Two-Year Recall of Lifetime Diagnoses in Offspring at High and Low Risk for Major Depression

The Stability of Offspring Reports

Michael Fendrich, PhD; Myrna M. Weissman, PhD; Virginia Warner, MPH; Laura Mufson, PhD

• **Stability of recall of *DSM-III* diagnoses was assessed at two interviews 2 years apart in a sample of 150 offspring, aged 6 to 23 years, at high and low risk for major depression. Stability of recall was good for major depression with the use of *DSM-III* criteria and fair for major depression with the use of "strict" criteria (based on 4 weeks' duration of illness and an impairment in a major social role). Stability of recall was good for substance abuse and conduct disorder. Stability of recall was generally poor for anxiety disorder, regardless of subtype. For all major disorders except anxiety disorder, the difference in reported age at onset between the two interviews was small (<1 year) and not statistically significant. The most important correlates of stability of reports of major depression were previous psychiatric treatment and dysthymia and poor social functioning at the initial interview. This is the first study to evaluate long-term recall of *DSM-III* lifetime diagnoses in a nonreferred sample of children, adolescents, and young adults.**

(*Arch Gen Psychiatry*. 1990;47:1121-1127)

With a growing interest in the developmental course of psychopathology,[1,2] there is an increasing awareness of the need for valid and reliable procedures for assessing *DSM-III* psychiatric disorders in children, adolescents, and young adults.[3-7] Previous studies of the reliability of structured diagnostic interviews in this group have focused on the assessment of current diagnoses in patient samples.[5,6] The reliability of procedures for assessing the lifetime history of psychiatric disorders in both patient and nonreferred samples is of critical importance for epidemiologic research.

In a previous article, we reported on parent-child concordance with respect to the child's lifetime diagnosis in a sample of offspring at high and low risk for depression by virtue of their parents' clinical status.[4] This study reports on a 2-year follow-up of offspring in this same sample to evaluate the long-term stability (test-retest reliability) of reports of *DSM-III* lifetime diagnoses. Using information from interviews with offspring, this study addressed several questions about stability of diagnoses. How stable are child reports of their lifetime *DSM-III* diagnoses over a 2-year period? Are the child's age, sex, parental diagnostic status, or age at onset of disorder associated with higher (or lower) levels of stability? Are characteristics associated with diagnoses (eg, impairment in social functioning) or life events surrounding the onset of diagnoses (eg, parental divorce) related to diagnostic recall?

Accepted for publication November 21, 1989.

From the Departments of Psychiatry (Drs Fendrich, Weissman, and Mufson) and Public Health (Dr Weissman), College of Physicians and Surgeons of Columbia University, and New York State Psychiatric Institute (Drs Fendrich, Weissman, and Mufson and Ms Warner), New York, NY.

Reprint requests to College of Physicians and Surgeons of Columbia University, 722 W 168th St, Box 14, New York, NY 10032 (Dr Weissman).

SUBJECTS AND METHODS

Sample

Offspring were selected for the study by virtue of the presence or absence of a lifetime history of major depression (as defined by the Research Diagnostic Criteria) in their parents. A complete description of the probands (parents) and their assessment has been published elsewhere.[4] The depressed probands had received treatment at the Yale University Depression Research Unit (New Haven, Conn). 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 Schedule for Affective Disorders and Schizophrenia–Lifetime criteria and Research Diagnostic Criteria) during an 8-year period. All probands were white and group matched by age and sex. The complete sample of offspring consisted of 220 offspring from 91 families who were between the ages of 6 and 23 years at the time of the initial interview (wave 1). Diagnostic information from parents and/or offspring is available for all 220 offspring at wave 1.

Two years after the initial interview, all 91 families were contacted for a second interview (wave 2). Eighty-five (93%) of the 91 families with a total of 203 offspring consented to participate at wave 2. Diagnostic information was obtained on all 203 offspring. Of these offspring, 160 were directly interviewed. In a separate study, we examined a subsample of 59 children with two waves of direct interviews with mothers and children.[8] This article focuses on offspring who were directly interviewed at two waves. Since 10 of the 160 offspring interviewed at wave 2 were not interviewed at wave 1, the sample for the present study consisted of 150 offspring.

Comparisons between these 150 offspring and the other 70 offspring who did not provide two direct interviews found no significant differences in age, sex, proband status, social class, overall social functioning (as measured by the Child's Global Assessment Scale [CGAS][9]), or child's diagnosis, with one exception. The 150 offspring were more likely to have received "any diagnosis" (see below) at wave 1 than were the other 70 offspring.

Table 1 describes the 150 offspring who provided reports at both waves. Interviews at both waves were provided by 81 offspring from 41 depressed proband families and by 69 offspring from 30 normal proband families. There were 67 boys and 83 girls. The overall age distribution was roughly similar across proband groups. Offspring in the youngest age group (6 to 11 years) represented 17% of the sample; about 37% of the sample was between 12 and 17 years old at wave 1; 46% of the sample was 18 years old or older.

Assessment of Offspring

A modified version of the Schedule for Affective Disorders and Schizophrenia for School-Aged Offspring, Epidemiologic Version (K-SADS-E[7]), formed the core of a comprehensive interview administered to the parent about the child and to the child about himself or 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. The analyses for the present study are based on diagnoses derived from child interviews only. To facilitate cooperation, an attempt was made to use the same interviewer for each child at both wave 1 and wave 2. Sixty-two (41%) of the offspring were

interviewed by the same person at both waves (the effects of using the same or a different interviewer are examined in this article). The battery of measures administered at wave 1 included an assessment of each child's intelligence by means of the Peabody Picture Vocabulary Test[10] and of social functioning by means of the CGAS.[9]

The modified K-SADS-E used in this study can generate most of the major *DSM-III* Axis I psychiatric disorders. In addition to *DSM-III* diagnoses of major depression, anxiety disorder, attention deficit disorder, conduct disorder, and substance abuse disorder, we employed two other diagnostic outcomes, including major depression "strict criteria" and "any diagnosis." A diagnosis based on "strict criteria" was used in a previous report that used this sample[4] and is defined as a lifetime episode of major depression with a duration of at least 4 weeks and an impairment in a major social role. A child who reported a lifetime history of any of the major *DSM-III* psychiatric disorders was classified as having "any diagnosis." Diagnoses were assigned only if a subject met "definite" diagnostic criteria.

The interviewers, all of whom had a minimum of 5 years' clinical experience with children, included two child psychologists with PhD degrees, two child psychiatry fellows with MD degrees, and two masters-level school psychologists. The interviewers all went through approximately 30 hours of training in the research assessments, during which interrater reliability was checked with the field supervisor by co-rating interviews before the study began. Interrater reliability was monitored throughout the study with the field supervisor used as the standard.

Table 1.—Demographic Characteristics of Offspring With Two Waves of Direct Interviews

Wave 1 Age of Offspring, y	Parent Proband Status, No. (%)		
	Depressed	Normal	Total
Boys			
6-11	11 (28.2)	4 (14.3)	15
12-17	9 (23.1)	15 (53.5)	24
18-23	19 (48.7)	9 (32.1)	28
Total	**39 (100.0)**	**28 (100.0)**	**67**
Girls			
6-11	6 (14.3)	5 (12.2)	11
12-17	15 (35.7)	16 (39.0)	31
18-23	21 (50.0)	20 (48.8)	41
Total	**42 (100.0)**	**41 (100.0)**	**83**
Total Offspring	**81**	**69**	**150**
Total families	41	30	71

Other Measures

Parents and offspring were asked to complete a battery of self-report measures at wave 1 and wave 2. Some of these measures were used to construct family risk factors described in detail elsewhere by Fendrich et al.[11] Parent reports on the Short Marital Adjustment Test (SMAT)[12] were used to construct an index of poor marital adjustment. Mothers' reports on a family life events questionnaire were used to construct an index of parent-child discord. Parent reports of marital history were used to construct an index of family divorce. Child reports on the cohesion subscale of the Family Adaptability and Cohesion Evaluation Scale (FACES)[13] and on the caring and overprotection subscales of the Parental Bonding Index[14] were used to construct an index of "low cohesion" and "affectionless control." Indexes of current depressive symptoms were derived from child reports on a child's version of the Center for Epidemiological Studies Depression Scale (CES-D).[15]

Data Analysis

Coefficient κ, an index of chance corrected agreement, was calculated to estimate the stability of recall for offspring.[16] Coefficients were calculated separately for each major diagnostic category and for demographic subgroups. Standards used for evaluating diagnostic stability were as follows: coefficients below 0.40 were considered poor stability; between 0.40 and 0.59, fair stability; between 0.60 and 0.74, good stability; and 0.75 and above, excellent stability.

In the present study, a response was counted as an agreement on the presence of a diagnosis if the 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 answers on the K-SADS-E did not meet criteria for a particular diagnosis at both waves. In the absence of wave 1 diagnoses, reports of diagnoses at wave 2 dating age at onset as occurring between wave 1 and wave 2 (ie, new onsets between follow-up) were treated as agreements on the absence of diagnosis. Intraclass correlation coefficients[17] were calculated to evaluate the consistency of reports of age at onset for diagnoses. Cross-tabulations and logistic regressions were used to investigate correlates of consistent reports of major depression.

RESULTS

Stability of Child's Recall

Table 2 assesses the stability of offspring's recall of lifetime diagnoses. On inspection of the magnitude of the coefficients for all 150 offspring combined, recall was good for major depression (*DSM-III*

Table 2.—Recall of Lifetime Diagnoses Over Two Years by Offsp

DSM-III Diagnosis in Offspring	Parent Proband Group		
	Overall	Depressed	Normal
MDD (*DSM-III*)	0.61 (42, 27, 24)†	0.65 (28, 20, 18)†	0.50 (14, 7, 6)†
MDD (strict criteria)	0.46 (29, 22, 14)†	0.50 (21, 15, 11)†	0.33 (8, 7, 3)†
Any anxiety disorder	0.23 (49, 20, 13)†	0.27 (29, 13, 9)†	0.17 (20, 7, 4)
Phobia	0.24 (23, 14, 6)†	0.43 (11, 9, 5)†	0.02 (12, 5, 1)
Separation anxiety	0.17 (25, 4, 3)†	0.24 (14, 2, 1)†	0.27 (11, 2, 2)†
Panic disorder	0.39 (3, 2, 1)†	0.49 (2, 2, 1)†	. . .
Obsessive-compulsive disorder	0.28 (6, 1, 1)†	0.39 (4, 1, 1)†	0.00 (2, 0, 0)
Overanxious disorder	0.14 (8, 4, 1)	0.18 (5, 4, 1)	0.00 (3, 0, 0)
Attention deficit disorder	0.38 (4, 6, 2)†	0.49 (2, 2, 1)†	0.31 (4, 2, 1)†
Conduct disorder	0.76 (28, 36, 26)†	0.78 (19, 24, 18)†	0.72 (9, 12, 8)†
Substance abuse	0.63 (16, 11, 9)†	0.71 (13, 11, 9)†	0.00 (3, 0, 0)
Any diagnosis	0.54 (94, 82, 71)†	0.50 (56, 47, 42)†	0.56 (38, 35, 29)†
No. of offspring	150	81	69
No. of families	71	41	30

*MDD indicates major depressive disorder. Numbers in parentheses represent wave 1 reports, wave 2 reports, overlap.
†Coefficient significantly different from 0 with $P<.01$.
‡Coefficient significantly different from 0 with $P<.05$.

criteria only; 0.61) and substance abuse (0.63) and excellent for conduct disorder (0.76). The coefficients for major depression by strict criteria (0.46) and any diagnosis (0.54) suggested fair recall. With the possible exception of panic disorder (0.39), recall of anxiety disorders was very poor, regardless of subtype. Recall of attention deficit disorder was also poor (0.38).

Offspring's Recall by Parent Proband Group

Table 2 suggests that offspring of depressed probands showed consistently better recall than did offspring of normal subjects. Offspring of depressed probands demonstrated considerably better recall of major depression (*DSM-III* criteria, 0.65 vs 0.50; strict criteria, 0.50 vs 0.33), any anxiety disorder (0.27 vs 0.17), any phobia (0.43 vs 0.02), obsessive-compulsive disorder (0.39 vs 0.00), attention deficit disorder (0.49 vs 0.31), and substance abuse (0.71 vs 0.00). Both groups of offspring showed fair recall for any diagnosis (offspring of depressed probands, 0.50; offspring of normal probands, 0.56).

Child's Recall by Sex of Child

The coefficients suggested a trend toward better recall of diagnoses for boys than for girls for all disorders except anxiety. Boys' recall of *DSM-III* major depression was excellent (0.76) and considerably better than girls' recall (0.53). Boys also demonstrated better recall for major depression than girls when strict criteria were used (0.58 vs. 0.40). Boys demonstrated considerably better recall of attention deficit disorder than girls did (0.47 vs −0.01).

Child's Recall by Age of Child

The coefficients for diagnostic recall were further examined in two separate age groups (Table 2). The first group consisted of 81 offspring who were 6 to 17 years old at the time of the initial interview. The second group consisted of 69 offspring who were 18 to 23 years old at the time of the initial interview. The coefficients for the younger group of offspring suggested a trend toward slightly better recall of major depression by *DSM-III* or strict criteria (0.63 vs 0.54 and 0.49 vs 0.40, respectively) and considerably better recall of attention deficit disorder (0.79 vs −0.04) than for the older group of offspring. The older group of offspring demonstrated considerably better recall of anxiety disorder (0.41 vs 0.05), substance abuse (0.66 vs 0.00), and any diagnosis (0.62 vs 0.44).

Additional analyses (not shown) suggested that sex differences in recall of major depression (*DSM-III* criteria) were more pronounced in the offspring 18 years old and older. In the older age group, boys showed considerably better stability of recall than girls for major depression by means of both *DSM-III* (0.79 vs 0.41) and strict criteria (0.61 vs 0.29). In the younger age group, boys demonstrated only slightly better stability of recall than girls for major depression by means of both *DSM-III* (0.72 vs 0.56) and strict criteria (0.53 vs 0.46).

Stability of Diagnosis by Same or Different Interviewer

Table 3 compares κ coefficients obtained for the 62 offspring interviewed by the same rater at both waves 1 and 2 with the coefficients obtained for the 88 offspring interviewed by different raters. Overall, offspring interviewed by the same rater tended to show better diagnostic treatment than offspring interviewed by different raters. Offspring showed considerably better recall of major depression (strict criteria only), attention deficit disorder, and any diagnosis when they were interviewed by the same person at both waves than when they were interviewed by the different interviewers (0.60 vs 0.36, 0.65 vs −0.02, and 0.64 vs 0.46). The coefficients from the overall sample (second column, Table 2) are much closer in magnitude to the coefficients for offspring rated by different interviewers (third column, Table 3) than they are to the coefficients for offspring rated by the same interviewer (second column, Table 3).

Stability of Reports on Age at Onset

The fifth column of Table 4 lists the intraclass correlation coefficients for age at onset of each disorder for offspring reporting definite diagnoses at both wave 1 and wave 2. With a small number of offspring reporting diagnoses of substance abuse (n = 9) and anxiety disorder (n = 13) at both waves, age at onset statistics for these disorders should be interpreted with caution. The intraclass correlation coefficient for conduct disorder was good (.73). The coefficients for major depression (.58) and any anxiety were both fair (.53). The coefficients for substance abuse (.39) and any diagnosis (.33) suggested poor reliability.

The second and third columns of Table 4 indicate the mean age at onset reported for each diagnosis at wave 1 and wave 2. For all of the diagnoses, there was a tendency for mean age at onset to be higher at the second interview than at the first. With the use of paired *t* tests, only two mean differences, anxiety disorder (3.2 years) and any diagnosis (2.6 years), were significantly different from 0. The mean paired differences in reported age at onset for major depression (0.83 year), conduct disorder (0.77 year), and substance abuse (0.56 year) were all small and not significantly different from 0.

Factors Associated With Consistent Reports of Major Depression

Parental diagnosis, family risk factors, and diagnostic characteristics were explored as correlates of reporting consistency for major

κ by Parents' Clinical Status and by Sex and Age of Offspring)*

Sex of Offspring		Age of Offspring, y	
M	F	≤17	18+
0.76 (10, 10, 8)†	0.53 (32, 17, 16)†	0.63 (12, 6, 6)†	0.54 (30, 21, 18)†
0.58 (7, 9, 5)†	0.40 (22, 13, 9)†	0.49 (10, 5, 4)†	0.40 (19, 17, 10)†
0.26 (18, 8, 5)‡	0.21 (31, 12, 8)‡	0.05 (26, 7, 3)	0.41 (23, 13, 10)†
0.27 (5, 7, 2)‡	0.23 (18, 7, 4)‡	0.06 (8, 6, 1)	0.33 (15, 8, 5)†
0.16 (10, 1, 1)‡	0.17 (13, 3, 2)‡	0.10 (15, 1, 1)‡	0.26 (10, 3, 2)
. . .	0.49 (2, 2, 1)†	0.00 (2, 0, 0)	0.66 (1, 2, 1)†
0.00 (3, 0, 0)	0.49 (3, 1, 1)†	0.00 (4, 0, 0)	0.66 (2, 1, 1)†
−0.02 (2, 1, 0)	0.18 (6, 3, 1)	0.00 (4, 0, 0)	0.20 (4, 4, 1)
0.47 (5, 3, 2)†	−0.01 (1, 1, 0)	0.79 (3, 2, 2)†	−0.04 (3, 2, 0)
0.75 (15, 17, 3)†	0.77 (13, 19, 13)†	0.80 (7, 10, 7)†	0.71 (21, 26, 19)†
0.63 (9, 6, 5)†	0.64 (7, 5, 4)†	0.00 (2, 0, 0)	0.66 (14, 11, 9)†
0.52 (39, 35, 29)†	0.55 (55, 47, 42)†	0.44 (43, 36, 28)†	0.62 (51, 46, 43)†
67	83	81	69
28	43	26	45

Table 3.—Recall of Lifetime Diagnoses in Offspring by Interviewer Status at Wave 2 (κ Values)*

	Wave 2 Interviewer	
	Same as Wave 1 Interviewer	Different From Wave 1 Interviewer
MDD (*DMS-III*)	0.66 (19, 11, 11)†	0.58 (23, 16, 13)†
MDD (strict criteria)	0.60 (12, 9, 7)†	0.36 (17, 13, 7)†
Any anxiety disorder	0.26 (22, 8, 6)‡	0.21 (27, 12, 7)‡
Attention deficit disorder	0.65 (3, 3, 2)†	−0.02 (3, 1, 0)
Conduct disorder	0.84 (13, 17, 13)†	0.71 (15, 19, 13)†
Substance abuse	0.63 (7, 5, 4)†	0.64 (9, 6, 5)†
Any diagnosis	0.64 (39, 32, 30)†	0.46 (55, 50, 41)†
No. of offspring	62	88

*MDD indicates major depressive disorder. Numbers in parentheses represent wave 1 reports, wave 2 reports, overlap.
†Coefficient significantly different from 0 with $P<.01$.
‡Coefficient significantly different from 0 with $P<.05$.

Table 4.—Age at Onset Reports in Offspring: Mean Scores and Reliability (Intraclass Correlation Coefficient [ICC]) for Offspring Reporting Diagnoses at Both Waves

Lifetime *DSM-III* Diagnosis in Offspring	Age at Onset, y				
	Wave 1 (Mean ± SD)	Wave 2 (Mean ± SD)	Difference* (Mean ± SE)	ICC	N†
Major Depression	15.3 ± 0.81	16.1 ± 0.67	0.83 ± 0.68	.58	24
Any anxiety disorder	7.5 ± 1.84	10.7 ± 1.44	3.23 ± 1.43‡	.53	13
Attention deficit§	. . .	. . .	. . .	. . .	. . .
Any substance abuse	16.1 ± 0.73	16.7 ± 0.94	0.56 ± 0.96	.39	9
Conduct disorder	13.3 ± 0.64	14.1 ± 0.53	0.77 ± 0.42	.73	26
Any diagnosis	9.1 ± 0.73	11.7 ± 0.68	2.56 ± 0.79‖	.33	71

*Wave 2 ages at onset minus wave 1 ages at onset for paired reports.
†Number reporting diagnosis at wave 1 and wave 2.
‡$P<.05$ (using paired t test).
§Unable to calculate since there were only two diagnoses at both wave 1 and wave 2.
‖$P<.01$ (using paired t test).

Table 5.—Relationship Between Consistency* in Offspring Reports of Lifetime Major Depression Over Two Years by Demographic, Psychosocial, Family Risk, and Diagnostic Variables†

Major Depression, Offspring Report (N = 45)	Consistency, No. (%)	P
Demographic characteristics		
Social class		
Upper-middle (1-3)	10/23 (43.5)	NS
Lower (4, 5)	14/22 (63.6)	
Sex of offspring		
M	8/12 (66.7)	NS
F	16/33 (48.5)	
Age at wave 1, y		
6-11	0/1 (00.0)	NS
12-17	6/11 (54.6)	
18-23	18/33 (54.6)	
Proband status		
Depressed	18/30 (60.0)	NS
Normal	6/15 (40.0)	
Psychosocial/family risks (wave 1)		
Poor marital adjustment		
No	7/16 (43.8)	NS
Yes	12/16 (75.0)	
Low family cohesion		
No	7/15 (46.7)	NS
Yes	17/29 (58.6)	
Affectionless control		
No	15/26 (57.7)	NS
Yes	9/18 (50.0)	
Parent-child discord		
No	15/30 (50.0)	NS
Yes	9/15 (60.0)	
Parental divorce		
No	14/28 (50.0)	NS
Yes	10/17 (58.8)	
IQ (PPVT)		
<100	8/15 (53.3)	NS
100+	14/27 (51.9)	

(cont)

depressive disorder (Table 5). Since the analysis was an attempt to ascertain correlates of consistency in reports and not correlates of onset, offspring without a positive report at either wave were eliminated from the analysis. Power considerations limited analyses to reports of major depression. Offspring who reported lifetime major depression with onset before wave 1 at only one wave (either wave 1 or wave 2) but with no new onset between wave 1 and wave 2 were considered inconsistent reporters of major depression. Offspring who reported lifetime major depression at both waves were considered consistent reporters.

Table 5 suggests that none of the family risk factors or demographic variables were important correlates of reporting consistency. Several clinical variables were associated with consistency. Offspring with wave 1 CGAS scores of 60 or less (ie, moderately or highly impaired offspring), offspring who reported having been in treatment for any psychiatric disorder before wave 1, and offspring who reported having a comorbid diagnosis for dysthymia at wave 1 were significantly more likely to be consistent reporters of major depression. Additional analyses revealed that consistent reporters of major depression had a larger mean number of diagnoses at wave 1 than inconsistent reporters (5.0 vs 3.4; $t[43]=2.42$; $P<.05$).

Following research by Aneshensel et al[18] on adult community samples, we explored whether recent depressive episodes or current depressive symptoms prompted recall of lifetime reports of major depression. Offspring diagnosed as having current depressive episodes at wave 1 were not more likely to recall lifetime depression at wave 2 than were offspring with wave 1 lifetime (but not current) depressive diagnoses. Most offspring who provided consistent reports of major depression at both waves did not report the occurrence of an interim depressive episodes between wave 1 and wave 2. Of the 24 offspring who "remembered" having a lifetime diagnosis of depression at wave 2, only 7 (29%) reported having an episode in between wave 1 and wave 2. Recall was also unaffected by reports of current depressive symptoms at wave 2, as measured by the Center for Epidemiological Studies Depression Scale for Children (CES-DC) (mean CES-DC score at wave 2 for offspring who remembered wave 1 major depression, 18.4; mean score for offspring who forgot wave 1 major depression, 15.4; $t[35]=0.685$; not significant).

Logistic regressions including treatment history, CGAS score (dichotomized at 60), reports of dysthymia at wave 1, child's age at wave 1, and child's sex as covariates were analyzed with consistency in reports of major depression as the outcome (Table 6). Treatment history proved to be the most powerful predictor of recall. Compared with others, offspring with a treatment history had 8.5 times the odds of providing consistent reports of major depression at both waves (95% confidence interval 1.8 to 40.2).

In earlier research assessing agreement of parents and offspring on child's psychiatric diagnosis at wave 1, we found that the girls reporting depression with first onset at age 15 years or older reported much higher rates than did offspring with younger ages at onset and showed the highest level of disagreement with parents; this was true regardless of parents' clinical status.[4] A nonsignificant trend suggested that girls with first onset of major depression in late adolescence were less likely than other offspring with depression (ie, girls with early onset and boys with early and late onset) to recall major depression at the wave 2 interview ($\chi^2[1]=2.93$). A logistic regres-

Table 5.—Relationship Between Consistency* in Offspring Reports of Lifetime Major Depression Over Two Years by Demographic, Psychosocial, Family Risk, and Diagnostic Variables† (cont)

Clinical characteristics		
GAS wave 1 offspring report		
61+	5/18 (27.8)	<.01
<61	19/27 (70.4)	
Ever treated before wave 1		
No	4/17 (23.5)	<.01
Yes	20/28 (71.4)	
Age at onset of MDD (offspring report wave 1), y		
<12	8/11 (72.7)	NS
12-17	8/20 (40.0)	
18+	8/11 (72.7)	
Current MDD wave 1 (offspring report)		
No	20/33 (60.6)	NS
Yes	4/9 (44.4)	
Dysthymia (offspring report wave 1)		
No	10/26 (38.5)	<.05
Yes	14/19 (73.7)	
Parent-offspring agreement, wave 1 depression		
Yes	7/11 (63.6)	NS
No	16/27 (59.3)	
Any anxiety disorder (wave 1 offspring report)		
No	10/20 (50.0)	NS
Yes	14/25 (56.0)	
Any substance abuse (wave 1 offspring report)		
No	16/35 (45.7)	NS
Yes	8/10 (80.0)	
Any conduct disorder (wave 1 offspring report)		
No	16/30 (53.3)	NS
Yes	8/15 (53.3)	

*Consistency indicates lifetime report of major depression at wave 1 and wave 2.

†PPVT indicates Peabody Picture Vocabulary Test; GAS, Global Assessment Scale; MDD, major depressive disorder; and NS, not significant.

Table 6.—Logistic Regressions Predicting Consistent Reports of Lifetime Major Depression in Offspring*

	Coefficient ± SE (OR 95% CI†)	
Variable	Model 1	Model 2
Parent's proband status§	1.19 ± 0.87 (0.60, 18.02)	1.27 ± 0.83 (0.70, 18.12)
Child's age at wave 1	0.06 ± 0.10 (0.87, 1.30)	. . .
Child's sex‖	−0.58 ± 0.87 (0.10, 3.11)	. . .
Child's previous treatment history	1.91 ± 0.98 (1.00, 45.74)	2.14 ± 0.79¶ (1.80, 40.22)
Child's dysthymia at wave 1	1.46 ± 0.84 (0.83, 3.11)	1.59 ± 0.77# (1.04, 23.20)
GAS score <61 (impaired social functioning)	0.43 ± 0.98 (0.22, 10.58)	. . .

*Analysis includes only children who reported major depression at either wave 1 or wave 2 (N = 45). GAS indicates Global Assessment Scale.

†95% Confidence interval for the odds ratio; odds ratios assess the odds of consistent reports of major depression.

‡Excludes variables with $P > .3$ in model 1.

§Codes are 1 for depressed proband and 0 for normal proband.

‖Codes are 1 for girl and 0 for boy.

¶$P < .01$.

#$P < .05$.

sion model controlling for other critical predictors of consistency identified above (ie, previous treatment history and previous report of dysthymia) did not confirm this trend.

COMMENT

Summary of the Findings

The major findings on children's recall of lifetime *DSM-III* diagnosis based on two K-SADS-E interviews given 2 years apart were as follows: Recall for major depression with the use of *DSM-III* criteria was good; it was fair when stricter criteria were imposed. Recall for substance abuse and conduct disorder was good. Recall for anxiety disorder was generally poor, regardless of subtype. Offspring of depressed probands showed better recall than offspring of normal subjects. Boys showed better recall than girls for major depression; this difference was more pronounced in the older group of offspring (18 to 23 years) than in the younger group (6 to 17 years). Offspring interviewed by the same rater at both waves tended to show better recall for most diagnoses than did offspring interviewed by different raters. The mean differences in reported age at onset for major depression, conduct disorder, and substance abuse were small (<1 year) and not significantly different from 0.

The most important correlates of consistency of child reports of depression and any diagnosis were factors associated with the severity of the child's impairment at wave 1, including poor social functioning, previous treatment history, and comorbidity (ie, number of diagnoses and lifetime diagnosis of dysthymia). A nonsignificant trend suggested that girls reporting onsets of major depression at age 15 years or older may be more likely to forget their diagnosis.

Limitations

One possible source of bias in this study may stem from interviewers' awareness of a child's wave 1 diagnostic history at the wave 2 interview. We noted that for less than half of the sample (63 offspring [42%]), the same interviewer interviewed the subject at both waves. Interviewer awareness of previous diagnostic history may have led to more extensive probing of certain subjects. If such probing had occurred, it may have inflated κ coefficients. We showed that, overall, offspring interviewed by the same rater showed better diagnostic recall than offspring interviewed by different raters. Nevertheless, we also showed that interviewer bias had a negligible impact with respect to overall conclusions about the stability of recall; the coefficients of agreement obtained from the overall sample were closer in magnitude to the coefficients obtained from offspring with different raters than they were to those obtained from offspring with the same rater.

Shrout and his colleagues[19] emphasized that reliability coefficients obtained from one particular sample may not be generalizable to samples drawn from other populations. Our findings with respect to recall stability are limited in their generalizability. The sample of informants in the present study was offspring at high and low risk for depression by virtue of their parents' diagnostic status. Although our findings suggest that nonreferred samples of offspring provide relatively stable reports of most lifetime diagnoses, replication of these findings in other nonreferred samples is necessary before firm conclusions can be drawn.

Previous Research on Diagnostic Stability in Offspring

Table 7 displays the results of six studies (including this one) assessing stability of *DSM-III* diagnoses from semistructured interviews with offspring. To our knowledge, the present study is the first to assess 2-year stability of recall of lifetime diagnoses in a nonreferred sample of offspring. In contrast to the present study, four of the other studies assessed the short-term reliability of current (and not lifetime) diagnoses in patient samples.[3,5,6,20] An additional study assessed recall of current diagnoses occurring 6 months to 2 years earlier in a small group of outpatients.[7] Two of the studies[5,7] used summary ratings based on information provided by both parents and offspring. These four design features—assessment of current diagnosis, shorter duration of reinterview time, a more seriously impaired group of subjects, and the use of multiple informants—should have increased the magnitude of the reliability coefficients for five other studies relative to the present study.

Table 7 lists κ values for major depression, anxiety disorder, and conduct disorder. In general, the reliability of reports of depressive disorder (or symptoms) was good to excellent, regardless of instrument, informant, or reinterview time frame. The reliability coefficients for depression from the present study are slightly higher than the coefficient generated by the short-term study using the K-SADS–Present Episode Version,[5] similar in size to the coefficients generated by the Diagnostic Interview Schedule for Children,[3] and smaller than the coefficients generated by the Diagnostic Interview Schedule for Children–Revised Version,[20] the Diagnostic Interview for Children and Adolescents,[6] and the long-term study using the K-SADS–Present Episode Version and the K-SADS-E.[7] The coefficients for anxiety disorder generated from the present study are similar to the coefficient generated from the K-SADS–Present Episode Version[5] and much smaller than coefficients generated from the Diagnostic Interview for Children and Adolescents, the Diagnostic Interview Schedule for Children, and the Diagnostic Interview Schedule for Children–Revised Version.[3,6,19] With one exception (the Diagnostic Interview Schedule for Children–Revised Version[19]), coefficients for conduct disorder suggest excellent reliability.

We found no other published studies assessing the reliability of reports of age at onset of psychiatric disorder in children, adolescents, and young adults. A recent summary of studies of adult "mixed" (ie, patient and nonreferred) samples and community samples indicated that recall of age at onset for lifetime major depressive disorder (using Schedule for Affective Disorders and Schizophrenia criteria and Research Diagnostic Criteria) was good to excellent.[21] Intraclass correlation coefficients generally exceeded .60 (with a range of .51 to .88) regardless of sample composition and length of time between interviews. Age at onset recall in this study (.54) was not as good as recall obtained from most adult studies.

Compared with adult informants, informants in this study had a relatively narrow age range for onset reports. This may result in relatively smaller between subject variability and a decrease in the relative magnitude of the intraclass correlation coefficient for children compared with adults. Thus, mean difference in reported age at onset may be a useful indicator of reliability in this study. Using this criteria, informants in this sample, like adult informants in previous studies, showed excellent reliability in their recall of age at onset for major depression.

CONCLUSIONS

The present study provided a rigorous test of diagnostic stability. We reinterviewed informants after a 2-year period, a time interval much longer than that generally employed in studies assessing test-retest reliability. Thus, taken as a whole, the findings provide support for the stability of recall of lifetime psychiatric diagnoses generated from child reports using the K-SADS-E.

Comparisons between different disorders suggest that age at onset and symptom course are important variables affecting long-term recall of lifetime diagnoses. Offspring showed consistently poor recall for anxiety disorders, regardless of subtype. In comparison with other diagnoses, offspring with anxiety disorder reported a mean age at onset that was much younger than that reported for the other disorders evaluated in this study. The most prevalent subtypes of childhood anxiety disorder, separation anxiety and phobias, do not commonly persist beyond late childhood.[22] Older children and adoles-

Table 7.—Studies on Test-Retest Reliability of Children and Adolescent Diagnoses*

Source, y	Sample	Assessment	Time	κ/ICC, Diagnosis†			
				MDD	Anxiety	ADD	CD
Chambers et al,[5] 1985	52 inpatients aged 6-17 y	K-SADS-P; current "syndromal" *DSM-III* diagnoses (summary ratings)	3 d	.54	.24	. . .	.63
Edelbrock et al,[3] 1985	242 inpatients and outpatients aged 6-18 y	DISC; current symptom scores	9 d	.64	.62	.67	.73
Orvaschel et al,[7] 1982	17 ex-outpatients aged 6-11 y at initial interview	K-SADS-P at initial interview; K-SADS-E at follow-up; current *DSM-III* diagnoses at initial interview; past diagnoses at follow-up	6 mo-2 y	1.00	. . .	. . .	. . .
Welner et al,[6] 1987	27 inpatients aged 7-17 y	DICA-C; current *DSM-III* diagnoses	1-7 d	.90	.76‡	1.00	1.00
Shaffer et al,[20] 1988	41 inpatients and outpatients	DISC-R; 6-mo *DSM-III-R* diagnoses	1-3 wk	.77	.72§	.48	.55
Present study	150 offspring at high and low risk for depression, aged 6-23 y	K-SADS-E; lifetime *DSM-III* diagnoses	2 y	.61	.23	.38	.76

*MDD indicates depressive symptoms or disorder; ICC, intraclass correlation coefficient; Anxiety, anxiety symptoms or disorders; ADD, attention deficit symptoms or disorders; CD, conduct disorder symptoms or conduct disorder; K-SADS-P, Schedule for Affective Disorders and Schizophrenia for Children, Present Episode Version; DISC, Diagnostic Interview Schedule for Children; K-SADS-E, Schedule for Affective Disorders and Schizophrenia in Children, Epidemiologic Version; DICA-C, Diagnostic Interview for Children and Adolescents–Child; and DISC-R, DISC–Revised Version.

†When syndrome scales were used for diagnoses, ICCs were reported; when *DSM-III* diagnoses were obtained, κ values were reported.

‡Separation anxiety disorder/phobic disorder/overanxious disorder.

§Separation anxiety disorder.

cents may forget lifetime reports of disorders with early onset and relatively brief course. Investigators tracing the developmental course of such disorders would be advised not to rely on retrospective recall of older children and adolescents for case definition.

There are several other possible reasons for lack of the stability of reports of anxiety disorder in this study. Nonreferred samples of children (and parents) may not be good informants with respect to anxiety disorders. The poor reliability of anxiety disorders yielded by K-SADS assessments on an inpatient sample,[5] similar to the findings from this study and in contrast to findings generated from patient samples assessed with other instruments, may suggest limitations in the anxiety disorder sections of K-SADS. Previous reliability studies assessed either symptoms of anxiety disorder[3,5,6] or specific anxiety disorder subtypes, such as separation anxiety.[20] The test-retest reliability of four of the five subtypes we assessed (ie, phobia, panic disorder, obsessive-compulsive disorder, and overanxious disorder) has not been described elsewhere. Replication of our findings with respect to these subtypes may be indicative of problems with *DSM-III* criteria used to define them.

Aneshensel et al[18] concluded that life events were important correlates of long-term recall of lifetime reports of major depression in a community sample of adults. Although the present study did not assess life events per se, it did assess indexes of family discord that may be associated with life events. These measures of family discord were not important correlates of recall. In a previous study of major depression in an adult community sample conducted by Bromet et al,[23] indexes of life events were not correlated with long-term diagnostic recall. Bromet et al and we in the present study used a semistructured interview to derive diagnoses meeting *DSM-III* criteria. Aneshensel et al used much less restrictive criteria; all respondents with lifetime reports of depressed moods were classified as having the disorder. These discrepancies suggest that life events may be less important as predictors of recall of clinical diagnoses of depressive disorder than they are of recall of depressive symptoms.

Surprisingly, the overall stability of recall was better for major depression defined according to *DSM-III* criteria than according to stricter criteria. Our use of the stricter criteria was an attempt to come to terms with the relatively high rates of *DSM-III* major depression observed for offspring in this and other studies using semistructured interviews.[24] Questions about the reliability of diagnoses based on strict criteria, along with the finding that best-estimate CGAS ratings were among the best predictors of consistency in reports of *DSM-III* major depression, suggest that use of impairment criteria based on the CGAS may be a superior approach to enhancing diagnostic validity. We are currently exploring this issue in greater detail.

The present study has important methodologic implications for epidemiologic research. To some extent, recall is a function of the assessment process. Our study suggests that more consistent diagnostic reports (enhanced reliability) will be obtained if semistructured interviews are administered by the same interviewer at both waves. The finding that long-term diagnostic recall is associated with higher levels of impairment and treatment history suggests that long-term recall may be an indicator of diagnostic validity; young people with more severe forms of disorder remember their disorder. If reporting consistency is an index of caseness, diagnoses based on multiple waves of interviews may facilitate more accurate conclusions about prevalence rates for psychiatric disorder in young people. It should be stressed, however, that the clinical implications of long-term recall in children, adolescents, and young adults are not fully understood. Further confirmation of our results in studies with other similar nonreferred samples is needed before firm conclusions about the role of recall in case definition can be drawn.

This research was supported in part by National Institute of Mental Health (Bethesda, Md) grants MH36197 and MH28274 and by grant 86-213, Child and Adult Depressive Disorders: A Test of Continuities Using Family-Genetic Data, from the John D. and Catherine T. MacArthur Foundation Mental Health Research Network on Risk and Protective Factors in the Major Mental Disorders.

References

1. Rutter M, Garmezy N. Developmental psychopathology. In: Hetherington EM, section ed. *Socialization, Personality and Social Development: Mussen's Handbook of Child Psychology.* 4th ed. New York, NY: John Wiley & Sons Inc; 1983;4:775-911.

2. Rutter M. Epidemiological approaches to developmental psychopathology. *Arch Gen Psychiatry.* 1988;45:486-495.

3. Edelbrock CS, Costello AJ, Dulcan MK, Kalas R, Conover NC. Age differences in the reliability of the psychiatric interview of the child. *Child Dev.* 1985;56:265-275.

4. Weissman MM, Wickramaratne PJ, Warner V, John K, Prusoff BA, Merikangas KR, Gammon GD. Assessing psychiatric disorders in children: discrepancies between mothers' and children's reports. *Arch Gen Psychiatry.* 1987;44:747-753.

5. Chambers WJ, Puig-Antich J, Hirsch M, Paez P, Ambrosini PJ, Tabrizi MA, Davies M. The assessment of affective disorders in children and adolescents by semistructured interview: test-retest reliability of the Schedule for Affective Disorders and Schizophrenia for School-Age Children, Present Episode Version. *Arch Gen Psychiatry.* 1985;42:696-702.

6. Welner Z, Reich W, Herjanic B, Jung KG, Amado H. Reliability, validity, and parent-child agreement studies of Diagnostic Interview for Children and Adolescents (DICA). *J Am Acad Child Adolesc Psychiatry.* 1987;26:649-653.

7. Orvaschel H, Puig-Antich J, Chambers LW, Tabrizi MA, Johnson RA. Retrospective assessment of prepubertal major depression with the Kiddie-SADS-E. *J Am Acad Child Adolesc Psychiatry.* 1982;21:392-397.

8. Fendrich M, Weissman MM, Warner V. Longitudinal assessment of major depression and anxiety disorders in children. *J Am Acad Child Adolesc Psychiatry.* In press.

9. Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, Aluwahlia S. A children's global assessment scale (CGAS). *Arch Gen Psychiatry.* 1983;40:1228-1231.

10. Dunn LM, Dunn LM. *Peabody Picture Vocabulary Test—Revised.* Circle Pines, Minn: American Guidance Service; 1981.

11. Fendrich M, Warner V, Weissman MM. Family risk factors, parental depression and childhood psychopathology. *Dev Psychol.* 1990; 26:40-50.

12. Locke HJ, Wallace KM. Short-term marital adjustment and prediction tests: their reliability and validity. *J Marriage Fam Living.* 1976;38:15-25.

13. Olson DH, Spenkle DH, Russell CS. Circumplex model of marital and family systems, 1: cohesion and adaptability dimensions, family types and clinical applications. *Fam Process.* 1979;18:3-78.

14. Parker G, Tupling H, Brown LB. A parental bonding instrument. *Br J Med Psychol.* 1979;52:1-10.

15. Fendrich M, Weissman MM, Warner V. Screening for depressive disorder in children and adolescents: validating the CES-DC. *Am J Epidemiol.* 1990;131:538-550.

16. Fleiss JL. *Statistical Methods for Rates and Proportions.* New York, NY: John Wiley & Sons Inc; 1988.

17. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420-428.

18. Aneshensel CS, Estrada AL, Hansell MJ, Clark VA. Social psychological aspects of reporting behavior: lifetime depressive episode reports. *J Health Soc Behav.* 1987;28:232-246.

19. Shrout PE, Spitzer RL, Fleiss JL. Quantifications of agreement in psychiatric diagnosis revisited. *Arch Gen Psychiatry.* 1987;44:172-177.

20. Shaffer D, Schwab-Stone M, Fisher P, Davies M, Piacentini J, Gioia P. *A Revised Version of the Diagnostic Interview Schedule for Children (DISC-R)–Results of a Field Trial and Proposals for a New Instrument (DISC-2).* Report submitted to the Division of Epidemiology and Biometrics, National Institute of Mental Health, Washington, DC. 1988.

21. Prusoff BA, Merikangas KR, Weissman MM. Lifetime prevalence and age of onset of psychiatric disorders: recall four years later. *J Psychiatr Res.* 1988;22:107-117.

22. McDermott JF. Anxiety disorder in children and adults: coincidence or consequence? *Integr Psychiatry.* 1985;3:158-167.

23. Bromet EJ, Dunn LO, Connell MM, Dew MA, Schulberg HC. Long-term reliability of diagnosing lifetime major depression in a community sample. *Arch Gen Psychiatry.* 1986;43:435-440.

24. Weissman MM. Psychopathology in the children of depressed parents: direct interview studies. In: Dunner DL, Gershon ES, Barret J, eds. *Relatives at Risk for Mental Disorders.* New York, NY: Raven Press; 1988:143-159.