

Validation of a new diagnostic procedure for DSM IV axis I disorders

JEAN-YVES LOZE, BRUNO FALISSARD, FREDERIC LIMOSIN, CHRISTOPHE RECASENS, ANNE-SOPHIE HORREARD, FREDERIC ROUILLON, Centre Hospitalier Universitaire Albert Chenevier, Service de Psychiatrie, Créteil, France

ABSTRACT *Despite the fact that, in today's psychiatric research and especially in epidemiological studies, diagnostic assessments are made with reliable standardized clinical interviews, recent articles have shown discrepancies in prevalence rates of DSM IV axis I disorders assessed with different, yet reliable, clinical standardized interviews, raising the problem of the clinical relevance of some of these instruments.*

Within an epidemiological study, we developed a simple method for evaluating DSM IV axis I disorders with the aim of improving the clinical relevance of assessed diagnoses. This method is based on an evaluation performed by two clinicians. The first one used a short structured clinical interview (MINI v 5.0) and the second one completed the procedure with an open clinical interview, intended to be more clinically relevant. Finally, a consensus diagnosis is given by the two investigators. We conducted a survey in order to validate this method by measuring the agreement of diagnoses reported by two pairs of clinicians on a population of 20 inpatients. Results show that this double evaluation led to a high agreement (kappa ranging between 0.76 and 1.00) suggesting that the proposed evaluation procedure, which is intended to be more clinically relevant, is also highly reliable.

Key words: validation, diagnostic procedure, kappa coefficient, DSM IV axis I disorder, standardized interview

Introduction

In usual clinical practice, the evaluation of DSM IV axis I disorders (American Psychiatric Association, 1994) is based on an open interview performed by a clinician. Despite the fact that this approach is supposed to be the most clinically relevant, it cannot be used in clinical trials or epidemiological studies because of its lack of reliability. The use of standardized clinical interviews increases diagnostic reliability, making comparisons between studies possible (Anthony et al., 1985; Helzer et al., 1985).

In epidemiological studies, discrepancies in prevalence rates of DSM IV axis I disorders have led some authors to question the clinical relevance of standardized instruments (Narrow et al., 2002). Although the reliability of these instruments is globally satisfying, recent studies have shown that inter-rater agreement between some standardized clinical interviews may be low.

Eaton et al. (2000) showed that inter-rater agreement between a standardized clinical interview administered by a non-clinician (Diagnostic Interview Schedule (DIS) (Robins et al., 1981)) and a standardized clinical interview administered by a psychiatrist (Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (Wing et al., 1990)) was as low as 0.20 for depressive disorders. Murphy et al. demonstrated that the inter-rater agreement between the DIS and the depression component of version 2 of author's customary assessment method (DPAX-2) (Murphy et al., 1985; Murphy et al., 1998), both administered by graduate students, was 0.40 for current depression and 0.33 for lifetime depression.

These results outline the problem of clinical relevance encountered with such instruments and suggest the possibility that new evaluation methods should be developed for psychiatric disorders in order to improve

clinical relevance without hampering reliability or feasibility.

Therefore, we focused on a method designed to evaluate DSM IV axis I disorders in large community samples using a double evaluation performed by a pair of clinicians. One of them evaluates DSM IV axis I disorders with a short, reliable, standardized clinical interview, the Mini International Neuropsychiatric Interview (MINI) French version v5.0.0 (Sheehan et al., 1998) and the other uses an open interview, which is thought to be more clinically relevant. Our hypothesis was that the combination of both evaluations by the two clinicians would lead to a more reliable and relevant consensus diagnosis compared with either one taken separately. We aimed to demonstrate and validate this procedure on a sample of 20 inpatients.

Methods

Description of the procedure (cf. Figure 1)

The proposed procedure is based on a double evaluation that is performed during a single interview by a pair of clinicians. One of them should be a senior clinician with at least five years experience, while the other one should be a junior clinician displaying less experience and who could be a graduate student. The senior clinician performs a clinical open interview, thought to be more clinically relevant, while the junior clinician administers the MINI, a short, reliable, standardized clinical interview, designed to assess current DSM IV axis I diagnoses except for affective disorders, panic disorder and psychotic disorders, where lifetime diagnoses are assessed.

The evaluation is performed as follows: during a period of approximately 5 minutes, the senior clinician introduces both investigators to the patient and gives general information about the procedure. Then, in the presence of the senior clinician, the junior clinician starts his assessment with the MINI interview, which will last for approximately 30 minutes. Then, in the presence of the junior clinician, the senior clinician performs his own assessment for DSM IV axis I disorders using an open interview for a maximum period of 20 minutes. In order to avoid diagnoses discrepancies with the MINI, the senior clinician is instructed to evaluate the same periods of patient's history as those assessed by the MINI and for each diagnostic category. The main purpose of the second evaluation performed

by a senior clinician is to improve the accuracy of diagnoses made with the MINI interview by clarifying points that remain uncertain or questionable.

After these two evaluations, the junior clinician documents the set of diagnoses made with the MINI. Additionally, both investigators independently document their own set of diagnoses in terms of DSM IV criteria, which is slightly reformulated in order to take into account specificities of the French nosography (cf. Table 1).

Then, both clinicians consult each other with the purpose of finding a set of consensus diagnoses. Finally, four different sets of diagnoses are reported for each patient (the junior set, the senior set, the MINI set and the consensus set), with the consensus diagnoses being predominant.

Validation of the procedure

Validation of the proposed diagnostic procedure was based on the assessment of the inter-rater agreement. In order to make it possible, a sample of 20 pseudo-randomly selected inpatients was set up. This sample consisted of patients admitted within the last seven days in a general psychiatric department of the suburbs of Paris. Among the 20 patients, half of them ($n = 10$) were females, mean age was 40.85 years ($sd = 16.67$; minimum = 18; maximum = 78), 50% ($n = 10$) were single and 40% ($n = 8$) were working. Patients' admission diagnoses were psychotic disorders ($n = 6$), affective disorders ($n = 15$), substance abuse/dependence ($n = 5$), anxiety disorders ($n = 6$).

Once the sample was selected, two pairs of clinicians were created (JYL – FL and ASH – CR), each pair involving one junior and one senior clinician. Each patient was assessed twice within five days according to the previously described procedure. The first assessment was performed by one of the two pairs of clinicians and the second performed by the other pair. Each investigator was blind to the patient's previously documented diagnoses. For each patient, four sets of diagnoses were collected twice (eight in total).

Statistical analysis

The study variables were binary (diagnostic present or not). Cohen's kappa coefficients were calculated with SPSS v11.0.1 (r) to assess inter-rater agreement for diagnostic categories between the two pairs of clinicians; 95% confidence intervals were calculated using the bootstrap method.

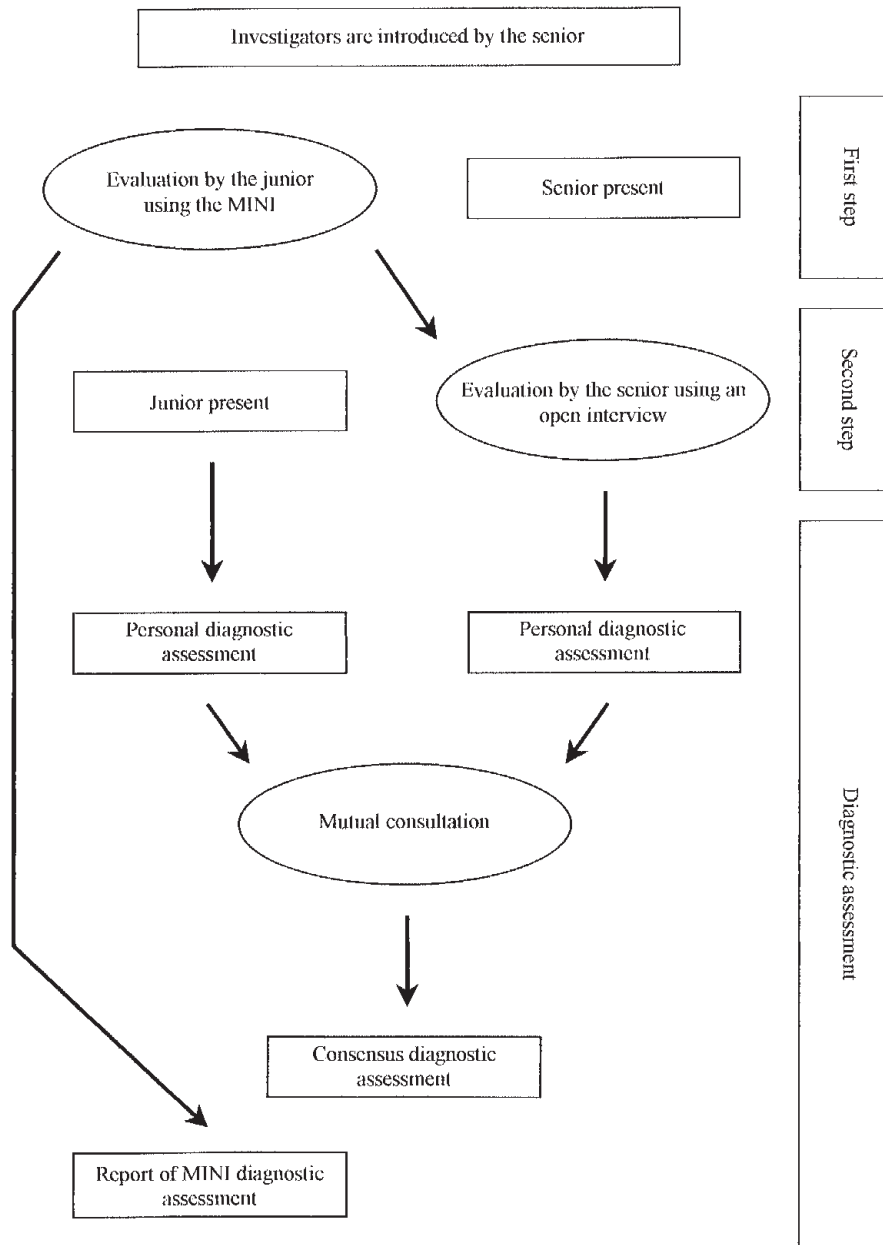


Figure 1. Summary of the diagnostic procedure.

In order to provide additional results regarding inter-rater agreement, a multidimensional scaling principal component analysis was performed with R v1.4.0 software (Ihaka et al., 1996).

Results

Results of inter-rater agreement assessed with the computation of kappa coefficients are presented in Table 2.

In order to make results more intelligible and to reduce the number of diagnostic categories, most of these categories have been consolidated according to Table 1. Results show that inter-rater agreement between the two pairs of clinicians is satisfactory overall. Especially for the consensus diagnoses (range: 0.77 to 1.00) and for diagnoses made with the MINI (range: 0.78 to 1.00), kappa coefficients are higher in comparison with

those observed with diagnoses made by the junior clinicians (range: 0.38 to 1.00) or the senior clinicians (range: 0.50 to 1.00). The high kappa values obtained for diagnoses made with the MINI were not surprising because the MINI was supposed to be more reliable. Concerning the consensus diagnoses, which are thought to be more clinically relevant, the high kappa values indicate that our procedure seems to be reliable. Lower kappa values observed with diagnoses made by the junior and senior clinician confirm that diagnostic assessment performed with an open interview only is a less reliable procedure, which was expected.

Results of the multidimensional scaling analysis performed with a principal component analysis, are reported in Figure 2, providing a visual depiction of inter-rater agreement.

It appears that diagnostic categories are well individualized with a close expected association between all affective disorders (major depression and bipolar disorders) and a weaker association between affective disorders and anxiety disorders. This is consistent with the literature.

More surprising is the absence of association between affective disorders and substance abuse/

dependence (points are orthogonal). Our explanation for this result is that diagnoses of substance abuse/dependence were assessed only over the past year, whereas affective disorders were lifetime diagnoses, therefore the association between the two diagnosis categories might be underestimated.

Confirming the results obtained previously with the kappa coefficients, it appears that for each pooled diagnostic category, distances between points representing consensus diagnoses made by the two pairs of clinicians, as well as those representing MINI's diagnoses, are closer in comparison with those observed for junior and senior clinicians' diagnoses.

Moreover, this analysis enables us to show if there is a 'pair-effect': in other words a grouping of diagnoses coming from a particular pair of clinicians. This phenomenon is observed for anxiety disorders where points representing the pair of clinicians number 1 can be isolated from those representing pair number 2.

As for which evaluation is more influential in the consensus diagnoses, it appears that none of the points representing junior, senior or MINI assessments are regularly or constantly close to those representing consensus assessments.

Table 2. Inter-rater agreement assessed on a 20 in-patients population between two pairs of clinicians, each one being constituted by a junior and a senior clinician

	Agreement for juniors' diagnoses		Agreement for seniors' diagnoses		Agreement for MINIs' diagnoses		Agreement for consensus diagnoses				
	Kappa	95% CI	Kappa	95% CI	Kappa	95% CI	n	Kappa	95% CI		
Psychotic disorder	0.86	0.64 – 1.00	0.88	0.70 – 1.00	0.83	0.56 – 1.00		+	-	0.88	0.70 – 1.00
							+	5	0		
							-	14	1		
Affective disorder	1.00	-	0.88	0.70 – 1.00	1.00	-		+	-	1.00	-
							+	15	0		
							-	0	5		
Major depression	0.89	0.73 – 1.00	0.60	0.39 – 0.80	1.00	-		+	-	0.89	0.73 – 1.00
							+	13	1		
							-	0	6		
Bipolar disorder	0.38	0.03 – 0.68	0.50	0.20 – 0.78	0.88	0.70 – 1.00		+	-	0.86	0.64 – 1.00
							+	4	1		
							-	0	15		
Substance or alcohol abuse/dependence	0.74	0.49 – 0.94	1.00	-	1.00	-		+	-	0.88	0.70 – 1.00
							+	5	0		
							-	1	14		
Anxiety disorder	0.53	0.24 – 0.76	0.63	0.35 – 0.84	0.78	0.59 – 0.95		+	-	0.77	0.56 – 0.95
							+	5	2		
							-	0	13		

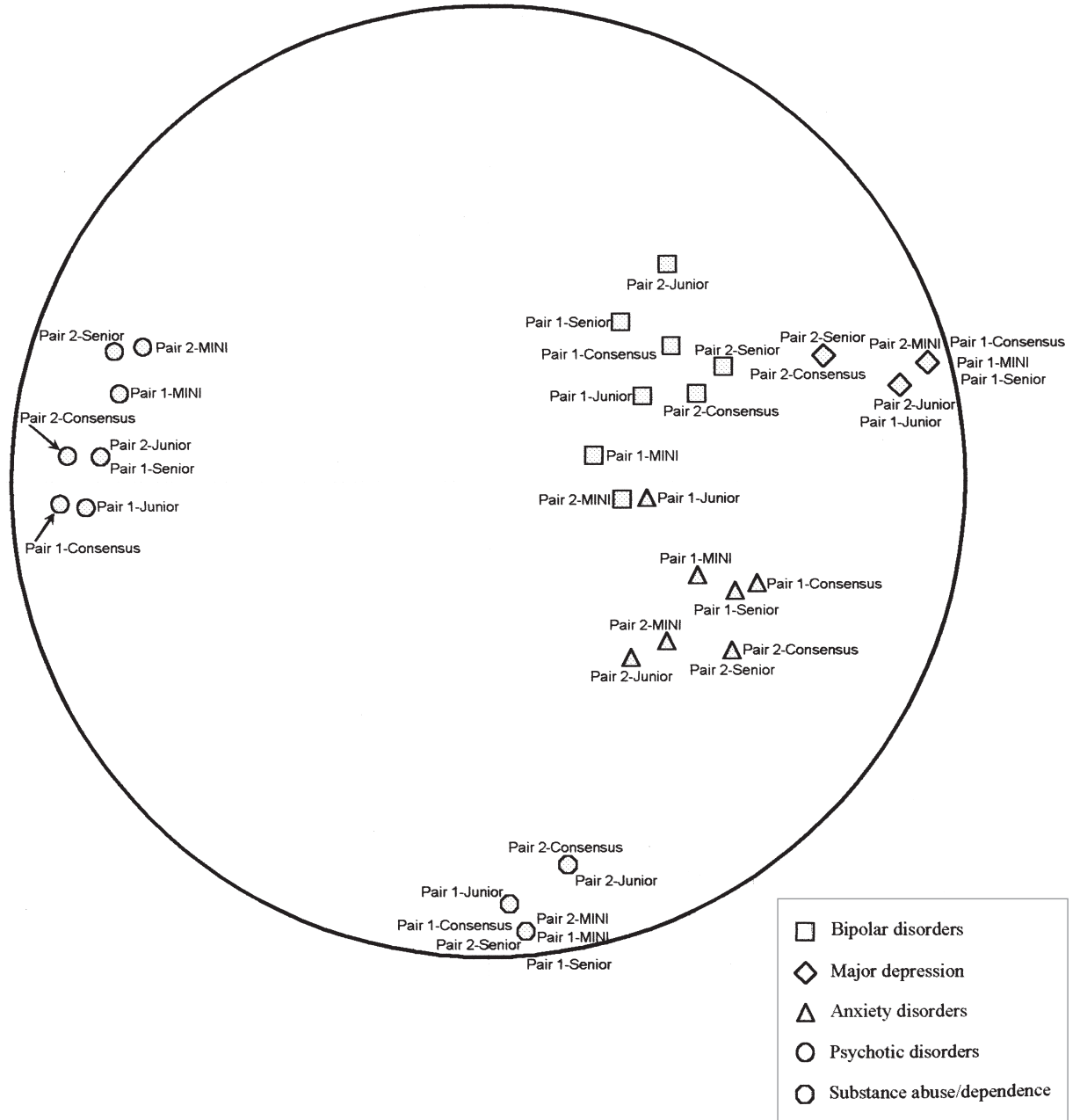


Figure 2. Multidimensional scaling – principal component analysis of rater agreement performed on a population of 20 inpatients by two pairs of clinicians.

Discussion

Results show that the diagnostic procedure for DSM IV axis I disorders that we propose seems to be reliable. The main benefit of this procedure lies in the fact that

it leads to more clinically relevant diagnoses than classical structured clinical interviews. This improvement is achieved through the combination of the administration of a structured clinical interview and an

open interview, which allows the investigator freely to propose the diagnoses he believes to be the most appropriate. This point was considered critical when we developed this procedure.

We used a structured clinical interview like the MINI because it systematically scans the whole axis I psychiatric pathology, giving suggestions to the clinicians, and more specifically to the senior clinician, for exploring some diagnostic categories that could otherwise have been underestimated. This certainly increases the reliability of assessments made by clinicians and our results emphasize this point.

The sample from which the validation study has been carried out, although consisting of inpatients, has been selected in a general psychiatric department treating patients with no or limited social disabilities. We think, therefore, that our diagnostic method of evaluation can be used in community samples.

Our procedure is only applicable for DSM-IV axis I disorders and does not apply to axis II disorders. This represents an important limitation. Moreover, the small size of the sample limits the interpretation of our results. Therefore, validation of this procedure should be made on a larger sample. Another limitation regarding feasibility is that the procedure requires two clinicians.

Nevertheless, our procedure is easy to implement and duration does not exceed one hour, generating an appreciable time saving compared with classical standardized clinical interviews. Especially for epidemiological studies, in which evaluations involve large samples and should be performed in a minimum amount of time, the benefit of using such a procedure is obvious.

Conclusion

This validation study, although limited to 20 inpatients, shows that our diagnostic procedure leads to reliable diagnosis assessments that are more clinically relevant. This method of evaluating DSM IV axis I disorders requires two clinicians per subject, which is certainly a limitation of feasibility, but the entire duration of evaluation per subject is much lower in comparison with durations observed with classical standardized clinical instruments.

We therefore suggest that the benefits of this diagnosis procedure, generating a more accurate clinical diagnosis, are superior to its limitations and that it can

minimize evaluation biases, which are a major problem in epidemiological studies.

Acknowledgements

This study was aiming at the validation of a diagnostic procedure for DSM IV axis I psychiatric disorders. It was conducted within the scope of an epidemiological study supported by the French government (department of justice and department of health) that will evaluate DSM IV axis I psychiatric disorders among prisoners in France.

References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington DC: American Psychiatric Association, 1994.
- Anthony JC, Folstein MF, Romanoski AJ, Von Korff MR, Nestadt G S, Kramer M, Gruenberg E. Comparison of the lay Diagnostic Interview Schedule and a standardized psychiatric diagnosis: experience in eastern Baltimore. *Arch Gen Psychiatry* 1985; 42: 667–75.
- Eaton Ww, Neufeld K, Chen LS, Cai G. A comparison of self report and clinical diagnostic interviews for depression: diagnostic interview schedule and schedules for clinical assessment in neuropsychiatry in the Baltimore Epidemiologic catchment area follow-up. *Arch Gen Psychiatry* 2000; 57:217–22.
- Helzer JE, Robins LN, Mcevoy LT, Spitznagel EL, Stoltzman RK, Farmer A, Brockington IF. A comparison of clinical and diagnostic interview schedule diagnoses: re-examination of lay-interviewed cases in the general population. *Arch Gen Psychiatry* 1985; 42: 657–66.
- Ihaka R, Gentleman RR. A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* 1996; 5(3): 299–314.
- Murphy JM, Laird NM, Monson RR, Sobol AM, Leighton AH. A comparison of diagnostic interviews for depression in the Stirling County Study: challenges for psychiatric epidemiology. *Arch Gen Psychiatry* 2000; 57: 230–6.
- Murphy JM, Monson RR, Laird NM, Sobol AM, Leighton AH. Identifying depression and anxiety in a forty-year epidemiological investigation: the Stirling County Study. *Int J Methods Psychiatr Res* 1998; 7: 89–109.
- Murphy JM, Neff RK, Sobol AM, Rice JX, Olivier DC. Computer diagnosis of depression and anxiety: the Stirling County Study. *Psychol Med* 1985; 15: 99–112.
- Narrow WE, Rae DS, Robins LN, Regier DA. Revised prevalence estimates of mental disorders in the United States. *Arch Gen Psychiatry* 2002; 59: 115–123.
- Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. *Arch Gen Psychiatry* 1981; 8: 381–9.
- Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. The

Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998; 59 Suppl 20: 22–33; quiz 34–57.

Correspondence: Jean-Yves Loze, Centre Hospitalier Universitaire Albert Chenevier et Henri Mondor Service de Psychiatrie, 40, rue de Mesly, 94000 Creteil, France.

Telephone: (+33) 1 49 81 30 51.

Fax: (+33) 1 49 81 30 59.

Email: jean-yves.loze@wanadoo.fr.