Diagnostic Agreement between Clinicians and the Diagnostic Interview for Children and Adolescents—DICA-R—in an Outpatient Sample

L. Ezpeleta, N. de la Osa, J. M. Doménech, J. B. Navarro, and J. M. Losilla

Universitat Autònoma de Barcelona, Spain

J. Júdez

Consorci Hospitalari Parc Taulí, Sabadell, Spain

The paper examines the diagnostic agreement between clinicians and the Diagnostic Interview for Children and Adolescents. One hundred and thirty-seven outpatients—children and adolescents, and their parents—were diagnosed independently following DSM–III–R criteria by clinicians and by the DICA-R. The diagnostic concordance between clinicians and DICA-R ranged from low to moderate in the majority of the categories. The only exception was Conduct Disorder. Differences depending on the informant and the quality of the information (cognitive vs. observable) were observed. Combining the information from the child/adolescent and their parents ameliorates the concordance. The reasons for the scanty agreement found could be due to the fact that clinicians and structured interviews differ in *what* they evaluate (conditions on which they focus), *how* they evaluate (strictness in the criteria application, use of different informants and different information, etc.), and *when* they evaluate (present condition vs. lifespan). After analysing the pros and cons of both, the use of structured interviews is advisable for research purposes. There is a clear need for a variety of informants, and the combination of information from different sources is recommended, depending on the age of the children and the type of disorder.

Keywords: Diagnosis, child and adolescent psychopathology, structured interviews, outpatients, DICA.

Abbreviations: DICA: Diagnostic Interview for Children and Adolescents; DISC: Diagnostic Interview Schedule for Children; MDD: Major depressive disorder; OAD: Overanxious disorder; ODD: Oppositional defiant disorder; OCD: Obsessive-compulsive disorder; PTSD: Post-traumatic stress disorder; SAD: Separation anxiety disorder.

Introduction

The poor reliability of psychiatric diagnosis in adults (Anthony et al., 1985; Costello, 1992; Helzer et al., 1985; Matarazzo, 1983; Robins, 1985) and in children (Rey, Plapp, & Stewart, 1989; Vitiello, Malone, Buschle, Delaney, & Behar, 1990; Weinstein, Stone, Noam, Grimes, & Schwab-Stone, 1989) is widely documented in the literature.

The presence or absence of a disorder may not always be identified by a psychiatric evaluation (Helzer et al., 1985; Robins, 1985). Researchers in the field of psychopathology lack a gold standard to establish the presence of morbidity (Lane, Pollard, & Cox, 1990; Vailliant & Schnurr, 1988).

In recent decades, differences among clinicians in their application of diagnostic criteria have been reduced considerably through the application of stricter oper-ational diagnostic criteria (Helzer, Robins, Taibleson, Wooddruff, & Reich, 1977). However, some differences still remain, and there are a variety of reasons for them. Helzer et al. (1985) and Rey et al. (1989) suggest that, lacking clear diagnostic guidelines, clinicians give different judgements to apparently identical data because ambiguities are still present (criterion variance). Alternatively, informants can report different information at different times (occasion variance). Anthony et al. (1985) suggest that there are deficiencies in information gathering (information variance) or inadequate use of the information on which the diagnosis is based, including incomplete criterion coverage. Shrout, Spitzer, and Fleiss (1987, p.172) add "uncareful, inconsistent or incompetent inference on the part of the diagnostician" as another source of variability.

Finally, Cohen, O'Connor, Lewis, Velez, and Malachowski (1987a) list some design factors that could influence the extent of agreement between clinicans: agreement is higher when the population studied has severe psychopathology (inpatients); when the interval between interviews is short; if neither of the clinicians assess in the initial interview; when they use the same

Table 1

Criterion validity of child and adolescent structured diagnostic interviews

			Study								
	1	2	3						4		5	6
			Parent			Child			Parent	Child		
Disorder	k	k	k	Se	Sp	k	Se	Sp	Se	Se	k	k
ADHD		.09	.53	57	97	.48	1	94			.22	.50
CD		.09	.60	62	95	.21	50	83			.06	.43
ODD	.09	.09	.51	89	84	.28	1	83			.04	
MDD	0	.17	.36	42	92	.39	38	95	55	18	.19	
AFF		.09										.52
ANX		.03									.07	.03
SAD	.11											
OAD	—											
OCD									53	65		
Anorexia									44	78		
Bulimia									63	75		
Psicosis									83	33		
SUBS									75	75		
TICS									100	67		

						Study											
	7									8	9		10					
	DICA-C			DICA-P			K–SADS				Parent	Child	Parent			Child		
Disorder	k	Se	Sp	k	Se	Sp	k	Se	Sp	k	k	k	k	Se	Sp	k	Se	Sp
ADHD	.30	75	50	.18	1	22	.56	1	61	.48	.60	.36	.42	.57	.92	.36	.36	.97
CD	.75	84	88	.60	84	76	.69	84	82	.33	.60	.50	.15	.14	.98	.15	.20	.97
ODD	.31	42	87	.66	85	87	.16	14	1				.58	.63	.96	.16	.13	.98
MDD													.33	.56	.92	.54	.67	.96
AFF	.48	81	68	.43	81	63	.58	81	79	.54	.75	.52	.37	.47	.92	.43	.41	.96
ANX											.51	.37	.29	.51	.80	.38	.54	.85
SAD	.15	1	75	.50	1	1	.44	1	93		.44	.49	.07	.13	.94	−.02	.11	.97
OAD	.28	1	75	.05	0	96	.35	50	92		.09	0						

k: kappa coefficient; Se: sensitivity; Sp: specificity; ADHD: attention deficit/hiperactivity disorder; CD: conduct disorder; ODD: oppositional defiant disorder; MDD: major depressive disorder; DD: dysthymic disorder; AFF: general category of affective disorders; ANX: general category of anxiety disorders; SAD: separation anxiety disorder; OAD: Overanxious disorder; OCD: obsessive-compulsive disorder; Anorexia: anorexia nervosa; Bulimia: bulimia nervosa; SUBS: substance abuse; TICS: Tourette disorder.

Study 1: Cohen et al. (1987a). Diagnostic Interview Schedule for Children (DISC). 100 community aged 9 to 12. Criterion K-SADS.

Study 2: Weinstein et al. (1989). DISC-C. 163 inpatients aged 12 to 16. Clinicians' diagnoses.

Study 3: Piacentini et al. (1993). DISC-R. 74 outpatients aged 11 to 17. Criterion structured interview CAF.

Study 4: Fisher et al. (1993). DISC. 75 inpatients and outpatients aged 8 to 19. Clinician diagnosis.

Study 5: Aronen et al. (1993). DISC-C. 163 inpatients aged 12 to 16. Clinicians' discharge.

Study 6: Welner et al. (1987). Diagnostic Inteview for Children and Adolescents (DICA-C). 27 inpatients aged 7 to 17. Discharge diagnosis.

Study 7: Carlson et al. (1987). DICA-C/DICA-P/Schedule for Affective Disorders and Schizophrenia (K-SADS). 30 inpatients aged 7 to 12. Best estimate.

Study 8: Vitiello et al. (1990). DICA. 46 inpatients aged 6 to 13. Discharge diagnoses.

Study 9: Hodges et al. (1987). CAS. 30 clinic referrals aged 7 to 16. Criterion K–SADS.

Study 10: Rubio-Stipec at al. (1994). DISC.2. 170 community and clinical aged 9 to 17. Best-estimate clinical diagnosis.

assessment instruments; when they have the same training and backgrounds; and when they rely on the same sources of information. Poorest agreement occurs when the design includes epidemiological samples; when there are longer intervals between evaluations; if the clinicians being compared use different assessment instruments; or if they use different diagnostic methods and different sources of information.

Diagnostic determinations in psychology and psy-chiatry rely mainly on the results of a clinical interview. In the last 15 years, different structured interviews have been developed for children and adolescents as researchers recognised the need to reduce the bias of unstructured interviews. This has generated more efficient decision-making rules for determining diagnoses (Angold, 1994).

Table 1 shows a summary of the concordance between parent or child (as the course of information) and a clinician, for a range of common diagnoses derived from

the most frequently used interviews. In general, the agreement ranges from low to moderate (Apter, Orvaschel, Laseg, Moses, & Tyano, 1989; Aronen, Noam, & Weinstein, 1993; Carlson, Kashani, Thomas, Vaidya, & Daniel, 1987; Cohen et al., 1987a; Fisher et al., 1993; Hodges, McKnew, Burbach, & Roebuck, 1987; Piacentini et al., 1993; Rubio-Stipec et al., 1994; Vitiello et al., 1990; Weinstein, Stone, Noam, Grimes, & Schwab-Stone, 1989; Welner, Reich, Herjanic, Jung, & Amado, 1987). There are variations in diagnostic concordance depending on the informant, the sample studied, the different diagnostic categories, the status of the diagnosis (probable or definite), and the schedule used.

We adapted the Diagnostic Interview for Children and Adolescents—Revised (DICA-R; Reich, Shayka, & Taibleson, 1991) for use with Spanish children and adolescents. In this paper we address the issue of diagnostic agreement between clinicians and the DICA-R for parents and children interviewed separately, and for diagnoses made by combining the information obtained from both sources. It is the first study using the DICA-R to address this topic in outpatients.

Method

Subjects

The sample consisted of 137 child and adolescent outpatients and their parents. The younger children ranged in age from 6 to 12 years ($N = 69$; 50 %), with 44 boys (64 %) and 25 girls (36 %). The adolescents ranged in age from 13 to 17 ($N = 68$; 50 %), with 21 boys (31 %) and 47 girls (69 %). The mean age of the total sample was 12.1 years ($SD = 3.15$). They were recruited from three Primary Mental Health Care Centres for children and adolescents. Caucasian families comprised 98.5 % of the sample and the remaining 1.5 % were of other races. Hollingshead's socioeconomic status (Hollingshead, 1975) for children was distributed as follows: class I: 3.3 %; class II: 1.6 %; class III: 9.9 %; class IV: 39.3 %; class V: 45.9 %. For adolescents the distribution was: class I: 3.0 %; class II: 13.6 %; class III: 13.6 %; class IV: 24.2 %; class V: 45.6 %. Children with evidence of mental retardation were not included.

Measures

The revised version of the Diagnostic Interview for Children and Adolescents (DICA-R 7.2) was administered to participating parents and child/adolescent dyads.

The DICA was the first structured diagnostic interview for children and adolescents. It was developed by Herjanic and Campbell (1977) and Herjanic and Reich (1982). The DICA-R is a revised version that follows DSM-III-R criteria for making diagnoses (American Psychiatric Association, 1987). The interview covers the 6- to 17-year-old span in three versions: one for children aged 6 to 12, another for adolescents aged 13 to 17, and a single-parent version for both groups. The instrument is totally structured, and organised in such a way as to explore the presence or absence of each one of a range of diagnostic syndromes. The interview lasts between 60 and 90 minutes. DICA and DICA-R have high between-interviewer and test-retest reliability (Herjanic & Reich, 1982; de la Osa, Ezpeleta, Doménech, Navarro, & Losilla, 1996; Welner et al., 1987).

Overall agreement between parents and children interviewed with the schedule is moderate (Herjanic & Reich, 1982; Kashani, Orvaschel, Burk, & Reid, 1985; Reich, Herjanic, Welner, & Gandhy, 1982; Sylvester, Hyde, & Reichler, 1987; Welner et al., 1987), and best for externalising pathology. The DICA has been found to discriminate between paediatric and

psychiatric samples and has a moderate correlation with other measures of child psychopathology (including clinician diagnoses) (Carlson et al., 1987; Herjanic & Campbell, 1977; Sylvester et al., 1987; Welner at al., 1987).

Participating clinicians were asked to fill in a checklist specifically developed for the study, containing the DSM-III-R symptomatology for all disorders included in the DICA-R. For each symptom and for each disorder, clinicians had to indicate their presence or absence on a two-point scale. Frequency and intensity of symptoms were not specifically coded.

Procedure

The DICA-R was translated from English into Spanish by a bilingual Ph.D. psychologist. It was then revised by a child psychiatrist and the content and meaning of each item were checked with the main author of the interview. Questions about race, about credit card theft in the conduct disorder section, and the beverage list in the alcohol use section, were adapted to be applicable to Spanish juveniles.

The structural interviews were done for each member of the dyad by different trained interviewers, who were blind to the diagnosis given by the clinician. Parents and children were interviewed on the same day and at the same time. The interview team consisted of 17 psychology students and 2 Ph.D. psychologists with clinical experience. Their training consisted of home study of the interview, practice interviews among trainees, coding of taped interviews, and field interviews. Interviewers were allowed to use the instrument on their own when they had been practising with it for at least 3 months. The average agreement between interviewer and another member of the team—the observer—in five interviews was equal or superior to a kappa value of .80.

In all cases, clinicians had the first contact with the child/ family. They had to provide cases for the DICA-R interview as close to the clinic intake as possible to avoid treatment effects. Clinic cases were included in the study if the child/adolescent was considered by the clinician to be suffering from any of the disorders covered in the DICA-R.

Collaboration by the clinicians was limited by the pressure produced by long waiting lists in the public health services and the stressful nature of the work. For this reason, it was not possible to introduce diagnostic guidelines for them that were different from their usual practice. They were encouraged to use all available information to complete the checklist and to derive DSM-III-R diagnoses at their own pace. The same clinician assessed the child/adolescent and the parent. All the clinicians participating (Ph.D. and M.D.) were familiar with DSM-III-R criteria and used them in their regular practice.

In order to examine the extent to which interviewers and clinicians agreed with each other, percentage agreement and kappa (Cohen, 1960) were used. Kappa may be affected by low frequency of categories, in this case rare diagnoses. Sensitivity and specificity of the interviews were also measured to indicate the proportion of cases with each diagnosis detected, in comparison with the interview by the individual clinician. However, as we mentioned previously, clinical judgement does not represent a "gold standard". McNemar's Exact Test for repeated measures was used to assess the level of agreement between clinician and interview (present/absent) for each. Paired t-tests compared mean differences in the number of diagnoses for each subject.

In this study we used the DICA-R algorithms relevant to children, adolescent, and parent interviews separately. For the phobia section, the DICA-R definition was altered and impairment questions were included in the algorithm, since the diagnosis of phobia was greatly overestimated (Gratacós, 1993).

Furthermore, we created "combined diagnoses", in which a disorder was considered to be present when the parent or the

child/adolescent indicated the presence of DSM–III–R symptoms relevant to a diagnosis. In this instance, if either of the two informants indicated the presence of a relevant symptom it was counted as being present.

Results

Agreement between Clinicians and the DICA–R Child Version

The diagnostic agreement between clinicians and the DICA–R–C is presented in Table 2. Specificity was high for all the DSM–III–R disorders except enuresis. Sensitivity was generally low except for conduct disorder, anorexia, and enuresis.

There were significant differences in rates of diagnosis of attention deficit/hyperactivity disorder (ADHD) and enuresis between the DICA–R–C and the clinicians. The clinician made relatively more diagnoses of ADHD, whereas interviews with the DICA–R-C resulted in more diagnoses of enuresis.

Overall, there were no significant differences in the mean number of disorders identified by the DICA–R–C or the clinicians.

Agreement between Clinician and the DICA–R Adolescent Version

Table 2 presents the results for agreement in diagnostic terms between DICA–R–A and clinicians. Specificity was high in all the categories. Sensitivity was good for conduct disorder and post-traumatic stress disorder. In eight categories of diagnosis, about half of the cases diagnosed by the clinician were also diagnosed by the structured interview. Phobias, obsessive-compulsive disorders, and bulimia showed less diagnostic concordance. In general, the DICA–R–A gave rise to a higher apparent prevalence of psychiatric disorder than did the clinicians, with the difference in apparent prevalence reaching significance for separation anxiety disorder, overanxious disorder, and enuresis. The number of disorders identified was significantly higher with the use of the DICA–R–A than by the clinicians, with a mean difference of 1.06 ($P =$ .001; 95% confidence interval: 0.5 to 1.6).

Agreement between Clinician and the DICA–R Parent Version for Children 6–12 Years of Age

There were no significant differences between clinicians' diagnoses and those made by the DICA–R–P for

Table 2
Diagnostic Agreement between Clinician and DICA–R–C and DICA–R–A

| Diagnostic categories | | N | Prevalence | | Agreement | | | | DICA vs. Clin | |
			DICA	Clin	Se	Sp	Agree	κ	Diff	95% CI
ADHD	C	65	7.7	2.3	14	96	69	.12	−24.6	−30.0 to −11.3
	A	65	7.7	6.2	50	95	92	.40	1.5	−5.4 to 6.9
ODD	C	64	10.9	6.3	25	90	86	.11	4.7⁻	−5.7 to 12.0
	A	67	25.4	17.9	50	80	75	.25	7.5	−5.9 to 18.2
CD	C	65	4.6	3.1	100	98	98	.79	1.5	−1.5 to 1.5
	A	66	16.7	7.6	80	89	88	.44	9.1	−0.6 to 12.0
MDD	C	64	6.3	7.8	40	97	92	.40	−1.6	−7.0 to 5.5
	A	67	26.9	31.3	48	83	72	.31	−3.1	−16.9 to 9.4
DD	C	64	3.1	1.6	—	97	95	−.02	1.6	−3.8 to 4.6
	A	66	15.2	24.2	31	90	76	.24	−9.1	−18.9 to 4.2
SAD	C	65	24.6	13.8	56	80	77	.27	10.8	−2.4 to 19.5
	A	65	23.1	3.1	50	78	77	.07	*20.0*	8.3 to 23.0
OAD	C	64	12.5	6.3	50	90	87	.27	6.3	−3.8 to 11.7
	A	65	29.2	9.2	67	75	74	.21	*20.0*	7.1 to 25.4
PHOBIA	C	63	7.9	12.7	13	93	82	.06	−4.8	−13.6 to 6.7
	A	63	11.1	11.1	29	91	84	.20	0	
OCD	C	65	1.5	7.7	20	100	94	.32	−6.2	−6.2 to 1.3
	A	64	1.6	7.8	20	100	94	.31	−6.3	−6.3 to 1.3
PTSD	C	64	3.1	4.7	—	97	92	−.04	−1.6	−7.0 to 5.5
	A	63	7.9	3.2	100	95	95	.55	4.8	−2.0 to 4.8
Anorexia	C	63	1.6	1.6	100	100	100	1.00	0	
	A	62	9.7	9.7	50	95	90	.45	0	
Bulimia	C	62	—	3.2	—	100	97	—	−3.2	−3.2 to 2.2
	A	64	6.3	9.4	33	97	91	.35	−3.1	−8.6 to 5.2
Enuresis	C	65	36.9	23.1	80	76	77	.46	*13.8*	0.9 to 21.1
	A	64	21.9	3.1	50	79	78	.07	*18.8*	7.1 to 21.8
Encopresis	C	64	7.8	15.6	20	94	83	.18	−7.8	−15.1 to 3.8
	A	63	7.9	4.8	67	95	94	.47	3.2	−3.9 to 6.3

C: DICA–R–C; A: DICA–R–A; Se: sensitivity; Sp: specificity; Clin: clinician; Agree: % agreement; Diff: % difference; CI: confidence interval.

For other abbreviations see Table 1. Differences in *italic* are significant.

Table 3
Diagnostic Agreement between Clinician and DICA-R-P

| Diagnostic categories | | N | Prevalence | | Agreement | | | | DICA vs. Clin | |
			DICA	Clin	Se	Sp	Agree	κ	Diff	95% CI
ADHD	C	65	23.1	32.3	53	91	78	.47	−9.2	−17.9 to 3.5
	A	65	12.3	6.2	75	92	91	.46	6.1	−2.6 to 9.1
ODD	C	65	15.4	6.2	50	87	85	.36	9.2	−1.7 to 14.6
	A	66	10.6	18.2	42	96	94	.45	−7.5	−12.9 to 2.7
CD	C	65	3.1	3.1	100	100	100	1.00	0	
	A	66	7.6	7.6	60	97		.57	0	
MDD	C	66	3.0	7.6	40	100	95	.55	−4.5	−4.5 to 1.9
	A	65	12.3	30.8	20	91	69	.13	−18.5	−27.2 to −3.9
DD	C	66	6.1	1.5	—	94	92	−.02	−4.6	−3.3 to 7.5
	A	65	12.3	21.5	29	92	78	.25	−9.2	−17.9 to 3.5
SAD	C	66	13.6	13.6	67	95	91	.61	0	
	A	64	9.4	3.1	50	92	91	.21	6.2	−2.6 to 9.3
OAD	C	65	9.2	6.2	25	92	88	.14	3.1	−6.3 to 10.2
	A	64	14.1	9.4	67	91	89	.47	4.7	−4.7 to 10.1
PHOBIA	C	61	6.6	13.1	12	94	84	.09	−6.5	−14.2 to 5.0
	A	63	7.9	11.1	43	96	90	.45	−3.2	−8.7 to 5.2
OCD	C	65	0	7.7	—	100	92	—	−7.6	−7.6 to 0.3
	A	63	—	7.9	—	100	92	—	−7.9	−7.9 to 0.3
PTSD	C	65	1.5	4.6	33	100	95	.49	−3.1	−3.1 to 2.1
	A	63	1.6	3.2	50	100	98	.66	−1.6	−1.6 to 1.5
Anorexia	C	60	0	1.7	—	100	98	—	−1.7	−1.7 to 1.6
	A	62	8.1	9.7	50	96	92	.50	−1.6	−7.2 to 5.7
Bulimia	C	64	—	3.1	—	100	97	—	−3.1	−3.1 to 2.1
	A	64	3.1	9.4	17	98	91	.21	−6.3	−9.3 to 2.6
Enuresis	C	65	30.8	23.1	87	86	86	.65	7.6	−2.7 to 13.1
	A	64	21.9	3.1	100	81	81	.21	18.8	8.8 to 18.8
Encopresis	C	65	18.5	15.4	80	93	91	.67	3.1	−5.1 to 8.4
	A	63	9.5	4.8	100	95	100	.64	4.7	−1.9 to 4.7

C: DICA-R-C; A: DICA-R-A; Se: sensitivity; Sp: specificity; Clin: clinician; Agree: % agreement; Diff: % difference; CI: confidence interval.
For other abbreviations see Table 1.

children from 6 to 12 years of age, either in prevalence rates (Table 3) or for the mean number of disorders. Sensitivity was good for conduct disorder, enuresis, and encopresis. It was moderate for ADHD, oppositional defiant disorder (ODD), and separation anxiety disorder (SAD). Poor concordance was found for post-traumatic stress disorder (PTSD), obsessive-compulsive disorder (OCD), phobia, overanxious disorder (OAD) and major depressive disorder (MDD).

Agreement between Clinicians and the DICA-R Parent Version for Adolescents

The strongest agreement between clinicians and the DICA-R was achieved with adolescents, although there were significant differences in apparent prevalence rates according to clinicians and to the DICA-R-P for MDD and enuresis. Clinicians diagnosed more MDD whereas the DICA-R detected a higher prevelance of enuresis. Although clinicians recorded higher prevalence rates than the DICA-R-P, there were no significant differences in the mean number of disorders diagnosed by DICA-R-P or clinicians.

Good sensitivity was found for ADHD, enuresis, and encopresis, moderate sensitivity for conduct disorder,

SAD, OAD, PTSD, and anorexia, and poor sensitivity for ODD, MDD, phobia and bulimia.

Agreement between Clinicians and DICA-R "Combined Diagnoses"

As expected, the introduction of combined diagnoses caused an increase in reported prevalence rates and sensitivity, and a slight decrease in specificity. The consequences for kappa values were inconsistent but, in general, agreement between clinician and structural interview did not improve when the DICA-R-C, DICA-R-A, and combined diagnoses were compared (compared Table 4 with Tables 2 and 3).

The combination of information from children (6–12 years of age) and parents increased the number of identified cases of ADHD, MDD, SAD, and enuresis. Parents were better informants than children about ODD and encopresis, whereas children were better informants about OAD and anorexia. In this age group, conduct disorder is equally well reported by children and parents. The loss of specificity entailed by not combining information from the two sources of information would mainly, although not dramatically, affect ADHD, ODD, OAD, and SAD.

Table 4
Diagnostic Agreement between Clinician and Combined Diagnosis

Diagnostic categories		N	Prevalence		Agreement				DICA vs. Clin	
			DICA	Clin	Se	Sp	Agree	κ	Diff	95% CI
ADHD	C	65	36.9	32.3	76	82	80	.56	4.6	−7.4 to 14.5
	A	65	13.8	6.2	75	90	89	.41	7.7	−1.7 to 10.7
ODD	C	65	24.6	6.2	50	77	75	.11	*18.5*	5.7 to 23.8
	A	66	30.3	18.2	75	80	79	.43	12.1	−0.3 to 19.2
CD	C	66	6.1	3.0	100	97	97	.65	3.0	−2.1 to 3.0
	A	66	24.2	7.6	80	80	80	.30	*16.7*	5.5 to 19.6
MDD	C	66	7.6	7.6	60	97	94	.57	0	
	A	66	34.8	31.8	52	73	66	.25	3.0	−10.9 to 16.9
DD	C	66	9.1	1.5	—	91	89	−.03	7.6	−1.5 to 10.5
	A	66	16.7	22.7	33	88	77	.24	−6.1	−16.9 to 7.1
SAD	C	66	31.8	13.6	78	75	76	.34	*18.2*	5.7 to 23.5
	A	64	26.6	3.1	50	74	73	.05	*23.4*	11.3 to 26.5
OAD	C	65	18.5	6.2	50	84	82	.17	*12.3*	0.6 to 17.7
	A	64	32.8	9.4	83	72	73	.26	*23.4*	11.3 to 26.5
PHOBIA	C	64	14.1	12.5	25	88	80	.12	1.5	−9.4 to 11.6
	A	63	17.5	11.1	57	88	84	.36	6.4	−4.8 to 13.8
OCD	C	65	1.5	7.7	20	100	94	.32	−6.2	−6.2 to 1.3
	A	63	1.6	7.9	20	100	94	.32	−6.4	−6.4 to 1.3
PTSD	C	65	6.2	4.6	33	95	92	.25	1.5	−5.4 to 6.9
	A	63	7.9	3.2	100	95	95	.55	4.8	−2.0 to 4.8
Anorexia	C	64	1.6	1.6	100	100	100	1.00	˙0	
	A	62	11.3	9.7	67	95	92	.57	1.6	−5.7 to 7.2
Bulimia	C	65	—	3.1	—	100	97	—	−3.1	−3.1 to 2.1
	A	64	9.4	9.4	67	97	94	.63	0	
Enuresis	C	65	44.6	23.1	100	72	78	.54	*21.5*	11.6 to 21.5
	A	64	31.3	3.1	100	71	72	.13	*28.1*	17.7 to 28.1
Encopresis	C	65	21.5	15.4	80	89	88	.59	6.2	−3.7 to 11.5
	A	63	11.1	4.8	100	93	94	.57	6.4	−1.3 to 6.4

C: DICA–R–C; A: DICA–R–A; Se: sensitivity; Sp: specificity; Clin: clinician; Agree: % agreement; Diff: % difference; CI: confidence interval.
For other abbreviations see Table 1.

The agreement between clinicians and the combined diagnosis within this age group was better than between clinicians and a "separate" diagnosis (i.e. only DICA–R–C information) only for ADHD. Thus, combining the information from children and parents did not greatly improve agreement in the 6–12 year age group.

In adolescents, combining information from both sources enhanced the identification of cases with ODD, MDD, OAD, phobia, anorexia, and bulimia. Having available information from parents would permit detection of more cases of ADHD, enuresis, and encopresis, whereas the information provided by adolescents was found to be more helpful for conduct disorder and PTSD. For SAD, a similar number of cases was detected according to whether the information came from the adolescent, the parent, or from combined sources.

Specificity was reduced if combined information was lacking in ODD, anorexia, bulimia, and encopresis. Kappa values for agreement between clinicians and combined sources of information were increased for anorexia and bulimia.

Combined diagnoses using information from the DICA–R identified a significantly higher mean number of disorders than did clinicians. This was the case both for the DICA–R–C ($p = .000$: 95% confidence interval: 0.60

to 1.61) and for the DICA–R–A ($p = .000$: 95% confidence interval: 1.27 to 2.41).

Discussion

We found that the diagnostic concordance between clinicians and structured interviews with parents and children ranged from low to moderate. This is a similar result to that reported by previous authors.

Considering first the DICA–R–C, the ability of children between 6 and 17 years to report their own behaviour improves with age (Edelbrock, Costello, Dulcan, Calabro, & Kalas, 1986). Previously, Hodges et al. (1987), using the Child Assessment Schedule, Weinstein et al. (1989) and Piacentini et al. (1993), using the DISC, and Carlson et al. (1987) and Vitiello et al. (1990), using the DICA, have all found that diagnostic agreement between clinicians and their interview is low when the information on which diagnosis is based is obtained only from children. Our results are in accordance with these earlier reports.

The low sensitivity of the DICA–R–C structured interview compared to clinicians in diagnosing ADHD could be explained by the limited source of information: children deny ADHD symptoms because they are not

aware of them (Loeber et al., 1990), and clinicians diagnose using information from different sources such as children, parents, or teachers. Weinstein et al. (1989) experienced the same phenomena using the DISC–C. The DICA–R–C had good sensitivity for the diagnosis of enuresis, but only moderate specificity was achieved. The clinician probably "missed" true cases because the DICA–R–C enquires about symptomatology throughout the lifespan whereas clinicians focus on the presenting complaint. The DICA–R–C probably detected cases where there had been enuresis in the past, whereas the clinician did not explore past symptomalogy. This explanation is shared by Welner et al. (1987), who also used the DICA–C. Nevertheless, the power of the DICA–R–C to identify true negative cases is generally very good. The characteristic lack of insight in ODD children probably explains the low sensitivity obtained for this category, relative to the clinician's diagnosis.

For the diagnostic category of MDD, the slightly higher prevalence found by the clinicians could be justified, in our opinion, by the difficulty some children have in understanding the timing (onset/offset) and frequency questions of DICA–R–C. Piacentini et al. (1993) reached the same conclusion on the basis of experience with the DISC–R and proposed a revision of this section of the interview so that the item structure would be less complex. Brenton, Bergeron, and Valla (1994) ascertained that only 25% of the questions of the DISC–2 containing time concepts were understood by children aged 9 to 11.

Dysthymic disorder is a very difficult entity to diagnose, especially in children (Rey et al., 1989), because it is confounded with MDD and is rarely a cause of consultation. As we mentioned before, clinicians are accustomed to focusing only upon severe current disorders that have led to the consultation. Furthermore, we can also appreciate how, in similar disorders (MDD and dysthymic disorder) when the structure of the interview is less complex, more symptomatology is obtained. Children can give information more easily when they understand better what they are being asked.

The higher prevalences of SAD and OAD diagnosed by the DICA–R–C than were detected by clinicians could be explained by the interview's enquiry about lifespan symptomatology. Also, as Welner et al. (1987) and Weinstein et al. (1989) suggest, clinicians tend to use a hierarchical system to make diagnoses, ignoring mild disorders when more severe ones are present. Comparing our results with those from previous studies using the DICA–C (Carlson et al., 1987; Welner et al. 1987), we found clinician–interview agreement (as measured by kappa and sensitivity) to be lower for ADHD and ODD, but higher for conduct disorder, in the present study. The earlier studies were, however, carried out with inpatient samples.

Older children, in adolescence, could provide relatively more information when interviewed with the DICA–R–A. This could explain the increased prevalence of DICA–R diagnoses in the adolescent age group. Questioning systematically increased the amount of information obtained from adolescents. This phenomenon was also observed by Weinstein et al. (1989) in their use of the DISC. Some disorders were diagnosed by clinicians more often than were detected by the DICA–R–A (bulimia, OCD, MDD, dysthymic disorder). This apparent difference in prevalence could be due to the way in which the diagnostic algorithm of the structured interview operated. Although a full range of symptoms was reviewed by the structured interview, the definition for the symptom to be considered as definitely present was often very restrictive (since only severe non-normal behaviours/emotions were taken into account). Included in the definition of these disorders were frequency and duration questions, which the interviewers applied more precisely than did the clinicians, who were trying to replicate DSM–III–R criteria. Accordingly a valid diagnosis may have sometimes been "missed".

Helzer et al. (1985) and Robins (1985) highlighted the fact that clinicians did not apply DSM–III criteria in a consistent way. They often have different local traditions (Okasha, Sadek, Al-Haddad, & Abdel-Mawgoud, 1993) and use preferred diagnostic categories. Clinicians also use information that is not part of the formal diagnostic system when making diagnoses, in particular distress and impairment symptoms (Cohen, Velez, Kohn, Schwab-Stone, & Johnson, 1987b).

In general, in terms of the data obtained from the parent version of the DICA–R (DICA–R–P), the disorders presenting the least concordance between parents and children (6–12 years) were those with internalising symptomatology (e.g. all the anxiety disorders and depression). Conditions with a lot of observable behaviour such as ADHD, ODD, conduct disorder, enuresis, and encopresis showed better agreement. Parents are usually found to be better informants for externalising disorders (Achenbach, McConaughy, & Howell, 1987; Apter et al., 1989; Costello, Edelbrock, & Costello, 1985; Edelbrock et al., 1986; Ivens & Rehm, 1988; Reich & Earls, 1987) than for internalising disorders. When parents are the sole source of information, these disorders are underdiagnosed.

Compared to the Carlson et al. (1987) study, which also used DICA–P–C, we obtained better sensitivity for conduct disorder and OAD, ADHD, and ODD. Lower sensitivity was found in the current study for affective disorders and SAD. As Canino et al. (1987) and Cohen et al. (1987a) noted, the nature of the population is a determinant of the expected size of the diagnostic agreement. The more severely ill the cases are, the easier it is to distinguish them from non-cases.

The higher prevalence of MDD among adolescents, according to the diagnosis made by clinicians compared to parental-report DICA–Rs, could also result from the application of the diagnostic algorithms. Ivens and Rehm (1988) found children and adolescents reported more depressive symptoms than parents. MDD, phobia, and bulimia all have internalising and cognitive symptomatology that may be unknown to parents.

Finally, we found a significant increase in sensitivity of the DICA–R compared to the clinician's diagnostic procedure using a combination of information from different sources. This is probably due to the fact that clinicians usually rely on more than one source of information to derive diagnoses.

For the diagnosis of MDD the same sensitivity was obtained whether child or parent information was used

(40%). The combination of information from both sources increased the sensitivity of the interview to 60%. Although depression is an internalising disorder, some symptoms are observable (e.g. sadness, agitation, fatigue). Therefore, both sources are of value and important for the diagnosis of this syndrome.

A similar explanation can be given in the case of SAD—also an internalising disorder—in which the parents may be disturbed by the clinging behaviour of their children, and the children become distressed by their feelings of anxiety. Consequently, in this age group, both informants contribute valuable information. Combining parent and child information in enuresis probably made it possible to detect cases among children who were ashamed of admitting to the interviewer that they wet the bed.

For adolescents, ODD—a disorder with a marked lack of insight—was diagnosed in greater agreement by clinicians by combining information from the different sources. MDD was more easily detected with the aid of combined information; however, in this case, the improvement in identification was very slight, as sensitivity grew from 48% in adolescent interview to 52% in combined diagnosis (20% in parent interview alone). Thus, the most important informant of depressive symptoms is the adolescent. Nevertheless, the gains in sensitivity to detect OAD, phobia, anorexia, and bulimia by using combined diagnoses are remarkable, indicating the need to gather information about these disorders from both adolescents and their parents.

With respect to the comparison between children aged 6 to 12 years as informants, their parents, and combined information from both sources, the *child* would be a better informant for OAD and anorexia; the *child's parents* would help better to identify ODD and encopresis; and *combining information from child and parent* would be more appropriate for ADHD, MDD, SAD, and enuresis.

Adolescent reports are most valuable for identifying conduct disorder, MDD and PTSD; their *parents* are most informative about ADHD, enuresis, and encopresis; and the *combination of the information* from both would result in the best detection of ODD, phobia, anorexia, and bulimia.

Our results concerning combined diagnosis are similar to those of Piacentini et al. (1993), who used the DISC-R. We found that combining information in this way was both practical and consistent with clinical practice. However, the improvement in sensitivity in our study due to this practice was applicable to more disorders than in Piacentini et al.'s work, in which MDD was the only disorder detected with higher sensitivity by using the combined diagnosis. We agree with Rubio-Stipec et al. (1994) that the importance of the contribution of the information from parents and children varied with the diagnosis.

Summary

The goal of the study was to study the diagnostic agreement between a structured diagnostic interview, the DICA-R, and clinicians. The main problem in the development of a valid research interview is the lack of a

gold standard against which to compare different instruments. The degree of concordance we found between clinician and interview was low to moderate. Reasons include the following: (a) the more strict operational definition of the DICA-R algorithm than the procedure used by clinicians to arrive at a diagnosis; (b) the "tailored" use of DSM-III-R criteria by clinicians; (c) the fact that clinicians use a hierarchical system for diagnoses, ignoring comorbidity, past symptoms, and mild conditions; (d) the use of different and inconsistent sources of information by clinicians; (e) the use by clinicians of information that is not part of the formal diagnostic system; and finally (f) the diagnostic algorithms of the interview, which excluded distress and impairment symptoms. On the whole, however, we feel that the structured diagnostic interview is superior for research purposes.

Future research should address the problem of diagnostic information being obtained from different sources to make a diagnosis, the symptom agreement vs. syndrome agreement discrepancy, and the inclusion of "near-miss cases" in the study of the concordance between clinicians and structured interviews. Furthermore, clear instructions should be given about the period of time during which symptoms are considered to contribute toward a particular diagnosis. A review of the diagnostic criteria used, bearing in mind the feasibility of diagnostic comorbidity and the varying ways in which information can be combined from different sources, must be established, and the process of psychiatric assessment should be kept under review throughout the study, with the participation of clinicians.

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.

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd edn.–revised). Washington, DC: American Psychiatric Association.

Angold, A. (1994). Clinical interviewing with children and adolescents. In M. Rutter, E. Taylor, & L. Hersov (Eds.), *Child and adolescent psychiatry* (pp. 51–63). Oxford: Blackwell Scientific Publications.

Anthony, J. C., Folstein, M., Romanoski, A. J., Von Korff, M. R., Nestadt, G. R., Chahal, R., Merchant, A., Brown, H., Shapiro, S., Kramer, M., & Gruenberg, E. M. (1985). Comparison of the lay Diagnostic Interview Schedule and a standardised psychiatric diagnosis. *Archives of General Psychiatry, 42*, 667–675.

Apter, A., Orvaschel, H., Laseg, M., Moses, T., & Tyano, S. (1989). Psychometric properties of the K-SADS-P in an Israeli adolescent inpatient population. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*, 61–65.

Aronen, E. T., Noam, G. G., & Weinstein, S. R. (1993). Structured diagnostic interviews and clinicians' discharge

diagnoses in hospitalized adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 32*, 674–681.

Boyle, M. H., Offord, D. R., Racine, Y., Sanford, M., Szatmari, P., Fleming, J. E., & Price-Munn, N. (1993). Evaluation of the Diagnostic Interview for Children and Adolescents for use in the general population samples. *Journal of Abnormal Child Psychology, 21*, 663–681.

Brenton, J. J., Bergeron, L., & Valla, J. P. (1994). Do children aged 9 to 11 understand questions from the DISC-2? *Journal of the American Academy of Child Adolescent Psychiatry, 33*, 919.

Canino, G. J., Bird, H. R., Rubio-Stipec, M., Woodbury, M. A., Ribera, J. C., Huertas, S. E., & Sesman, M. J. (1987). Reliability of child diagnosis in a Hispanic sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 560–565.

Carlson, G. A., Kashani, J. H., Thomas, M. F., 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.

Cohen, P., O'Connor, P., Lewis, S., Velez, C. N., & Malachowski, B. (1987a). Comparison of DISC and K-SADS-P interviews of an epidemiological sample of children. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 662–667.

Cohen, P., Velez, N., Kohn, M., Schwab-Stone, M., & Johnson, J. (1987b). Child psychiatric diagnosis by computer algorithm: Theoretical issues and empirical tests. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 631–638.

Costello, C. G. (1992). Research on symptoms versus research on syndromes. *British Journal of Psychiatry, 160*, 304–308.

Costello, E. J., Edelbrock, C., & Costello, A. (1985). Validity of the NIMH Diagnostic Interview Schedule for Children: A comparison between pediatric and psychiatric referrals. *Journal of Abnormal Child Psychology, 13*, 579–595.

Edelbrock, C., Costello, A., Dulcan, M. K., Calabro, N., & Kalas, R. (1986). Parent–child agreement on child psychiatric symptoms assessed via structured interview. *Journal of Child Psychology and Psychiatry, 27*, 181–190.

Fisher, P. W., Shaffer, D., Piacentini, J. C., Lapkin, J., Kafantaris, V., Leonard, H., & Herzog, D. B. (1993). Sensitivity of Diagnostic Interview Schedule for Children, 2nd edn. (DISC-2.1) for specific diagnoses of children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 32*, 666–673.

Gratacós, Ll. (1993). *Psicopatología infantil y variables familiares: Factores de riesgo y factores protectores*, Unpublished doctoral dissertation, Universitat Autònoma de Barcelona, Barcelona, Spain.

Helzer, J. E., Robins, L. N., McEvoy, L. T., Spitznagel, E. L., Stolzman, R. K., Farmer, A., & Brockington, I. F. (1985). A comparison of clinical Diagnostic Interview Schedule diagnoses. *Archives of General Psychiatry, 42*, 657–666.

Helzer, J. E., Robins, L. N., Taibleson, M., Wooddruff, R., & Reich, T. (1977). Reliability of psychiatric diagnosis. *Archives of General Psychiatry, 34*, 122–141.

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 parent on individual symptoms. *Journal of Abnormal Child Psychology, 10*, 307–324.

Hodges, K., McKnew, D., Burbach, D. J., & Roebuck, L. (1987). Diagnostic concordance between the Child Assess-ment Schedule (CAS) and the Schedule for Affective Disorders and Schizophrenia for School-age children (K-SADS) in an outpatient sample using lay interviewers. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 654–661.

Hollingshead, A. B. (1975). *Four-factor index of social status*. Unpublished manuscript, Department of Sociology, Yale University, U.S.A.

Ivens, C., & Rehm, L. P. (1988). Assessment of childhood depression: Correspondence between reports by child, mother and father. *Journal of the American Academy of Child and Adolescent Psychiatry, 27*, 738–741.

Kashani, J. H., Orvaschel, H., Burk, J. P., Reid, J. C. (1985). Informant variance: The issue of parent–child disagreement. *Journal of the American Academy of Child and Adolescent Psychiatry, 24*, 437–441.

Lane, J. W., Pollard, A., & Cox, G. I. (1990). Validity study of the anxiety symptoms interview. *Journal of Clinical Psychology, 46*, 52–57.

Loeber, R., Green, S. M., & Lahey, B. B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. *Journal of Clinical Child Psychology, 19*, 136–143.

Matarazzo, J. D. (1983). The reliability of psychiatric and psychological diagnosis. *Clinical Psychology Review, 3*, 103–145.

Okasha, A. S., Sadek, A., Al-Haddad, M. K., & Abdel-Mawgoud, M. (1993). Diagnostic agreement in psychiatry. *British Journal of Psychiatry, 162*, 621–626.

de la Osa, N., Ezpeleta, L., Doménech, E., Navarro, J. B., & Losilla, J. M. (1996). Fiabilidad entre entrevistadores de la Entrevista Diagnóstica Estructurada para Niños y Adolescentes (DICA-R). *Psicothema, 8*, 359–368.

Piacentini, J. C., Shaffer, D., Fisher, P., Schwab-Stone, M., Davies, M., & Gioia, P. (1993). The Diagnostic Interview Schedule for Children-Revised Version (DISC-R): III. Concurrent criterion validity. *Journal of the American Academy of Child and Adolescent Psychiatry, 32*, 658–665.

Reich, W., & Earls, F. (1987). Rules for making diagnoses in children on the basis of multiple sources of information: Preliminary strategies. *Journal of Abnormal Child Psychology, 15*, 601–616.

Reich, W., Herjanic, B., Welner, Z., & Gandhy, P. R. (1982). Development of a structured psychiatric interview for children: Agreement on diagnosis comparing child and parent interviews. *Journal of Abnormal Child Psychology, 10*, 325–336.

Reich, W., Shayka, J. J., & Taibleson, Ch. (1991). *Diagnostic Interview Schedule for Children and Adolescents—DICA-R*. Unpublished manuscript, Division of Child Psychiatry, Washington University, U.S.A.

Rey, J. M., Plapp, J. M., & Stewart, G. W. (1989). Reliability of psychiatric diagnosis in referred adolescents. *Journal of Child Psychology and Psychiatry, 30*, 879–888.

Robins, L. (1985). Epidemiology: Reflections on testing the validity of psychiatric interviews. *Archives of General Psychiatry, 42*, 918–924.

Rubio-Stipec, M., Canino, G. J., Shrout, P., Dulcan, M., Freeman, D., & Bravo, M. (1994). Psychometric properties of parents and children as informants in child psychiatry epidemiology with the Spanish Diagnostic Interview Schedule for Children (DISC.2). *Journal of the Abnormal Child Psychology, 22*, 703–720.

Shrout, P. E., Spitzer, R. L., & Fleiss, J. L. (1987). Quantification of agreement in psychiatric diagnosis revisited. *Archives of General Psychiatry, 44*, 172–177.

Sylvester, C. E., Hyde, T. S., & Reichler, R. J. (1987). The Diagnostic Interview for Children and the 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,* 668–686.

Vailliant, G. E., & Schnurr, P. (1988). What is a case? *Archives of General Psychiatry, 45,* 313–319.

Vitiello, B., Malone, R., Buschle, P. M., Delaney, M. A., & Behar, D. (1990). Reliability of DSM–III diagnoses of hospitalized children. *Hospital and Community Psychiatry, 41,* 63–67.

Weinstein, S. R., Stone, K., Noam, G. G., Grimes, K., & Schwab-Stone, M. (1989). Comparison of DISC with clin-

icians' DSM–III diagnoses in psychiatric inpatients. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 53–60.

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.