Eur Arch Psychiatry Clin Neurosci (1999) 249 : 174–179

ORIGINAL PAPER

Martin Preisig · Brenda T. Fenton
Marie-Louise Matthey · Alexandre Berney
François Ferrero

Diagnostic interview for genetic studies (DIGS): inter-rater and test-retest reliability of the French version

Received: 23 March 1998 / Accepted: 12 March 1999

Abstract The National Institute of Mental Health developed the semi-structured Diagnostic Interview for Genetic Studies (DIGS) for the assessment of major mood and psychotic disorders and their spectrum conditions. The DIGS was translated into French in a collaborative effort of investigators from sites in France and Switzerland. Inter-rater and test-retest reliability of the French version have been established in a clinical sample in Lausanne. Excellent inter-rater reliability was found for schizophrenia, bipolar disorder, major depression, and unipolar schizoaffective disorder while fair inter-rater reliability was demonstrated for bipolar schizoaffective disorder. Using a six-week test-retest interval, reliability for all diagnoses was found to be fair to good with the exception of bipolar schizoaffective disorder. The lower test-retest reliability was the result of a relatively long test-retest interval that favored incomplete symptom recall. In order to increase reliability for lifetime diagnoses in persons not currently affected, best-estimate procedures using additional sources of diagnostic information such as medical records and reports from relatives should supplement DIGS information in family-genetic studies. Within such a procedure, the DIGS appears to be a useful part of data collection for genetic studies on major mood disorders and schizophrenia in French-speaking populations.

Key words Reliability · Psychiatric diagnoses · Semi-structured interview · Genetic studies

Introduction

Diagnostic misclassification as well as heterogeneity provide potential explanations for recent failures to replicate findings in psychiatric linkage and association studies. With respect to misclassification, standardized diagnostic criteria and (semi)structured interviews have been developed to reduce the risk of inaccurate assessment. To date, the diagnoses of subjects in most family and genetic linkage studies in Europe and in the United States were based on the semi-structured Schedule for Affective Disorders and Schizophrenia – Lifetime Version (SADS-L, Endicott and Spitzer 1978) or a modification thereof. Test-retest reliability studies on this instrument (interval 1 – 2 days) suggested difficulties in diagnosing lifetime bipolar and schizoaffective bipolar disorders. Spitzer et al. (1978) and Keller et al. (1981) reported kappa coefficients of 0.40 and 0.52 for bipolar disorder and 0.47 for schizoaffective disorder (Keller) according to RDC criteria. When the interval was extended to 3 months, Leboyer et al. (1991) observed moderate test-retest agreement (kappa = 0.49) for DSM-III-R bipolar disorder but poor agreement for DSM-III-R schizoaffective disorder (kappa = 0.27).

In order to obtain more accurate psychiatric diagnoses, several semi-structured interviews have recently been developed, including the Schedules for Clinical Assessment in Neuropsychiatry – SCAN (Wing et al. 1990), the Comprehensive Assessment of Symptoms and History – CASH (Andreasen et al. 1992), the Semi-Structured Assessment for the Genetics of Alcoholism – SSAGA (Bucholz et al. 1994), and the Diagnostic Interview for Genetic Studies – DIGS (Nurnberger et al. 1994).

The DIGS was developed by collaborators from the National Institute of Mental Health (NIMH) Genetics Initiative in order to more accurately assess phenotypes of schizophrenia and mood disorders through 1) a semi-structured design corresponding to a wide spectrum of DSM-III-R Axis I criteria and 2) collection of extensive information on the course and chronology of comorbid conditions. In addition to questions in each diagnostic module, there is a chapter consecrated to the description of the temporal relationships between substance disorders and other major Axis I disorders (mood, anxiety, psychotic). This assessment enables diagnostic subtyping of a given disorder as a function of the typical onset sequence of its

M. Preisig (✉) · B. T. Fenton · M.-L. Matthey · A. Berney
University Department of Adult Psychiatry, Hôpital de Cery, CH-1008 Prilly-Lausanne, Switzerland

B. T. Fenton · F. Ferrero
University Hospital of Geneva, Department of Psychiatry, CH-1225 Chêne-Bourg, Switzerland

episodes and those of the comorbid condition. An updated version of the DIGS includes DSM-IV criteria (NIMH Molecular Genetics Initiative 1995).

The original American instrument was extensively tested in 5 sites (Nurnberger et al. 1994). The test-retest reliability (interval 4–10 days in the cross-site phase) was found to be high for the DSM-III-R diagnoses of major depression, bipolar disorder, and schizophrenia with kappa coefficients of 0.94, 0.96, and 0.75, respectively. The agreement for schizoaffective disorders was much lower (kappa = 0.31). The low reliability of DSM-III-R schizoaffective disorders has been attributed to low diagnostic sensitivity and its high confusability with schizophrenia and, to a lesser extent, with bipolar disorder (Faraone et al. 1996). Comparison between DIGS and referral diagnoses revealed high agreement for major depression and bipolar disorder, moderate agreement for schizophrenia, and no agreement for schizoaffective disorder.

The DIGS was translated into French in a multi-site endeavor by a group of bilingual collaborators of the INSERM in Paris and the University Department of Adult Psychiatry in Lausanne (Leboyer et al. 1995). The goals of this study were to:

- test the inter-rater and test-retest reliability of the mood and psychosis sections of the French version of the DIGS; and
- compare the DIGS diagnoses with reviewed clinical diagnoses.

Methods

Sample

The reliability of the DIGS diagnoses of mood disorders and psychosis was tested in 136 consecutive patients from psychiatric inpatient and outpatient facilities of the University Department of Adult Psychiatry in Lausanne with referral diagnoses of bipolar disorder, major depressive disorder, schizoaffective disorder, schizophrenia/schizophreniform disorder, alcohol or drug disorder. Selection for a specific diagnosis stopped after recruitment of 25 patients with that diagnosis. Fifty-two percent of the patients were males and the mean age of the sample was 39.5 years (range 18.5–65.5 years). Patients were required to give written consent. Seventy-four percent of the interviewed patients also participated in the retest phase (N = 101; 2 dropped at analysis phase).

Interview

The development of the DIGS and its specific sections have been extensively described by the authors of the instrument (Nurnberger et al. 1994). The French version of the DIGS was created by a team of bilingual psychologists and psychiatrists from the INSERM in Paris and the University Department of Adult Psychiatry in Lausanne. Establishment of the French version was performed in three stages: 1) a first draft was produced by the Paris team, 2) each section was assigned to a primary reviewer in Lausanne who verified the translation, and 3) the responsible psychiatrist and senior psychologist of the Lausanne group examined the preterminal Lausanne draft of the entire instrument. Ambiguous items which had undergone major revision between the first Parisian draft and the preterminal Lausanne draft were discussed with the team. Several modifications were made in the DIGS. All additional questions were clearly labeled in the interview.

Modifications of the French version involved the following sections:

- An optional screening question for mania was added in order to lower the threshold for entering this chapter by asking whether there was objective evidence of elated mood (i.e., friends or family members have observed that the subject's mood was higher than normal).
- Optional questions were added to allow a better temporal assessment of the last episode in both the major depression and mania sections. This makes it easier to determine whether the last episode is still ongoing, or came to an end within the last month or earlier.
- A chapter on Generalized Anxiety Disorder was added to the anxiety section using questions from the SADS. The assessment of this diagnosis was not included in the original DIGS.
- The brief DIGS phobia chapter was replaced by the corresponding and more extensive sections from the SADS.
- An optional screening question on 'energy lost' was added in the major depression section to also allow diagnosis of major depression according to ICD-10 criteria.
- Given that the Mini-Mental State examination has already been translated into French (Hoff 1990), it was omitted from this validation study.

Procedure

A team of interviewers (8 psychologists and psychiatrists) was trained over a four month period. Training included videotaped interviews which were supervised by experienced clinicians. All interviewers were blinded as to the recruitment source (specific clinical unit) and referral diagnosis of the subject. Raters performed interviews for the inter-rater and test-retest reliability study in a balanced way, exchanging roles between first-interviewer, observer at the first interview, and retest interviewer. Patients were interviewed by a member of the team in the presence of an observer (co-rater) who simultaneously and independently completed a DIGS. At the end of the interview, interviewer and observer independently assigned DSM-IV lifetime diagnoses.

With respect to schizoaffective disorders, operationalized definitions were established for the interpretation of DSM-IV criteria. As disagreement may stem from differences in rating of the schizoaffective C criterion (which requires mood symptoms not be 'brief' compared to psychotic syndromes), we followed the suggestions of Nurnberger et al. (1994) and considered mood symptoms less than 30% of total time with psychosis as 'brief'.

Six weeks later, subjects were contacted again for the retest interview which was conducted by a third member of the team who had no information regarding the first interview. A longer time interval (6 weeks) between the test and retest was selected in comparison to that used to test the original instrument (4 to 14 days) in order to better estimate the stability of the diagnoses. One problem of selecting a longer test-retest period is the potential for the development of new symptoms that can change the diagnosis (i.e., diagnostic discrepancies due to new episodes vs. a lack of reliability in reporting the same episodes). For this reason, we removed from the test-retest analyses those two cases which were found to involve the appearance of first manic episodes in the interval between test and retest (retest analysis N = 99).

Analysis

For both the analyses of inter-rater and test-retest data, 6×6 tables were constructed based on mutually exclusive diagnostic categories of 'psychosis', 'schizoaffective bipolar', 'schizoaffective unipolar', 'bipolar I/bipolar II', 'MDD/dysthymia', and 'no target diagnosis'. The category of psychosis included schizophrenia, schizophreniform disorder, brief psychotic disorder, delusional disorder, substance-induced psychotic disorder, and 'psychotic disorder due to a general medical condition'.

The 6 × 6 tables allowed the computation of overall kappa coefficients (Cohen 1960) for inter-rater and test-retest reliability. For the reliability of specific target diagnoses, kappas were calculated for each disorder based on the presence or absence of the specific target diagnosis (2 × 2 tables). According to Fleiss (1981), agreement was considered as "excellent" when kappas were greater than 0.75, "fair to good" when kappas were between 0.40 and 0.75, and "poor" when kappas were below 0.40. Considering the wide prevalence range of target disorders in our sample and the strong dependence of the kappa coefficient on the base rate of a specific diagnosis, we also calculated Yule's coefficients for specific disorders (Spitznagel and Helzer 1985).

Lastly, we compared diagnoses based on the DIGS interview with clinical diagnoses. Given the need for lifetime as well as comorbid diagnoses, existing clinical discharge diagnoses were found to be insufficient for our purposes. Therefore, we reviewed all clinical diagnoses based upon a standardized examination of symptomatology from all available medical records according to DSM-IV criteria. More specifically, clinical diagnoses were altered or added in the case of contradiction between medical record symptom information and assigned clinical discharge diagnosis (e.g., a diagnosis of schizophrenia was changed to schizoaffective bipolar disorder when substantial episodes of mania were described in the medical record).

Results

Inter-rater reliability

The frequencies and inter-rater agreements for target diagnoses are shown in Table 1. The overall kappa coefficient was 0.87. With the exception of schizoaffective bipolar disorder, the kappa coefficients for specific diagnoses were all in the excellent range between 0.85 (bipolar disorder) and 0.93 (MDD/dysthymia). The agreement for the absence of a target diagnosis was perfect. Within the category of psychosis, schizophrenia was predominant (approximately 68% of psychotic cases). The kappa coefficient for schizophrenia (kappa = 0.87, Yule = 0.91) was almost identical to that for the overall diagnosis of psychosis. Except for delusional disorder (2 cases rated by the interviewer, 1 case rated by the observer), inter-rater agreement was perfect for the rare cases of psychotic disorders other than schizophrenia. With respect to schizoaffective disorders, the considerable discrepancy between the values of the kappa and the Yule's coefficient for schizoaffective bipolar disorder indicates that the relatively low kappa coefficient was partially attributable to the low prevalence of this diagnosis in our sample. The separate analysis of bipolar I and bipolar II disorders revealed a kappa coefficient of 0.73 for bipolar I and 0.53 for bipolar II disorder. The discrepancy between the kappa coefficients of the two types of bipolar disorder was mainly attributable to the relatively rare frequency of bipolar II disorder in our sample (i.e., interviewers assigned this diagnosis to only 6 cases and observers to only 5 cases). In fact, when Yule's coefficients were considered, the reliability estimates for both subtypes of bipolar disorder were almost identical (0.77 for bipolar I, 0.78 for bipolar II).

Test-retest reliability

The results of the test-retest analysis are presented in Table 2. The computed overall kappa of 0.60 was markedly lower than that of the inter-rater analysis. With the exception of schizoaffective bipolar disorder (kappa = 0.38), the kappa's for all individual target diagnoses were in a range of fair to good agreement, ranging from 0.48 (schizoaffective unipolar) to 0.65 ('no target diagnosis'). When Yule's coefficients were considered, the test-retest reliability coefficients of schizoaffective disorders were close to those of other diagnoses, again suggesting that the relatively low kappas were essentially due to the low sample prevalence. Within the category of psychosis, test-retest reliability was high for the predominant diagnosis of schizophrenia (kappa = 0.72, Yule = 0.81). When bipolar disorder was subtyped into bipolar I and bipolar II subtypes, the kappa coefficient for bipolar I diagnosis was 0.48. However, no agreement was obtained for the rare bipolar II patients. Among the 6 cases with that diagnosis at the first interview: 4 received the diagnosis of bipolar I, one of psychosis and one of MDD at the second interview. Conversely, two patients with bipolar I disorder were assigned as bipolar II at the retest interview.

Table 1 Inter-rater reliability

Diagnosis	Prevalence		Kappa	95% C.I.	Yule
	Interviewer	Observer			
Psychosis	26/136	27/136	0.88	0.71–1.00	0.91
Schizoaffective bipolar	11/136	11/136	0.60	0.44–0.77	0.76
Schizoaffective unipolar	8/136	8/136	0.87	0.70–1.00	0.94
Bipolar I/ Bipolar II	37/136	35/136	0.85	0.68–1.00	0.87
MDD/Dysthymia	46/136	47/136	0.93	0.78–1.00	0.94
No target diagnosis	8/136	8/136	1.00	0.83–1.00	1.00

Table 2 Test-retest reliability

Diagnosis	Prevalence		Kappa	95% C.I.	Yule
	Interviewer	Retester			
Psychosis	19/99	18/99	0.63	0.44–0.83	0.70
Schizoaffective bipolar	8/99	2/99	0.38	0.23–0.53	0.58
Schizoaffective unipolar	5/99	3/99	0.48	0.29–0.67	0.77
Bipolar I/ Bipolar II	30/99	27/99	0.63	0.43–0.83	0.66
MDD/Dys-lar II	32/99	42/99	0.62	0.42–0.81	0.67
No target diagnosis	5/99	7/99	0.65	0.45–0.84	0.83

Table 3 Agreement between reviewed clinical and digs diagnoses

Diagnosis	Prevalence		Kappa	95% C.I.	Yule
	Inter-viewer	Clinical			
Psychosis	27/136	26/136	0.55	0.39–0.72	0.63
Schizoaffective bipolar	11/136	11/136	0.51	0.34–0.67	0.69
Schizoaffective unipolar	8/136	3/136	0.34	0.19–0.49	0.73
Bipolar I/ Bipolar II	35/136	30/136	0.58	0.41–0.74	0.63
MDD/Dys-thymia	47/136	54/136	0.54	0.38–0.71	0.56
No target diag-nosis	8/136	12/136	0.35	0.19–0.52	0.59

Interview versus reviewed clinical diagnoses

Table 3 shows the agreement between the DIGS and the reviewed clinical diagnoses. The overall agreement (kappa = 0.52) was fair. With the exception of two categories, 'schizoaffective unipolar disorder' (kappa = 0.34) and 'no target diagnosis' (kappa = 0.35), the agreement for specific diagnoses was fair and within a narrow range (i.e., from 0.51 for schizoaffective bipolar disorder to 0.58 for bipolar disorder). However, in contrast to their low kappa values, the Yule's coefficients for schizoaffective disorders and 'no target diagnosis' were in the same range as those of other specific diagnoses, suggesting that the poor kappa's resulted from the low prevalence of these groups in the sample.

Discussion

The inter-rater and test-retest reliability of major mood and psychotic diagnoses in the French version of the DIGS have been established for psychiatric patients. In order to approximate naturalistic conditions, we have tested the instrument using a consecutive patient sample covering a full range of psychiatric disorders from major depression to schizophrenia. In this type of sample, many subjects were close to the diagnostic borders between mood disorders-schizoaffective disorders and schizoaffective disorders-schizophrenia. With respect to these borders, the clinical referral diagnosis was uncertain for several patients. This approach contrasts that of other studies whose patients have particularly clear clinical presentations and produces more conservative estimates of reliability. Similarly, we limited the source of diagnostic information to the DIGS interview itself which also estimates the lower bound of diagnostic reliability for family studies since best-estimate diagnoses are generally used.

Despite this, the inter-rater study revealed excellent agreement between interviewer and observer for all major target diagnoses with the exception of schizoaffective bipolar disorder. The relatively low kappa coefficient for this diagnosis was still in the fair to good range. The excellent inter-rater agreement indicates that the subjects' responses could be recorded in a highly reliable way leading to identical diagnostic assignment for most cases. Table 4 shows that these inter-rater kappa coefficients are comparable to those of other structured and semi-structured interviews. However, a comparison with the original DIGS was not possible, since inter-rater reliability was not assessed.

Test-retest reliability, which involves two independent interviews and relies on the subject's recall of symptoms,

Table 4 Comparison of inter-rater reliability (kappa coefficients)

	Spitzer et al. (1978)	Leboyer et al. (1991)	Andreasen et al. (1992)	Wittchen (1994)	Preisig et al.
Study design					
Sample	Inpatients	In-/Outpts.	Inpatients	In-/Outpts Gen. Pop.	In-/Outpts.
N	150	38	30	575	136
Instrument	SADS	SADS-LA*	CASH	CIDI	DIGS
Diagnostic Criteria	RDC	DSM-III DSM-III-R	DSM-III-R	DSM-III-R	DSM-IV
Diagnoses					
MDD	0.91	1.00	0.65	–	0.93
MDD single	–	–	–	0.97	–
MDD recurrent	–	–	–	0.93	–
Dysthymia	–	–	1.00	0.96	–
BP		1.00	1.00	–	0.85
BP I	0.95	–	–	0.92	0.73
BP II	0.85	–	–	0.94	0.53
SA	–	0.75	0.45	–	0.69
SAM	–	–	–	–	0.60
SAD	0.87	–	–	–	0.87
Schizophrenia	–	1.00	0.61	0.91	0.87

* French and German versions

is generally lower than inter-rater reliability. As phenotype assessment in genetic studies is usually based on lifetime diagnoses, standardized interviews employed in such studies must provide accurate diagnostic assessment months, or even years, after psychiatric episodes. In order to estimate the stability of the assessment of major mood diagnoses and psychosis, we chose a relatively extended interval of 6 weeks between the two interviews. Compared to most previous studies which used intervals of less than one week, our relatively long time-span minimizes the risk of the subject's recall of his/her previous responses.

Despite the long test-retest interval, we obtained fair to good test-retest reliability for major mood disorders and psychosis with the highest test-retest agreement for schizophrenia. The only diagnosis with poor test-retest reliability was schizoaffective bipolar disorder. However, according to the Yule's coefficient, the relatively low agreement for schizoaffective bipolar diagnosis was attributable to its low sample prevalence, rather than to particular difficulties in the assessment of the DSM-IV criteria.

The analysis of cases with diagnostic discrepancies between the first and second interviews generally revealed less severe diagnoses at retest (if a diagnostic hierarchy considers schizophrenia as the most severe diagnosis followed by diagnoses of schizoaffective, bipolar and major depressive disorder). The finding of a general decrease in reported symptomatology resulted from incomplete symptom recall or reporting at retest with either negative answers to all screening questions of a section or a failure to have the required number of symptoms and/or required duration of symptomatology. It can also be hypothesized that some patients who had particularly long first interviews may have learned that negative answers accelerated the interview. However, although about a third of subjects with psychosis and bipolar disorder did not meet criteria for their respective diagnoses at retest, only three out of 94 subjects with a target diagnosis at first interview had no target diagnosis at retest.

Cross-study comparisons are difficult because of important methodological differences, namely the length of the test-retest interval but also recruitment strategy, the selected diagnostic classification system and the source of information used for diagnostic assignment (e.g., use of medical records/information to create best-estimate diagnoses). It is not surprising that studies with short intervals (less than 10 days) generally found higher test-retest reliability, whereas the study of Leboyer et al. (1991) using twice our test-retest interval reported slightly lower reliability coefficients (Table 5).

Comparisons also revealed that the length of the test-retest interval had an important effect on the reliability of mood diagnoses but not on that of schizophrenia. Despite the large range of test-retest intervals across studies, kappa coefficients for schizophrenia were almost identical across studies. This tendency was also supported by the comparison of test-retest reliability between the French and English versions of the DIGS where the kappa coefficients were similar or even slightly higher for schizophrenia and schizoaffective disorders in our study despite the longer test-retest interval, but substantial differences existed in the reliability of MDD and bipolar disorder.

In our study, schizophrenia had low inter-rater agreement (compared to mood disorders) but high test-retest agreement resulting in a relatively small difference between the two estimates. The relatively low inter-rater reliability for schizophrenia was essentially due to vague or inconsistent responses of psychotic patients regarding the duration of psychotic symptoms, whereas the high test-retest reliability may be the result of the greater persistence of psychotic symptoms as compared to mood symptoms which favors their recall.

Table 5 Comparison of test-retest reliability (kappa coefficients)

	Spitzer et al. (1978)	Keller et al. (1981)	Leboyer et al. (1991)	Andreasen et al. (1992)	Nurnberger et al. (1994)	Preisig et al.
Study design						
Sample	Inpatients	In-/Outpts.	In-/Outpts.	Inpatients	Outpt.	In-/Outpts.
N	60	25	36	30	81	99
Instrument	SADS	SADS	SADS-LA	CASH	DIGS	DIGS
Interval (days)	1–2	1	97	1	4–10	42
Diagn. Criteria	RDC	RDC	DSM-III-R	DSM-III-R	DSM-III-R	DSM-IV
Diagnoses						
MDD	0.71	0.82	0.46	0.65	0.94	0.59
BP	–	–	0.49	0.90	0.96	0.63
BP I	0.40	0.52	–	–	–	0.48
BP II	–	0.26	–	–	–	0.00
SA	–	–	0.27	0.52	0.31	0.40
SAM	–	0.47	–	–	–	0.38
SAD	0.70	0.67	–	–	–	0.48
Schizophrenia	0.73	0.60	0.71	0.72	0.75	0.72
No target	–	–	–	0.78	0.86	0.65

Similar to the original reliability study, we found fair agreement between DIGS and reviewed clinical diagnoses. However, in contrast, we found agreement (in terms of the Yule's coefficient) to vary only slightly across diagnoses, including schizoaffective disorders. This finding along with the higher test-retest reliability for these disorders in our study might reflect a stronger tradition of their assignment in our clinical setting.

In summary, the French version of DIGS demonstrated excellent inter-rater reliability and fair to good 6-week test-retest reliability for mood diagnoses and schizophrenia. In order to increase reliability, particularly for lifetime mood diagnoses in persons not currently affected, best-estimate procedures using additional sources of diagnostic information such as medical records and reports from relatives should supplement DIGS information in family-genetic studies. Within such a procedure, the DIGS appears to be a useful part of data collection for genetic studies on major mood disorders and schizophrenia in French-speaking populations.

Acknowledgments This work was supported by the Swiss National Science Foundation (Grant number 32-40677.94). We would like to thank the administrations of the participating institutions as well as the patients who agreed to be interviewed.

References

Andreasen NC, Flaum M, Aendt S (1992) The Comprehensive Assessment of Symptoms and History (CASH). Arch Gen Psychiatr 49: 615–623

Bucholz K, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger J, Reich T, Schmidt I, and Schuckit MA (1994) A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. Journal of Studies on Alcohol 55: 149–158

Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Measurement 20: 37–46

Endicott J & Spitzer RL (1978). A diagnostic interview: the Schedule for Affective Disorders and Schizophrenia – lifetime version (modified for the study of anxiety disorders). Arch Gen Psychiatr 35: 837–844

Faraone SV, Blehar MC, Pepple J, Modlin S, Norton J, Nurnberger JI, Malaspina, D, Kaufmann C, Reich T, Cloninger CR, De Paolo JR, Berg K, Gershon E, Kirch D, Tsuang MT (1996) Diagnostic accuracy and confusability analyses: An application to the Diagnostic Interview for Genetic Studies. Psych Med 26: 401–410

Fleiss JL (1981) Statistical Methods for Rates and Proportions, ed. 2. John Wiley, New York

Hoff X (1990) Intéret et limite du test de Folstein dans l'évaluation des performances de sujets âgés, non-déments et déments hospitalisés. Psychiatrie et Psychobiologie 5: 257–263

Keller MB, Lavori PW, McDonald-Scott P, Scheftner WA, Andreasen NC, Shapiro RW, Croughan J (1981) Reliability of lifetime diagnoses and symptoms in patients with a current psychiatric disorder. J Psychiat Res 16: 229–240

Leboyer M, Barbe B, Gorwood P, Teherani M, Allilaire JF, Preisig M, Matthey ML, Poyetton V, Ferrero F (1995) Interview diagnostique pour les études génétiques

Leboyer M, Maier W, Teherani M, Lichtermann D, D'Amato T, Franke P, Lépine JP, Minges J, McGuffin P (1991) The reliability of the SADS-LA in a family study setting. Eur Arch Psychiatry Clin Neurosci 241: 165–169

NIMH Molecular Genetics Initiative (1995) Diagnostic Interview for Genetic Studies (DIGS) – Version 2.0

Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler CY, Simpson SG, Harkavy-Friedman JH, Severe JB, Malaspina D, Reich T (1994) Diagnostic Interview for Genetic Studies. Rationale, unique features, and training. Arch Gen Psychiatr 51: 849–859

Spitzer RL, Endicott J, and Robins E (1978) Research Diagnostic Criteria: Rationale and reliability. Arch Gen Psychiatr 35: 773–782

Spitznagel EL and Helzer JE (1985) A proposed solution to the base rate problem in the kappa statistic. Arch Gen Psychiatr. 42: 725–728

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablensky A, Regier D, Sartorius N (1990) SCAN: Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatr 47: 589–593

Wittchen HU (1994) Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): a critical review. J Psychiat Res 28: 57–84