminative capabilities, the CDI was rescored by assigning a double weighting to the five items which tap the major symptom of dysphoria. It was reasoned that lending a greater degree of importance to this symptom, as is done with DSM-III criteria, could result in improved classification accuracy. The double-weighting procedure accurately classified 92% of the sample as either true depressed or true nondepressed when a cutoff of 12 was used. The four mis-

classified subjects (8%) included one false negative and 3 false positives. Thus, the double weighting of dysphoria items improved both the overall classification accuracy (88% to 92%) and the classification accuracy of the depressed group (76% to 94%) when compared to a cutoff of 12 without weighting.

DISCUSSION

The results of this study indicate that DSM-III criteria for major depressive episode offer a useful starting point in untangling the confusion surrounding the diagnosis and prevalence rate of childhood depression. Although earlier researchers have consistently ignored the issue of diagnostic reliability in their reporting of prevalence estimates, the present study's use of operational criteria produced an adequate degree of inter-interviewer reliability such that the prevalence rate reported can be considered an accurate estimate of depression in this population of children. The interrater agreement reported in this study is considerably higher than that reported in DSM-III (APA, 1980, p. 471). One explanation for the discrepancy is that the reliability reported in DSM-III was based upon agreement between professionals on the differential diagnosis of depression from the *entire* range of possible diagnostic categories. In the present study agreement was based on a more limited range of possibilities, i.e., presence or absence of depression.

Previous epidemiological studies have used widely differing diagnostic criteria in an attempt to establish childhood prevalence rates for this disorder. Not surprisingly, this practice has yielded widely discrepant and often conflicting estimates, even in comparable populations of children. For example, utilizing more liberal criteria, both Weinberg et al. (1973) and Petti (1978) reported prevalence estimates of 62% and 59%, respectively. In this investigation, the use of the more stringent DSM-III criteria resulted in prevalence rates of 22% and 34% (depending on informant) for the identical age and referral population. The prevalence rate increased to 50% when the current data were rescored according to Petti and Weinberg's et al. liberal criteria (dysphoria plus two of eight associated symptoms). This 50% prevalence rate was identical for both parent and child informants. An increase of this magnitude supports the observation by Petti (1978) and Carlson and Cantwell (1980) that use of less stringent diagnostic criteria fails to adequately discriminate between children who are truly depressed from those experiencing less severe and more transient forms of distress associated with the course of normal development. It appears that the use of DSM-III criteria results in a more conservative estimate of depression among children.

Separate interviewing of parents and children uncovered another source of variance in previously reported prevalence estimates. Notably, Carlson

and Cantwell (1980) found a rate of 28% when data from both parent and child interviews were presumably pooled to arrive at a diagnosis of depression in the child. This rate falls precisely between the 22% and 34% rates found in this investigation when parent and child data were reported separately. This distinction underscores the need for future epidemiological research to specify information source in reporting base rate data.

The lack of gender differences reported in this study is a finding inconsistent with epidemiological research in adult depression. Adult studies have, in general, found higher rates among men than among women (Weissman & Myers, 1978; Comstock & Helsing, 1976). Unfortunately, other studies of childhood depression do not report gender data. The results of this investigation suggest that gender is not a significant factor in the prevalence of depression among latency-aged children. However, this finding should be viewed as tentative, particularly because of the relatively small sample involved. Future longitudinal or cross-sectional research would provide useful data in confirming this finding and determining when gender differences begin to appear.

Additional support for the validity of the CDI and PIC-D scales was obtained when means of DSM-III depressed and nondepressed subjects were compared and found to be significantly different. Subsequent analyses revealed that the classification accuracy of these scales was somewhat limited. Midrange elevations (70-80T) on the PIC-D produced a large number of false positives. While use of 80T as a cutoff increased the overall classification accuracy by reducing the number of false positives, it also resulted in an increase in the number of false negatives. Use of either 70T or 80T as a cutoff represents a cost-benefit decision, i.e., better identification of true positives with 70T than 80T but lower overall classification accuracy with 70T due to the increased number of false positives. Results from the use of either cutoff score should be interpreted with caution. The weighting of dysphoria items on the CDI along with the use of 12 as a cutoff score demonstrated potential for improving its classification accuracy. In general, the results support additional research on the use of these scales as diagnostic and/or mass-screening devices.

REFERENCES

- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- Brumback, R. A., Jackoway, M. K., & Weinberg, W. A. (1980). Relation of intelligence to childhood depression in children referred to an educational diagnostic center. *Perceptual and Motor Skills*, 50, 11-17.

- Carlson, G. A., & Cantwell, D. P. (1980). A survey of depressive symptoms, syndrome and disorder in a child psychiatric population. *Journal of Child Psychology and Psychiatry*, 21, 19-25.
- Comstock, G. W., & Helsing, K. J. (1976). Symptoms of depression in two communities. *Psychological Medicine*, 6, 551-564.
- Friedman, R. J., & Butler, L. F. (1979). *Development and evaluation of a test battery to assess childhood depression*. Unpublished manuscript, Ontario Institute for Studies in Education.
- Froman, P. K. (1971). *The development of a depression scale for the Personality inventory for children (PIC)*. Unpublished manuscript, University of Minnesota.
- Hollingshead, A. B. (1957). *Two factor index of social position*. Unpublished manuscript, Yale University.
- Kashani, J. H., Husain, A., Shekim, W. O., Hodges, K. K., Cytryn, L., & McKnew, D. H. (1981). Current perspectives on childhood depression: An overview. *American Journal of Psychiatry*, 138 (2), 143-153.
- kashani, J., & Simonds, J. F. (1979). The incidence of depression in children. *American Journal of Psychiatry*, 136 (9), 1203-1205.
- Kovacs, M. (1978). *Children's depression inventory (CDI)*. Unpublished manuscript, University of Pittsburgh.
- Kovacs, M. (1980). Rating scales to assess depression in school aged children. *Acta Paedopsychiatrica*, 46, 305-316.
- Kovacs, M., Betof, N. G., Celebre, J. E., Mansheim, P. A., Petty, L. K., & Raynak, J. F. (1976). *Childhood depression*. Unpublished manuscript, Philadelphia.
- Ling, W., Oftedal, G., & Weinberg, W. (1970). Depressive illness in childhood presenting as severe headache. *American Journal of Diseases of Children*, 120, 122-124.
- Petti, T. A. (1978). Depression in hospitalized child psychiatry patients: approaches to measuring depression. *Journal of the American Academy of Child Psychiatry*, 17, 49-59.
- Puig-Antich, J., Blau, S., Marx, N., Greenhill, L., & Chambers, W. (1978). Prepubertal major depressive disorder: a pilot study. *Journal of the American Academy of Child Psychiatry*, 17, 695-707.
- Schulterbrandt, J. G., & Raskin, A. (1977). (Eds.). *Depression in childhood: diagnosis, treatment and conceptual models*. New York: Raven Press.
- Weinberg, W. A., Rutman, J., Sullivan, L., Penick, E. L., & Dietz, S. G. (1973). Depression in children referred to an educational diagnostic center: diagnosis and treatment. *Journal of Pediatrics*, 83, 1065-1072.
- Weissman, M.M., & Myers, J. K. (1978). Rates and risks of depressive symptoms in a United States urban community. *Acta Psychiatrica Scandinavica*, 57, 219-231.
- Wirt, R. D., Lachar, D., Klinedinst, J. K., & Seat, P. D. (1977). *Multidimensional description of child personality: A manual for the Personality Inventory for Children*. Los Angeles: Western Psychological Services.