Journal of Abnormal Child Psychology, Vol. 21, No. 6, 1993

Evaluation of the Diagnostic Interview for Children and Adolescents for Use in General Population Samples

Michael H. Boyle,[1,2] David R. Offord,[1] Yvonne Racine,[1] Mark Sanford,[1] Peter Szatmari,[1] Jan E. Fleming,[1] and Nancy Price-Munn[1]

This article presents evaluative data on the use of the Diagnostic Interview for Children and Adolescents-Revised (DICA-R) to classify DSM-III-R disorders in the general population. Data for the analyses came from a probability sample (N = 251) of parent–child/adolescent dyads aged 6 to 16 separately administered the DICA-R on two occasions, 10- to 20 days apart, by trained lay interviewers and child psychiatrists. Data are presented on prevalence, test–retest reliability, parent–child/adolescent agreement, and trained lay interviewer–child psychiatrist agreement. High prevalences of oppositional defiant disorder derived from parent assessments and overanxious disorder and dysthymia derived from adolescent assessments suggest that these disorders may be overidentified. Interview data provided by 6- to 11-year olds to classify the internalizing disorders were too unreliable to be useful. Agreement between parent–child/adolescent dyads was generally low while agreement between trained lay interviewers–child psychiatrists was generally high.

Manuscript received in final form April 20, 1993.
This work was supported by funds from the National Health Research and Development Program (grant number 6606-3760-42) and the Ontario Ministry of Community and Social Services, and was carried out by the Child Epidemiology Unit, Department of Psychiatry, McMaster University and the Child and Family Centre, Chedoke-McMaster Hospitals, Hamilton, Ontario. Dr. Boyle is supported by a William T. Grant Foundation Faculty Scholar Award; Dr. Offord by a National Health Scientist Award, Health and Welfare Canada; Dr. Szatmari by a Research Fellowship, Ontario Mental Health Foundation; and Dr. Fleming by a NARSAD Young Investigator Award. The authors wish to thank the Hamilton Board of Education, particularly Mr. Keith Reilly and Mr. Owen Jackson; principals and teachers of the Board; and parents and students for their help on this project. Our appreciation also extends to Ms. Barb Hall who assisted in the analysis of the data.
[1]Department of Psychiatry, McMaster University, Hamilton, Ontario, Canada L8N3Z5.
[2]Address all correspondence to Michael H. Boyle, Child Psychiatry Research Centre, Chedoke McMaster Hospitals, Chedoke Division, Patterson Building, P.O. Box 2000, Station "A," Hamilton, Ontario, Canada, L8N 3Z5.

0091-0627/93/1200-0663$07.00/0

A prerequisite for studying childhood psychopathology in the general population is the availability of instruments to classify psychiatric disorder. Although a number of different assessment methods might be considered (e.g., self-report, questionnaires, direct observation, psychological testing, clinical evaluations), structured interviews administered in person by trained lay people have come into favor (see Edelbrock & Costello, 1988). A face-to-face interview provides an opportunity to reduce unwanted sources of error that might arise from respondent lack of interest or motivation and problems with reading or following instructions. In addition, lay administration of an interview is far less expensive than clinical assessments.

For studies in the general population, particularly when large samples are involved, instruments to classify childhood disorders should be simple to apply, acceptable to respondents, relatively brief to administer, and easily converted to a score or classification. Other desirable features include broad coverage of disorders listed in the existing nosology, the ability to obtain assessments from different respondents (e.g., parents and children), applicability of the assessment across a fairly wide age span and a high level of structure to facilitate standardization and reduce training costs. At the present time, there are two structured interviews that meet most, if not all, of these practical requirements: the Diagnostic Interview Schedule for Children (DISC) and the Diagnostic Interview for Children and Adolescents (DICA). Although these two instruments are potentially suitable for use in general population studies, evidence about their reliability and validity is incomplete (Edelbrock & Costello, 1988). Not only is there a lack of measurement evaluation studies that have been reported but also none of the published studies has been done using a general population sample. In addition, these instruments tend to be moving targets, constantly undergoing revisions that could alter their performance characteristics.

The overall purpose of this study was to provide information for evaluating the revised version of the Diagnostic Interview for Children and Adolescents (DICA-R; Reich & Welner, 1988) administered to parents and children for purposes of classifying childhood psychiatric disorders in the general population. The DSM-III-R disorders selected for study included conduct disorder (CD), oppositional defiant disorder (ODD), attention-deficit hyperactivity disorder (ADHD), overanxious disorder (OAD), separation anxiety disorder (SAD), major depressive disorder (MDD), and dysthymia (DYS). The following information is provided on the DICA-R: (1) prevalence estimates of disorder, (2) test–retest reliability estimates, (3) estimates of interrespondent agreement between classifications of disorder arising separately from parent and child/adolescent assessments, and (4) the sensitivity and specificity of the DICA-R administered by trained lay

interviewers vs. the DICA-R administered by child psychiatrists. Data for this report come from a measurement evaluation study done during 1989–1990 (see also Boyle, Offord, Racine, et al., (1993).

METHODS

Sample

Figure 1 shows the sampling design for the study. A simple random sample of 2317 children aged 6 to 16 years was taken in the year 1989 from among those attending public schools in an industrialized, urban setting (94 schools with 27,629 students). There were 70 families in which two siblings were included in the sample and one was selected at random from each family. Twenty-eight other children were removed from the sample because of ineligibility (e.g., living outside the region, severe mental handicap in the child, etc.). Among the eligible sample of 2219 children, 1751 (79%) parents participated in stage 1 assessments by completing problem checklists and giving informed consent to obtain teacher assessments. The reasons for non-participation included refusal, 15% (324/2219); language problems, 4% (93/2219); and miscellaneous, 2% (51/2219). In the 6- to 11-year-old age group, 898/1122 (80%) participated, and in the 12- to 16-year-old age group, 853/1097 (78%) participated. With parental consent, teachers were mailed questionnaires to complete. Homeroom teachers were asked to complete the assessments for children in elementary school. For children on a rotary system or in high school, principals were asked to have the teacher (or guidance counselor) who best knew each child to complete the assessment.

A stratified random sample of participants was selected for stage 2 assessments at 1- to 3 months after completing stage 1. The strata were based on age (6 to 11, 12 to 16), sex, and high vs. low symptom scores in the index child based on the original Ontario Child Health Study (OCHS) scales (Boyle et al., 1987). Children with high symptom scores were oversampled to increase the number of children likely to meet DSM-III-R criteria for disorder. Among the 329 selected for stage 2, 251 (76%) agreed to participate. These included 114/142 (80%) in the 6- to 11-year-old age group and 137/187 (73%) in the 12- to 16-year-old age group.

Administration of the DICA-R

Parent and child/adolescent dyads participating in stage 2 ($N = 251$) were administered the revised version of the Diagnostic Interview for Chil-

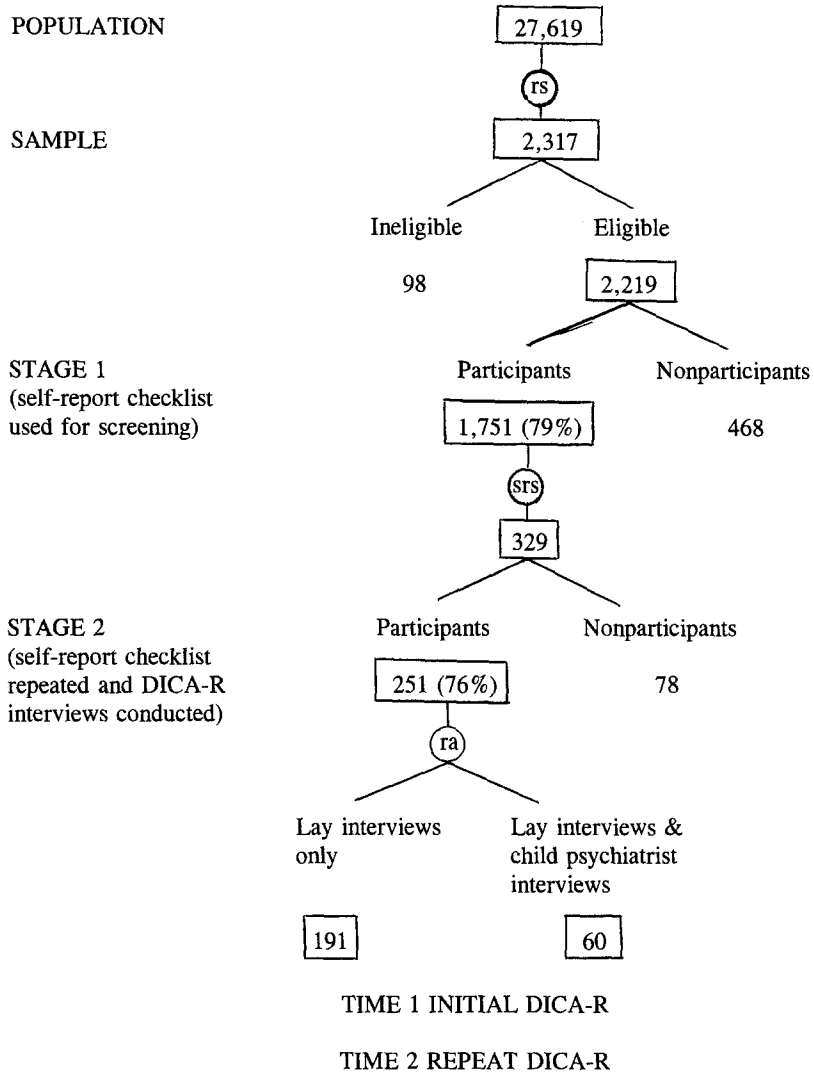

Fig. 1. Survey design for measurement study. The letters in the circles: rs, random sample; srs, stratified random sample; ra, random allocation. DICA-R, Diagnostic Interview for Children and Adolescents—Revised.

dren and Adolescents (DICA-R) on two separate occasions, 3 weeks apart. Stage 2 families were divided randomly into two groups. In the first group (n = 191), the time 1 and time 2 DICA-R interviews with parents and

children/adolescents were done by different trained lay interviewers. In the second group (n = 60), the time 1 and time 2 DICA-R interviews with parents and children were done by either two lay interviewers, a lay interviewer and a child psychiatrist, or two child psychiatrists. The purposes of these latter two groups were to be able to compare assessments obtained by trained lay interviewers vs. child psychiatrists and to examine agreement between child psychiatrists independently administering the DICA-Rs. Block randomization was used to achieve balance in the chronological ordering of interviews (i.e., lay then psychiatrist vs. psychiatrist then lay) and in the number of time 1 and time 2 interviews done with parents and children: 36 pairings of lay interviewer and child psychiatrist for both parent and child/adolescent interviews, and 12 pairings of different child psychiatrists for both parent and child/adolescent interviews. Although the block randomization worked well, scheduling conflicts prevented it from working exactly as planned. The final distribution of interviews included the following: 32 pairings of lay interviewer and child psychiatrist for both parent and child/adolescent interviews; 15 pairings of different child psychiatrists for parent interviews; and 13 pairings of different child psychiatrists for child adolescent interviews. Balanced chronological ordering of interviews was almost attained. Of the 32 parent interviews, 15 followed the lay then psychiatrist ordering and 17 followed the psychiatrist then lay ordering. Of the 32 child/adolescent interviews, 17 followed the lay then psychiatrist ordering and 15 followed the psychiatrist then lay ordering. Parents and children/adolescents were informed when a child psychiatrist was the interviewer.

DICA-R

The DICA-R is a revised version of the DICA. The DICA was developed by Herjanic and her associates at Washington University in St. Louis (Herjanic & Campbell, 1977; Herjanic & Reich, 1982) to classify childhood psychiatric disorders. The classification of disorders in the DICA was based originally on the International Classification of Psychiatric Disorders in combination with the Feighner criteria (Feighner et al., 1972). In 1981, a revised version of the DICA was developed, patterned after the National Institute of Mental Health Diagnostic Interview Schedule (Robins, Helzer, Croughan, & Ratcliff, 1981) and based upon DSM-III criteria. A further revision in 1988 produced the DICA-R, developed to classify DSM-III-R categories of disorders. The DICA-R consists of three versions nearly identical in structure and content: one for children aged 6 to 12, another for adolescents aged 13 to 17, and a parent version that covers the 6- to

17-year-old age span. In comparison with other interview schedules used to classify childhood psychiatric disorders, the DICA-R is very structured (Edelbrock & Costello, 1988; Hodges, 1993).

Modifications

The DICA-R was modified to serve the objectives and constraints of the present study. To reduce the burden to respondents, only those disorders relevant to the objectives of the measurement evaluation exercise were retained. These disorders included CD, ODD, ADHD, OAD, SAD, MDD, and DYS. Avoidant disorder and somatization disorder were also retained for measurement evaluation purposes not related to this report. The DICA-R is worded to assess lifetime as well as current classifications of disorders. Question wording was changed to focus on the present. In most instances, this meant removing the word *ever* and using the present tense. The exception to this was CD where question wording was left intact and included the word *ever*. When questions were answered positively, a followup question asked about duration of the particular symptom. Changes to the physical layout of the DICA-R questionnaire were also made to facilitate the correct observance of branching sequences and to make data capture more efficient. Other than verb tense, no changes were made to the content and sequence of questions. The response options of *yes, sometimes*, and *no* were maintained from the DICA-R. Also maintained were instructions around the use of probes and the diagnostic algorithms for classifying each disorder.

Interviewers and Training

The lay interviewers consisted of 16 women who possessed extensive interviewing experience from working on federally sponsored surveys such as the Canada Labour Force Survey and Census. Training included five steps: (1) home study of the interviewer manual and interview, (2) a large group interview (with interviewers taking turns asking questions) of an adult using scripted responses, (3) practice interviews among the interviewers, (4) observation and coding of a taped interview of an experienced interviewer administering the DICA-R to an adolescent, and (5) a field interview with a parent or child drawn from stage 1 participants reserved for study training purposes. Steps (2) and (3) above were used to instill and reinforce the need to ask questions exactly as worded, to use probes only as directed, and to refrain from giving any judgmental cues about responses. In addition, interviewers were trained to respond appropriately to parents or children exhibiting distress. Step (4) provided an opportunity to assess

the ability of interviewers to correctly code the interview based on respondent answers contained on tape. The practice interviews [step (5)] done with parents and children were all reviewed by one of the investigators (MB) to identify coding errors (e.g., missed check items or wrong sequencing) with a view to improving interviewer transcription of responses.

In addition to lay interviewers, there were five child psychiatrists who interviewed parents and children using the DICA-R. All of the child psychiatrists were practicing clinicians with academic appointments. Training included (1) home study of the interviewer manual and interview, (2) practice interviews with patients which were videotaped and critically evaluated, and (3) an interview with a parent or child drawn from stage 1 participants reserved for study training purposes. The child psychiatrists were instructed to complete each section of the interview as though they were lay interviewers. After completing each section they were free to use their clinical skills to clarify subject responses. Clarifications and changes were made in red ink to distinguish them from their initial findings made in black pencil. Each child psychiatrist interview was then scored as a regular interview (no additional probing) and an enriched interview (includes clinical probing). Analysis of agreement between child psychiatrist interviews scored as "regular" and as "enriched" revealed few differences in classification.

Analyses

For prevalence estimates, the assessments provided by each respondent ($N = 251$) during the first interview were weighted inversely to the probability of respondent selection from stage 1 participants ($N = 1751$). This was done to generate unbiased estimates for the population. Only definite *yes* responses contributed to the identification of disorder. CD, ADHD, and ODD were subdivided into mild, moderate, and severe according to the scoring rules specified in the DICA-R. Interviews done by child psychiatrists scored as regular interviews were combined with interviews done by trained lay interviewers.

Percent agreement and kappa (Cohen, 1960), a measure of agreement corrected for chance, provided the basis for evaluating 1- to 3-week test–retest reliability and levels of agreement between parent and child assessments. The term *reliability* is used to emphasize the focus on temporal stability (time 1 and time 2). The term *agreement* is used to focus attention on data provided by different respondents (parent–child/adolescent) and on data elicited by different interviewers (lay–child psychiatrist). The measures used to assess reliability and agreement were the same. For these and subsequent analyses, classifications of CD, ODD, and ADHD were col-

lapsed from mild, moderate, and severe into disorder present. Respondent assessments (N = 251) were weighted to reflect the large sample (N = 1751) experience, and once again interviews done by child psychiatrists scored as regular interviews were combined with interviews done by trained lay interviewers.

The extent to which trained lay interviewer assessments corresponded with child psychiatrist assessments was evaluated by estimates of sensitivity, specificity, and kappa. Percents of children classified with disorders were compared between the lay interviewers and child psychiatrists to examine bias (i.e., systematic over or under identification of disorder by interviewers). These analyses used child psychiatrist interviews that were scored as enriched interviews. Respondent assessments were not weighted for these analyses because of the small sample contributing to the estimates. Any inaccuracy in the weights could lead to a substantial distortion of results.

RESULTS

Table I gives the prevalence estimates of disorder by respondent and child/adolescent age. The levels of mild, moderate, and severe for CD, ODD, and ADHD reflect the classifications accompanying the scoring algorithms in the DICA-R. For each disorder, the mild level corresponds to the minimum number of symptoms specified in DSM-III-R needed to reach threshold (e.g., three symptoms for CD). The moderate and severe levels are defined by the presence of additional symptoms (e.g., five or more symptoms for CD moderate and seven or more symptoms for CD severe). Symptoms were coded as present only if the respondent provided an unequivocal *yes* response to the question; *sometimes* responses were recoded as *no*, symptom not present. In addition to presenting estimates for the individual disorders, estimates are given as well for disorder groupings (e.g., externalizing and internalizing disorders).

Inspection of Table I reveals a marked variation in the prevalences of disorders. For example in the 6- to 11-year-old age group, prevalences of CD (parent, child), SAD (parent) and MDD (parent, child) are very low while the prevalence of ODD (parent) is very high. In the 12- to 16-year-old age group, the prevalence of ADHD (adolescent) and SAD (parent, adolescent) are very low, while the prevalences OAD (adolescent) and DYS (parent, adolescent) are very high. In general, parent assessments vs. child/adolescent assessments yield higher prevalences of the externalizing disorders (CD, ODD, ADHD) while child/adolescent assessments vs. parent assessments yield higher prevalences of the internalizing disorders (OAD, MDD, DYS).

Table I. DSM-III-R Diagnostic Classifications Based on Parent and Child/Adolescent Assessments Derived from the DICA-R: Prevalence per 100 by Child Age[a]

Classification levels	6 to 11 years		12 to 16 years	
	Parent	Child	Parent	Adolescent
CD				
Mild	0.0	0.0	0.2	0.4
Moderate	0.4	0.6	3.2	1.4
Severe	0.0	0.4	0.0	0.0
ODD				
Mild	5.9	1.3	2.7	2.6
Moderate	0.9	0.7	2.2	3.2
Severe	6.5	4.5	5.5	0.7
ADHD				
Mild	2.9	1.6	0.9	1.1
Moderate	0.9	0.7	0.8	0.0
Severe	1.0	0.5	3.6	0.0
EXT	17.4	7.8	10.8	7.9
OAD	8.4	11.1	6.5	20.5
SAD	1.5	8.5	0.5	1.0
MDD	1.1	0.6	4.0	5.4
DYS	13.5	14.5	13.8	21.3
INT	18.5	25.8	20.5	32.0

[a]*Note*: Individual respondent classifications are weighted inversely to the probability of respondent selection to reflect the total sample experience (N = 1751). Sample sizes are 114 in the 6- to 11-year-old age group and 137 in the 12- to 16-year-old age group. CD, conduct disorder; ADHD, attention-deficit hyperactivity disorder; ODD, oppositional defiant disorder; OAD, overanxious disorder; SAD, separation anxiety disorder; MDD, major depressive disorder; DYS, dysthymia; EXT, one or more of CD, ODD, ADHD; INT, one or more of OAD, SAD, MDD, DYS; DICA-R, Diagnostic Interview for Children and Adolescents—Revised.

Table II provides estimates of 1- to 3-week test–retest reliability by respondent and age. These estimates include interviewer error because the two interviews were conducted by different interviewers. Kappa values tend to be high for parental assessments of their 12- to 16-year-olds. For example, 10 of the 12 estimates exceed .40. In contrast, kappa values tend to be low for child assessments in the 6- to 11-year-old age group where 11 of the 12 estimates are below .40. The distribution of kappa values for parental assessments in the 6- to 11-year-old age group (7 of the 12 estimates exceed .40) and adolescent assessments in the 12- to 16-year-old age group (6 of the 12 estimates exceed .40) are comparable.

In addition to differences in test–retest reliability associated with child age and respondent, there are differences associated with the disorders.

For example, disorders with prevalences < 1.0% such as CD (parent, 6- to 11-year-old age group) and SAD (parent and adolescent, 12- to 16-year-old age group) yield very high percent agreements but kappa estimates around .00 (Table II). Among child assessments of disorder, reliability estimates for the internalizing disorders are particularly low, going from −.00 to .21 (Table II). Among 12- to 16-year-olds, parental assessments of the externalizing disorders tend to be more reliable than their assessments of the internalizing disorders, whereas the reverse is true for adolescent assessments in this age group (Table II).

Table III gives percent agreement and kappa between parent and child/adolescent assessments of disorder at time 1 (the first interview) and at time 2 (the second interview). These estimates include interviewer error because the two interviews were conducted by different interviewers but they do not include error associated with timing because the interviews were conducted at the same time. Although the kappa estimates encompass a wide range (i.e., −.03 to .68), agreement levels are generally low. For example, 43 of the 48 (90%) kappa values in Table III are below .40 and only 1 of the 48 (2%) exceeds .60.

Table II. Test-Retest Reliability of DICA-R Assessments by Respondent and Child Age[a]

	6 to 11 years				12 to 16 years			
	Parent		Child		Parent		Adolescent	
Classification	%	*K*	%	*K*	%	*K*	%	*K*
CD	99	−.01	99	.37	99	.87	99	.92
ODD	84	.32	94	.33	94	.67	92	.28
CD–ODD	84	.32	93	.36	94	.67	92	.39
ADHD	98	.78	96	.43	99	.86	99	.24
EXT	84	.37	92	.32	93	.68	91	.38
OAD	86	.44	85	.01	96	.69	88	.54
SAD	99	.38	90	.01	99	–	99	−.00
OAD–SAD	86	.46	81	.04	96	.67	88	.56
MDD	99	.77	99	−.00	96	.31	95	.45
DYS	92	.57	86	.19	90	.51	84	.40
MDD–DYS	92	.57	84	.21	88	.47	83	.38
INT	81	.44	74	.06	87	.51	79	.47

[a]*Note*: % is percent agreement. *K* is kappa. Responses are weighted. See note, Table I. CD, Conduct disorder; ODD, oppositional defiant disorder; ADHD, attention-deficit hyperactivity disorder; EXT, one or more of CD, ODD, ADHD; OAD, overanxious disorder; SAD, separation anxiety disorder; MDD, major depressive disorder; DYS, dysthymia; INT, one or more of OAD, SAD, MDD, DYS; CD–ODD, CD and ODD combined; OAD–SAD, OAD and SAD combined; MDD-DYS, major depressive disorder and dysthymia combined; DICA-R, Diagnostic Interview for Children and Adolescents—Revised.

Table III. Parent–Child/Adolescent Agreement on DICA-R Classifications of DSM-III-R Disorders at Time 1 and Time 2 Assessments[a]

	6 to 11 years				12 to 16 years			
	Time 1		Time 2		Time 1		Time 2	
Classification	%	*K*	%	*K*	%	*K*	%	*K*
CD	99	−.01	99	.43	95	−.02	95	−.02
ODD	88	.39	89	.11	87	.16	84	.02
CD–ODD	88	.41	89	.21	86	.14	84	.06
ADHD	95	.34	94	.29	93	−.02	94	−.01
EXT	88	.48	88	.27	85	.14	82	.05
OAD	83	.03	84	.31	77	.06	86	.06
SAD	90	−.03	98	−.01	99	.68	99	—
OAD–SAD	77	.02	83	.30	77	.08	86	.06
MDD	98	−.01	99	−.00	91	−.05	95	.13
DYS	78	.10	89	.09	80	.32	88	.24
MDD–DYS	79	.11	90	.12	79	.37	89	.28
INT	74	.24	82	.30	78	.48	81	.18

[a]*Note*: % is percent agreement. *K* is kappa. Responses are weighted. See note, Table I. CD, conduct disorder; ODD, oppositional defiant disorder; CD–ODD, CD and ODD combined; ADHD, attention-deficit hyperactivity disorder; EXT, one or more of CD, ODD, ADHD; OAD, overanxious disorder; SAD, separation anxiety disorder; OAD–SAD, OAD and SAD combined; MDD, major depressive disorder; DYS, dysthymia; MDD–DYS, MDD and DYS combined; INT, one or more of OAD, SAD, MDD, DYS; DICA-R, Diagnostic Interview for Children and Adolescents—Revised.

Table IV displays the results of comparing child psychiatrist vs. lay administrations of the DICA-R. These analyses used the child psychiatrist interviews which were scored as enriched interviews (included clinical probing). Not all classifications add up to 32 because of missing responses. Because actual (unweighted) responses are shown, kappa values (unweighted) for comparing lay vs. lay administrations of the DICA-R are provided under the heading K_{LL}. The estimates under K_{LL} can be contrasted with the child psychiatrist vs. lay administration ($K_{CPE\text{-}L}$) because they are unweighted (unlike the estimates in Table II). In addition, kappa values are shown comparing child psychiatrist interviews which were scored as regular interviews vs. lay administration of the DICA-R ($K_{CPR\text{-}L}$).

With the exception of OAD–SAD, the estimates of sensitivity, specificity, and kappa tend to be high. For example, sensitivity exceeds 65% and specificity 70% for all disorders except OAD–SAD. Similarly, kappa values are greater than .55 (Table IV: $K_{CPE\text{-}L}$). When child psychiatrist interviews are scored as regular interviews, agreement with the lay-administered DICA-R is almost identical among parent respondents ($K_{CPR\text{-}L}$). However, kappa estimates for ADHD when the respondent is a child or

Table IV. Child Psychiatrist vs. Lay Administration of the DICA-R[a]

Respondent classification	+	+−	−+	−−	SENS	SPEC	K_{CPE-L}	P_{CPE}	P_L	K_{CPR-L}	K_{LL}
Parent											
CD–ODD	8	0	2	21	100	91	.84	25.8	32.3	.84	.52
ADHD	3	1	1	27	75	96	.71	12.5	12.5	.71	.75
EXT	9	1	3	18	90	86	.72	32.3	38.7	.72	.58
OAD–SAD	4	3	2	23	57	92	.52	21.9	18.8	.52	.58
MDD–DYS	5	2	2	21	71	91	.63	23.3	23.3	.62	.48
INT	7	3	3	17	75	70	.55	33.3	33.3	.54	.49
Child/adolescent											
CD–ODD	3	0	1	28	100	97	.84	9.4	12.5	.84	.48
ADHD	2	0	1	28	100	97	.78	6.5	9.7	.26	.34
EXT	3	0	2	26	100	93	.72	9.7	16.1	.52	.42
OAD–SAD	1	4	1	25	20	96	.21	16.1	6.5	.27	.40
MDD–DYS	7	1	2	19	88	90	.75	27.6	31.0	.84	.36
INT	6	3	2	17	67	89	.58	32.1	28.6	.74	.32

[a]*Note*: ++, both assessments positive; +−, child psychiatrist positive and lay interviewer negative; −+, child psychiatrist negative and lay interviewer positive; −−, both interviewers negative. SENS is sensitivity; SPEC is specificity; these data compare the DICA-R administered by a child psychiatrist and enriched by clinical judgment vs. a lay-administered DICA-R. K_{CPE-L} is the kappa estimate between child psychiatrist (enriched DICA-R) and lay interviewer administrations of the DICA-R. K_{CPR-L} is the kappa estimate between child psychiatrist (regular DICA-R) and lay interviewer administrations of the DICA-R. P_{CPE} prevalence based on child psychiatrist (enriched DICA-R). P_L prevalence based on lay interviewers. K_{LL} is the kappa estimate between lay interviewer administration of the DICA-R (n = 191) without weighting the responses back to the original sample. See note, Table I. CD, conduct disorder; ODD, oppositional defiant disorder; CD–ODD, CD and ODD combined; ADHD, attention-deficit hyperactivity disorder; EXT, one or more of CD, ODD, ADHD; OAD, overanxious disorder; SAD, separation anxiety disorder; OAD–SAD, OAD and SAD combined; MDD, major depressive disorder; DYS, dysthymia; MDD–DYS, MDD and DYS combined; INT, one or more of OAD, SAD, MDD, DYS. DICA-R, Diagnostic Interview for Children and Adolescents—Revised.

adolescent drop from .78 to .26. The distribution of classifications yielding the kappa estimate of .26 is (++) 1, (+−) 2, (−+) 2 and (−−) 26. It is noteworthy that none of the prevalences of disorders differs by more than 10% for child psychiatrists vs. lay interviewers, suggesting that under- or overreporting of disorders by lay interviewers does not seem to be occurring.

DISCUSSION

This study provides information for evaluating the usefulness of the DICA-R for classifying childhood psychiatric disorders in the general population according to DSM-III-R criteria. The information focuses on prevalence, test–retest reliability, agreement between parent and child/adolescent respondents, and agreement between trained lay interviewers and child psychiatrist classifications of disorders using the DICA-R.

Prevalence

A comparison of prevalence findings in this study with estimates generated from other community surveys (e.g., Brandenberg, Friedman, & Silver, 1990; Costello, 1989) suggests that some categories of disorders may be overidentified. For example, the prevalence of parent-identified ODD was 15.4% (6- to 11-year-old age group) and 10.4% (12- to 16-year-old age group). The high prevalence of ODD could be corrected by reclassifying those with mild and/or moderate ODD as nondisordered. However, such a simple solution is not available to OAD and DYS where prevalences are very large particularly when based on interviews with adolescents aged 12 to 16.

Bringing the prevalences "into line" with other community surveys could be accomplished in a number of ways: by changing the threshold for classifying disorder so that more or fewer symptoms are needed for positive identification, by changing question wording to make it more or less difficult to obtain positive responses for individual symptoms, or by invoking or not invoking severity criteria such as the need for evidence of academic or social impairment to qualify a child as disordered. The need for such strategies to obtain "reasonable" prevalences attests to our lack of knowledge about where and how to draw the boundary between normal and abnormal, disorder and nondisorder (see Boyle, 1991). This problem of how to define useful boundaries goes beyond whether or not trained lay interviewers or child psychiatrists administer the interviews. In our study, the prevalences of disorder obtained by trained lay interviewers and child psy-

chiatrists were very similar. Allowing the child psychiatrists to use clinical probing, after completing each section of the interview following the standards imposed on the lay interviewers, had little or no effect on the frequency of classification (data not shown). Although the sample available for examining agreement between trained lay interviewers and child psychiatrists was small, the lack of difference in prevalence suggests that no undue inflation in prevalence was being caused by large numbers of false positives being identified by trained lay interviewers.

Using information on prevalence to evaluate the usefulness of the DICA-R or any other structured interview is a tricky business. Inherent differences in sampling frames, the composition of the sample obtained, and assessment methodologies make it difficult to use the results from different general population studies to determine guidelines for what constitutes a reasonable prevalence of a disorder (see Brandenburg et al., 1990; Costello, 1989). For example, in our study, subjects were drawn from a single, industrialized urban setting; the sample size (N = 251) was small and attrition reached 40%. In addition, the DICA-R has not been used in a general population sample to estimate prevalences of disorders, making it difficult to set expectations for prevalences. In two clinic-based studies, the DICA (earlier version of the DICA-R) consistently overidentified children as disordered compared to (1) hospital discharge diagnoses (Welner, Reich, Herjanic, Jung, & Amado, 1987) and (2) the Schedule for Affective Disorders and Schizophrenia (Carlson et al., 1987).

Test-Retest Reliability

Test–retest reliability of the DICA-R, based on kappa, showed considerable variability attributable to respondent and type of disorder. In general, parent assessments yielded more reliable classifications than child (age 6 to 11) or adolescent (age 12 to 16) assessments. Child and adolescent assessments of the externalizing disorders (CD, ODD, ADHD), achieved similar reliability in the "fair" range (i.e., .21 to .40: Landis & Koch, 1977); however, child assessments of the internalizing disorders (OAD, SAD, MDD, DYS) were too unreliable to be useful, compared with adolescent assessments of the same disorders which yielded kappa estimates in the "moderate" range (i.e., .41 to .60: Landis & Koch, 1977).

With some important exceptions, this general pattern resembles the findings of a study reported by Edelbrock, Costello, Dulcan, Kalas, and Conover (1985). This team examined the test–retest reliability of symptom scores obtained using the DISC in 242 children aged 6 to 18 and their parents referred to mental health services. That study found that parents

were generally more reliable than children/adolescents in reporting child/adolescent symptoms. However, test–retest reliabilities showed an opposite age pattern for parent and child. The reliability of the child's report increased with age while the reliability of the parent's report decreased. In our study, the reliability of respondent assessments depended on both type of disorder and on age. Parent assessments of the externalizing disorders (CD, ODD, ADHD) were more reliable in the older than the younger age group but this was not the case for parent assessments of the internalizing disorders (OAD, SAD, MDD, DYS). In contrast, adolescent assessments of the internalizing disorders were more reliable than child assessments of the same disorders but this was not true for adolescent assessments of the externalizing disorders.

Information on test–retest reliabilities of structured interviews such as the DICA and DISC for use in the general population is virtually nonexistent. A handful of studies has examined this issue among children using mental health services (Costello, Edelbrock, Dulcan, Kalas & Klaric, 1984; Shaffer et al., 1988; Sylvester, Hyde & Reichler, 1987) and produced kappa estimates of test–retest reliability through the entire range of .00 to 1.00. The findings from these studies are unlikely to be applicable to general population samples. In particular, the magnitude of kappa is sensitive to the marginal distributions or base rates for disorders (Soeken & Prescott, 1986) and these will be very different between clinic and general population samples. In addition, the experiences of clinic referral and engagement are very likely to affect respondents' perceptions of problems. It seems reasonable to anticipate that assessment data obtained in clinic settings would appear to be more reliable because respondent perceptions have coalesced around certain problem configurations that have received clinical attention. In addition, the extreme severity of problems dealt with in clinic settings may make it easier for everyone to agree on their existence. In general population samples, respondent perceptions are not subject to the same influences that accompany clinic attendance. This makes it imperative to be cautious about generalizing clinic findings to the general population.

Parent vs. Child/Adolescent

Examination of agreement between parent and child/adolescent assessments generated by the DICA-R suggests that overall levels of agreement are low. These results are consistent with most studies that have examined this issue for structured interviews in clinic samples (Costello et al., 1984; Shaffer et al., 1988; Sylvester et al., 1987) and for checklist self-reports in clinic and general population samples (see Achenbach,

McConaughy, & Howell, 1987). The exception to these findings is one study involving the administration of the DICA to 27 psychiatric inpatients, 7 to 17 years of age, and their parents (Welner et al., 1987). This study generated much higher estimates of parent–child agreement based on kappas from .52 to .80. However, these estimates may be inflated because the sample consisted of inpatients. As noted in the previous section, clinical attention may have focused the perceptions of respondents. In addition, the severity of the presenting problems may have made it relatively easy to obtain agreement on their existence. It is noteworthy as well that the small sample used (N = 27) would yield very wide confidence intervals around the estimates.

Child Psychiatrist vs. Trained Lay Interviewer

Data on the strength of agreement between trained lay interviewers and child psychiatrists administering the DICA-R suggest concordance in the "substantial" range (i.e., .61 to .80: Landis & Koch, 1977) except for the assessment of OAD–SAD among children and adolescents which generated a kappa estimate of .21. The high levels of agreement quantified by kappa are reflected in high levels of sensitivity and specificity. Interestingly, agreement levels between trained lay interviewers and child psychiatrists are higher than agreement levels between lay interviewers, particularly when interviewing children and adolescents (Table IV). This finding could arise if child psychiatrists administering a structured interview obtained more reliable data than trained interviewers. If this were true, then less random error would be available to attenuate agreement between a comparison of child psychiatrists vs. trained lay interviewers than between a comparison of trained lay interviewers vs. trained lay interviewers. If this finding is not due to higher levels of reliability among child psychiatrists than trained lay interviewers, other explanations are not readily apparent. The order in which the interviews were done (lay vs. psychiatrist) was randomized. Moreover, assessment data obtained by the interviewers were not shared.

The only other study to compare lay vs. clinician administration of a structured interview for children was reported by Costello et al. (1984). Using an earlier version of the DISC, Costello and colleagues obtained kappa estimates from .48 to .63 for parent responders. In clinic-based studies, agreement between earlier versions of the DICA administered by clinicians and hospital diagnoses yielded kappa estimates varying from .05 to .75 (Carlson et al., 1987) and from .03 to .50 (Welner et al., 1987).

To our knowledge, this is the first study to provide evaluation data on the use of the DICA-R in a general population sample. The study possessed several important strengths, notably the following: the use of probability sampling in a large unselected population of children; a large sample yielding fairly precise statistical estimates; independent assessments—all interviews were done by different interviewers with no prior knowledge of the respondent's status or functioning; an interview schedule design constructed to permit a variety of agreement studies; and a random sequencing of interviews among the different types of interviewers to prevent the tendency of respondents who report fewer symptoms at second interviews from systematically distorting the findings of one particular group (e.g., child psychiatrists). There were also some weaknesses: sample losses at screening, stage 1 and stage 2, when compounded, meant that only 60.2% (1751/2219 × 251/329) of the original sample participated fully in the study; in addition, there were too few interviews done by different child psychiatrists to examine interclinician agreement on classifications of disorders.

In summary, this study has examined the measurement potential of the DICA-R in a general population sample. Prevalence data indicate that some categories of disorders may have been overidentified because they occurred more frequently than might be expected in a general population sample. Agreement analyses indicate that differences in reliability occurred depending on type of respondent, type of disorder, and child age. Test–retest reliability estimates show that child (ages 6 to 11) reports of the internalizing disorders contained too much error to be useful while their reports of externalizing disorders exhibited enough reliability to demonstrate measurement potential. With the exception of SAD and MDD, parent assessments of adolescent disorders were very reliable. The low levels of agreement between parent and child/adolescent respondents observed in this study are typical of most investigations that focus on this issue. The very high levels of agreement between trained lay interviewers and child psychiatrists suggest that trained lay interviewers using a structured interview are quite capable of replicating data that would be obtained by child psychiatrists using a structured interview in a general population sample.

In our view, the DICA-R shows considerable measurement potential. There are, however, several categories of disorders that might be improved. It is our understanding that revisions to the DICA-R since 1989 have seen mandatory probes added to selective categories of disorders to ensure that the symptoms are abnormal. This would have the effect of reducing the prevalences of certain disorders (e.g., oppositional defiant disorder and overanxious disorder: Wendy Reich, personal communication, October 26, 1992). In addition, it is important to acknowledge that this study and virtually all studies that have evaluated structured interviews for classifying

psychiatric disorder have focused on issues of prevalence and reliability of measurement (e.g., see Malgady, Rogler, & Tryon, 1992). Future research on the DICA-R and other child interviews must also address the vexing issue of validity. Do these instruments yield classifications that are useful for understanding the epidemiology of childhood psychopathology? To accomplish this goal we must demonstrate that the informational value of the classifications goes beyond symptom content to reveal important data about etiology, current functioning, and natural history. Studies that focus on issues of prevalence and reliability of measurement are important but limited in what they can tell us about the usefulness of structured interviews for classifying childhood disorders.

REFERENCES

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213-232.

Boyle, M. H. (1991). Children's mental health issues. In L. C. Johnson & D. Barnhorst (Eds.), Children, families and public policy in the 90s (pp. 73-105). Toronto: Thompson Educational Publishing Inc.

Boyle, M. H., Offord, D. R., Hofmann, H. G., Catlin, G. P., Byles, J. A., Cadman, D. T., Crawford, J. W., Links, P. S., Rae-Grant, N. I., & Szatmari, P. (1987). Ontario Child Health Study: I. Methodology. *Archives of General Psychiatry, 44*, 826-831.

Boyle, M. H., Offord, D. R., Racine, Y., Fleming, J. E., Szatmari, P., & Sanford, M. (1993). Evaluation of the revised Ontario Child Health Study Scales. *Journal of Child Psychology and Psychiatry, 34*, 189-213.

Brandenburg, N. A., Friedman, R. M., & Silver, S. E. (1990). The epidemiology of childhood psychiatric disorders: Prevalence findings from recent studies. *Journal of the American Academy of Child and Adolescent Psychiatry, 29*, 76-83.

Carlson, G. A., Kashani, J. H., De Fatima, Thomas, M., Vaidya, A., & Daniel, A. E. (1987). Comparison of two structured interviews on a psychiatrically hospitalized population of children. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 645-648.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Costello, A. J., Edelbrock, C., Dulcan, R. K., Kalas, R., & Klaric, S. H. (1984). *Development and testing of the NIMH Diagnostic Interview Schedule for Children in a clinic population. Final report* (Contract No. RFP-DP-81-0027). Rockville, MD: Center for Epidemiologic Studies, National Institute of Mental Health.

Costello, E. J. (1989). Developments in child psychiatric epidemiology. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*, 836-841.

Edelbrock, C., & Costello, A. J. (1988). Structured psychiatric interview for children. In M. Rutter, H. Tuma, & I. Lann (Eds.), *Assessment and diagnosis in child psychopathology* (pp. 87-112). New York: Guilford Press.

Edelbrock, C., Costello, A. J., Dulcan, M. K., Kalas, R., & Conover, N. C. (1985). Age differences in the reliability of the psychiatric interview of the child. *Child Development, 56*, 265-275.

Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Jr., Winokur, G., & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry, 26*, 57-63.

Herjanic, B., & Campbell, W. (1977). Differentiating psychiatrically disturbed children on the basis of a structured interview. *Journal of Abnormal Child Psychology, 5*, 127-134.

Herjanic, B., & Reich, W. (1982). Development of a structured psychiatric interview for children: Agreement between child and parent on individual symptoms. *Journal of Abnormal Child Psychology, 10*, 307-324.

Hodges, K. (1993). Structured interviews for assessing children. *Journal of Child Psychology and Psychiatry, 34*, 49-68.

Landis, J. R., & Koch, G. G. (1977). The measure of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Malgady, R. G., Rogler, L. H., & Tryon, W. W. (1992). Issues of validity in the Diagnostic Interview Schedule. *Journal of Psychiatric Research, 26*, 59-67.

Reich, W., & Welner, Z. (1988). *Revised version of the Diagnostic Interview for Children and Adolescents (DICA-R)*. St. Louis, MO: Department of Psychiatry, Washington University School of Medicine.

Robins, L., Helzer, J. E., Croughan, J., & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule. *Archives of General Psychiatry, 38*, 382-389.

Shaffer, D., Schwab-Stone, M., Fisher, P., Davies, M., Piacentini, J., & Gioia, P. (1988). *A revised version of the Diagnostic Interview Schedule for Children (DISC-R)*. New York: New York State Psychiatric Institute, Columbia University College of Physicians and Surgeons, Division of Child Psychiatry.

Soeken, K. L., & Prescott, P. A. (1986). Issues in the use of kappa to estimate reliability. *Medical Care, 24*, 733-741.

Sylvester, C. E., Hyde, T. S., & Reichler, R. J. (1987). The Diagnostic Interview for Children and Personality Inventory for Children in studies of children at risk for anxiety disorders or depression. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 668-675.

Welner, Z., Reich, W., Herjanic, B., Jung, K. G., & Amado, H. (1987). Reliability, validity, and parent-child agreement studies of the Diagnostic Interview for Children and Adolescents (DICA). *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 649-653.