Longitudinal Assessment of Major Depression and Anxiety Disorders in Children

MICHAEL FENDRICH, PH.D., MYRNA M. WEISSMAN, PH.D., AND VIRGINIA WARNER, M.P.H.

Abstract. Two-year recall and mother–child agreement with respect to a child’s DSM-III lifetime diagnoses of major depression and anxiety disorder, based on K-SADS-E interviews with children, were assessed for a sample of 59 children, 6 to 16 years of age, at high and low risk for depression. The mothers had excellent recall and the children had good recall of a child’s major depression. Both mothers and children had poor recall of a child’s anxiety disorder. Mother–child agreement on major depression in children improved at the 2-year follow-up. A comparison of diagnoses based on mother and child reports with the psychiatrists’ best estimate diagnoses of major depression suggested that children were more informative than mothers at the initial interview. The children were slightly less informative than the mothers at follow-up. These findings underscore the importance of multiple informants and longitudinal assessment in research on childhood psychiatric disorder. J. Am. Acad. Child Adolesc. Psychiatry, 1991, 30, 138–42. Key Words: interviews, recall, agreement, psychopathology.

With a growing interest in the epidemiology and developmental course of psychiatric disorders in children (Rutter, 1988), considerable attention is being paid to the assessment methods in this population (e.g., Young et al., 1987). The contribution of both parents and children as informants with respect to child psychopathology is widely acknowledged. A number of studies have compared the relative contribution of parents and children as informants in semistructured interviews by cross-sectionally evaluating the parent–child agreement (Reich et al., 1982; Kashani et al., 1985; Sylvester et al., 1987; Weissman et al., 1987; Ivens and Rehm, 1988). The present study adds to our knowledge concerning the relative contribution of parents and children as informants in two ways. First, by evaluating agreement longitudinally; second, by comparing parent and child reports to a child psychiatrist’s “best estimate” diagnosis.

In previous papers, the authors have reported on parent–child agreement on DSM-III diagnosis at the initial interview (Weissman et al., 1987) and on 2-year diagnostic recall among children (Fendrich et al., in press) in a sample of children at high and low risk for depression by virtue of their parent’s clinical status. In this paper, the authors report data on a subsample of children about whom two waves of diagnostic information were obtained over a 2-year interval from independent interviews with both the parents and the children, in order to address the following questions: Does parent–child agreement with respect to a child’s major depression or any anxiety disorder vary over 2 years? When a child psychiatrist reviews all available assessment information, who is considered more informative, the parent or the child? Does psychiatrist–informant agreement on these diagnoses vary over 2 years? The analysis in the current study is limited to diagnoses of depression and anxiety disorder because of the relatively low prevalence of other disorders (i.e., conduct disorder, attention deficit disorder, and substance abuse) within this subsample.

Method

Sample

Children were selected for the study by the presence or absence of a lifetime history of major depression (as defined by the Research Diagnostic Criteria [RDC]) (Spitzer et al., 1978) in their parents. A complete description of the probands (parents) and their assessment has been described elsewhere (Weissman et al., 1987). The depressed probands received treatment at the Yale University Depression Research Unit. The normal controls came from a 1975 community survey conducted in New Haven and had no history of psychiatric illness, based on at least four direct interviews (the last two using the Schedule for Affective Disorders and Schizophrenia—Lifetime (SADS-L)) (Endicott and Spitzer, 1978; and RDC criteria) over an 8-year period. All probands were white and group matched by age and sex. The complete sample of children consisted of 220 children, who were between 6 and 23 years of age at the time of the initial interview (Wave I), from 91 families. Diagnostic information from parents and/or children was available on all 220 children at Wave I.

Two years after the initial interview, all 91 families were contacted for a second interview (Wave II). Eighty-five (93%) of the 91 families with a total of 203 children consented to participate at Wave II. One hundred sixty of 220 eligible children (73%) provided direct interviews. In order
to reduce the interview burden and enhance compliance, interviews were requested with mothers about children who were to 6 to 18 years of age at Wave I; mothers provided interviews at Wave II for 69 out of 127 children (54%) in this age group. The mothers provided five additional interviews about children 19 years old; these interviews were excluded from the present analysis.

The authors' previous paper on 2-year diagnostic recall focused exclusively on information derived from interviews with children and was based on a subsample of 150 children, 6 to 23 years of age, who completed two waves of diagnostic interviews over a 2-year interval. The present study focuses on a subsample of 59 children about whom two waves of diagnostic information were obtained over a 2-year interval from independent interviews with both parents and children. It should be noted that 32 of these children and their parents were interviewed by the same person at both waves.

Children in this subsample ranged in age from 6 to 16 years, with a mean age of 11.7 years. Since older children were excluded, these 59 children were significantly younger than the other 161 children in the study (mean age 18.5 years). There were no sex or proband status differences between the two groups of children. Other comparisons suggested that the 59 children in this subsample were significantly less impaired than the other 161 children on the Children's Global Assessment Scale (Shaffer et al., 1983), a measure of social impairment based on a psychiatrist's evaluation of the interviews. They were also significantly less likely to have received a DSM-III diagnosis of conduct disorder or substance abuse. Multiple regression analyses suggested that differences in impairment and diagnoses disappeared once age was controlled for. Table 1 describes the 59 children included in the present study. Two complete sets of direct interviews were provided by 31 children from 19 depressed proband families and 28 children from 18 normal proband families. There were 32 boys and 27 girls. Children in the youngest age group (6 to 11 years) represented 39% of the sample; 61% of the sample was between 12 and 16 years of age.

### Assessment of Children

A modified version of the Schedule for Affective Disorders and Schizophrenia for School-Aged Children, Epidemiologic Version (K-SADS-E) (Orvaschel et al., 1982) formed the core of a comprehensive interview administered to the parent about the child and to the child about him/herself. Interviewers who were blind to the parents' diagnoses interviewed a parent (usually the mother) about the child and, at a later time, the child.

This study uses diagnoses derived from three sources: K-SADS-E interviews with children, K-SADS-E interviews with parents about children, and a psychiatrist's 'best estimate' diagnosis (Leckman et al., 1982). To derive best estimate diagnoses, a clinical psychologist and a child psychiatrist, who were not involved in the interviewing, independently reviewed all sources of information and assigned a lifetime DSM-III diagnosis. Discrepancies in diagnoses by the independent evaluators were resolved by a third source, who also independently and blindly reviewed all available information. In previous reports, it was shown that the best estimate procedure has excellent interrater reliability between psychiatrists (Weissman et al., 1988).

### Data Analysis

Coefficient kappa, an index of change corrected agreement, was calculated to estimate the stability of recall and diagnostic agreement (Fleiss, 1981). Coefficients were calculated separately for each major diagnostic category. The standards used for evaluating stability and agreement were as follows: Coefficients below 0.40 were considered poor; coefficients between 0.40 and 0.59 were considered fair; coefficients between 0.60 and 0.74 were considered good; and coefficients above 0.74 were considered excellent.

In the analysis of the stability of recall, the classification of agreement proceeded in the following way: A response was counted as an agreement on the presence of a diagnosis if answers on the K-SADS-E met DSM-III criteria for a particular diagnosis at both waves. A response was counted as an agreement on the absence of a diagnosis if the answers on the K-SADS-E did not meet the criteria for a particular diagnosis at both waves. In the absence of Wave I diagnoses, a report of a diagnosis at Wave II, dating the age of onset as occurring between Wave I and Wave II (i.e., new onsets between initial interview and follow-up), was treated as an agreement on the absence of a diagnosis. For comparing Wave I and Wave II diagnostic agreement between parents and children, new onsets between the initial interview and follow-up were treated as noncases.

### Results

#### Stability of Reports

Table 2 describes the 2-year recall for parents and children. The mothers showed excellent recall for major depression (K = 0.88) and poor recall for anxiety disorder (K = 0.25). Children showed good recall for major depression (K = 0.63) and poor recall for anxiety disorder (K = -0.07).

#### Parent-Child Agreement

Table 3 shows parent-child agreement with respect to the child's diagnosis at both Wave I and Wave II. In a previous
Table 2. Two-Year Recall of Child’s Lifetime—
DSM-III Diagnosis for Mothers and Children

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Major depression</td>
<td>0.88 (4, 5, 4)</td>
<td>0.63 (8, 4, 4)</td>
</tr>
<tr>
<td>Any anxiety disorder</td>
<td>0.25 (10, 3, 2)</td>
<td>−0.07 (19, 2, 0)</td>
</tr>
</tbody>
</table>

* W = total cases identified at Wave I; W2 = total cases identified at Wave II; W, W2 = total cases identified at both waves.

Table 3. Parent–Child Agreement at Wave I and Wave II

<table>
<thead>
<tr>
<th>DSM-III Diagnosis in Child</th>
<th>Wave I</th>
<th>Wave II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kappa (M, C, MC)² (N = 59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major depression</td>
<td>0.45 (4, 8, 3)</td>
<td>0.64 (5, 4, 3)</td>
</tr>
<tr>
<td>Any anxiety disorder</td>
<td>0.25 (10, 19, 6)</td>
<td>−0.04 (3, 2, 0)</td>
</tr>
</tbody>
</table>

* Wave II excluding new onsets.

² M = total cases identified by mother; C = total cases identified by child; MC = total cases identified by both mother and child.

report based on the full sample of children, including children ages 6 to 23 (Weissman et al., 1987), it was reported that the agreement at Wave I between parents and children was poor for most disorders and that agreement did not vary according to the age of the child. For the subsample of children included in this study, agreement at Wave I for major depression by DSM-III criteria was fair (K = 0.45). Improvement over Wave-I agreement was observed at Wave II for major depression (K = 0.64), but not for anxiety disorder (K = 0.04).

At Wave I (Table 4), the children showed better agreement with psychiatrists and higher rates of sensitivity than the mothers for both depression (K: 0.93 versus 0.57; sensitivity: 89% versus 44%) and anxiety disorder (K: 0.74 versus 0.49; sensitivity: 77% versus 45%). The findings from comparisons at Wave I are in contrast to those obtained when lifetime reports at Wave II are compared with the best estimate diagnoses at Wave II. The mothers showed slightly better agreement with the psychiatrists and slightly higher levels of sensitivity than the children for major depression (K: 0.61 versus 0.53; sensitivity: 50% versus 42%). Both mothers and children showed poor agreement with psychiatrists and low levels of sensitivity for anxiety disorder.

It should be stressed that at Wave II psychiatrists generated best estimate diagnoses based on information provided by informants at both Wave I and Wave II; best estimate diagnoses were cumulative. If both parents and children were perfectly reliable, the cumulative diagnosis generated by their reports would be the same as their Wave-II lifetime reports. As pointed out, however, both mothers and children have imperfect long-term recall. Thus, a final comparison on Table 4 uses the cumulative diagnoses of mothers and children. These cumulative reports were generated by combining reports made at both Wave I and Wave II. The cumulative reports of mothers and children, when compared with the psychiatrists' Wave-II best estimate, look very similar to the comparisons made at Wave I. Using cumulative reports of mothers and children, children showed better agreement with psychiatrists and higher levels of sensitivity than mothers for both major depression (K: 0.83 versus 0.61; sensitivity: 75% versus 50%) and any anxiety disorder (K: 0.64 versus 0.51; sensitivity: 75% versus 50%).

Discussion

Summary of the Findings

Two-year recall and mother–child agreement, with respect to DSM-III lifetime diagnoses of major depression and any anxiety disorder, was assessed in a sample of 59 children, 6 to 16 years of age, at high and low risk for depression. Mothers showed excellent recall for a child’s major depression; children showed good recall for major depression. Both mothers and children showed poor recall for any anxiety disorder. At the follow-up interview, mother–child agreement for major depression improved. Mother–child agreement with respect to any anxiety disorder remained poor at both interviews and actually worsened at follow-up. A comparison of diagnoses based on mother and child reports with the psychiatrist’s best estimate diagnosis suggested that children were more informative to the psychiatrist at the initial interview. Mothers were slightly more informative to psychiatrists at the follow-up interview 2 years later. Cumulatively, children were more informative to psychiatrists.

Limitations

Shrout et al. (1987) emphasized that the reliability coefficients obtained from one particular sample may not be generalizable to samples drawn from other populations. Although these findings have relevance to all samples of non-referred children, the characteristics of the informants and children in this study limit its generalizability. The sample of informants for the present study consists of children at high and low risk for depression and their parents. A replication of the present analyses in other samples of children is necessary before firm conclusions about the long-term reliability of assessment in this population can be drawn.

Generalizability is also limited by the focus on only two disorders and by the small sample size. The analyses were restricted to major depression and anxiety disorder because of the relatively low prevalence of other disorders in this subsample. It should be pointed out that consistent trends with respect to parent–child agreement, i.e., improved agreement at the second interview (not shown here), were observed for two other disorders, attention deficit disorder and conduct disorder.

It should be stressed that the psychiatrists did not directly interview the child; their best estimate diagnoses were partially based on reports provided by parents and children. If either parent or child reports on the K-SADS-E met criteria for a DSM-III diagnosis, psychiatrists almost always assigned the same best estimate diagnosis to the child. Two consequences of this lack of independence in the assignment of ratings are noted. The magnitude of Kappa (and sensitivity) for evaluating the psychiatrist’s agreement with a
particular source varies depending on the proportion of cases identified by each source. Additionally, since Kappa is partially based on the number of agreements with respect to caseness, the Kappas generated here will be inflated in magnitude compared with those generated under circumstances of complete rater independence. These two consequences suggest caution in the interpretation of statistics with respect to psychiatrist-informant agreement.

Test–Retest Reliability

In Table 5, the present findings are compared with the results from three other studies (Edelbrock et al., 1985; Hodges et al., 1987; Shaffer et al., unpublished manuscript) that evaluated the test–retest reliability of parents and children on DSM-III diagnoses or symptom scores derived from semistructured interviews. All three previous studies assessed the short-term (i.e., 1 to 3 weeks) reliability of current diagnoses or symptom scores in patient samples. Since this study assessed lifetime diagnoses, a 2-year retest interval is appropriate. Nevertheless, the longer interval between interviews makes the present study a relatively stringent test of diagnostic recall. Consistent with previous studies, the current study suggests that parents are more reliable than children as informants with respect to a child’s major depression. Unlike previous studies comparing parents and children, the present study found poor reliability for parents and children with respect to anxiety disorder.

It should be noted that in this study, all six reports of major depression made by parents at either Wave I or Wave II were from mothers who had a lifetime history of major depression. Recent findings (Breslau et al., 1988) suggest that reports on children provided by these mothers may be biased; depressed mothers may overreport depressive symptomatology in their children. This raises questions about the validity of the parental reports provided in this study. On the other hand, parent–child concordance with respect to the diagnoses of depression at Wave II was good. This supports the validity of the mother reports of child’s depression within this study.

Parent–Child Agreement

The present findings of improved parent–child agreement with respect to the child’s depression at follow-up stands in contrast with the general findings from most previous cross-sectional work on agreement (e.g., Reich et al., 1982; Kasani et al., 1985; Sylvester et al., 1987; Weissman et al., 1987). Two possible explanations are offered for the improved concordance shown in this study. The first explanation is that changes occur in the nature of the reports provided by children over time. Parents recalled all four of the cases of depression that they identified in the initial interview. On the other hand, only a subset of cases identified at the initial interview by children was recalled at the second interview (i.e., four out of eight). In the authors’ previous study focusing on children, it was suggested that children who recalled their depression were those with the most impaired social functioning and a history of psychiatric treatment at the initial interview (Fendrich et al., in press). As a consequence, it was suggested that long-term recall by children may be an indicator of diagnostic validity. Improved concordance between parents and children may be a consequence of the enhanced validity of reports of major depression provided by the children at follow-up.

Another variable affecting improved concordance may be the interview process itself. As parents and children are interviewed at multiple points in time, they may become more aware of the symptomatology that they were asked to report on in the initial interview. This awareness may lead to an increased accuracy in reporting and an improved concordance in symptom reports.

Implications for Assessment

Even though both parents and children showed at least “good” recall reliability for major depression, parents in this study were clearly more reliable than children. Nevertheless, if the study had been limited to parent informants, three of the 12 children (25%) identified as having a lifetime diagnosis of major depression would not have received the diagnosis. Thus, while the use of parental informants may enhance a study’s reliability, the use of child self-reports is essential for adequate sensitivity. Achenbach et al. (1987) suggested that parents and children report on different domains of symptomatic behavior in children. This study suggests that parents and children each make different contributions to the validity of the child assessment process. The weight of the evidence from both studies suggests that a
more comprehensive picture of childhood psychopathology will emerge when interviews with both parents and children are obtained.

References


