BRIEF REPORT

Accuracy of psychiatric diagnosis performed under indirect supervision

Precisão do diagnóstico psiquiátrico realizado sob supervisão indireta

Cristina Marta Del-Ben,[1] Jaime Eduardo Cecílio Hallak,[1] Alcion Sponholz Jr,[1] João Mazzoncini de Azevedo Marques,[1] Cybelli Morelo Labate,[1] José Onildo Betioli Contel,[1] Antonio Waldo Zuardi[1]

Original version accepted in English

Abstract

This work aimed at comparing the accuracy of the psychiatric diagnoses made under indirect supervision to the diagnoses obtained through Structured Clinical Interview for DSM-III-R (SCID). The study was conducted in 3 university services (outpatient, inpatient and emergency). Data from the emergency service were collected 3 years later, after changes in the training process of the medical staff in psychiatric diagnosis. The sensitivity for Major Depression (outpatient 10.0%; inpatients 60.0%, emergency 90.0%) and Schizophrenia (44.4%; 55.0%; 80.0%) improved over time. The reliability was poor in the outpatient service (Kw = 0.18), and at admission to the inpatient service (Kw = 0.38). The diagnosis elaborated in the discharge of the inpatient service (Kw = 0.55) and in the emergency service (Kw = 0.63) was good. Systematic training of supervisors and residents in operational diagnostic criteria increased the accuracy of psychiatric diagnoses elaborated under indirect supervision, although excellent reliability was not achieved.

Keywords: *Mental disorders/diagnosis; Depression/diagnosis; Schizophrenia/diagnosis; Emergency services, psychiatric; Psychiatric status rating scales; Interview, psychological*

Resumo

O objetivo deste estudo foi comparar a precisão do diagnóstico psiquiátrico elaborado sob supervisão indireta com o diagnóstico obtido por meio da Entrevista Clínica Estruturada para o DSM-III-R (SCID). O estudo foi realizado em três serviços universitários (ambulatório, enfermaria e emergência). Os dados do serviço de emergência foram colhidos três anos mais tarde, após mudanças no treinamento da equipe médica em diagnóstico psiquiátrico. A sensibilidade do diagnóstico de Depressão Maior (ambulatório 10,0%; enfermaria 60,0%, emergência 90,0%) e de Esquizofrenia (44,4%; 55,0%; 80,0%) aumentou com o passar do tempo. A concordância diagnóstica foi insatisfatória no serviço ambulatorial (Kw = 0,18) e na admissão da enfermaria (Kw = 0,38), mas satisfatória na alta da enfermaria (Kw = 0,55) e na emergência psiquiátrica (Kw = 0,63). O treinamento sistemático de supervisores e médicos residentes em critérios diagnósticos e entrevistas estruturadas contribuiu para uma maior precisão do diagnóstico elaborado sob supervisão indireta, embora níveis excelentes de confiabilidade não tenham sido alcançados.

Descritores: *Transtornos mentais/diagnóstico; Depressão/diagnóstico; Esquizofrenia/diagnóstico; Serviços de emergência psiquiátrica; Escalas de graduação psiquiátrica; Entrevista psicológica*

[1] Department of Neurology, Psychiatry and Medical Psychology, Clinical Hospital of the Faculty of Medicine of Ribeirão Preto, São Paulo University, Brazil

Correspondence
Cristina Marta Del-Ben
Departamento de Neurologia, Psiquiatria e Psicologia Médica
Faculdade de Medicina de Ribeirão Preto da Universidade de S. Paulo
Av. Bandeirantes, 3900
14049-900 Ribeirão Preto, SP, Brasil
Phone: (16) 602-2533 Fax: (16) 602-2544
E-mail: delben@fmrp.usp.br

Financing and conflict of interests: non-existent
Submitted: 03 December 2003
Accepted: 07 October 2004

Introduction

The elaboration of clinical diagnosis and therapeutic plan under indirect supervision is still a common practice in academic services. In such situations, it is taken for granted that the supervisors know the diagnostic process well enough to compensate for possible mistakes made in the collection, interpretation and organization of the information, and then elaborate a diagnosis even without examining the patient.[1]

The supervision of reports may save time, as various case reports can be presented in the same supervision session. However, this method has been criticized because it relies only in observations made by students with relatively little training.[2] The theoretical formulations and clinical interventions are proposed based on reports of unskilled professionals, without a direct evaluation made by an experienced psychiatrist. Transcribed or reported data eliminate one of the most important sources of evaluation, which is the observer's perception of the patient's psychopathological alterations.[3]

Some studies[1,4] have suggested that the psychiatric diagnosis made under indirect supervision presents low reliability levels. Patients' ratings made by residents under indirect supervision, as compared to supervisors' direct ratings, are usually associated to less accurate diagnoses, lower engagement in the treatment and higher difficulty in managing more complex clinical situations.[5-6]

The present study aimed at verifying the accuracy of the psychiatric diagnosis obtained under indirect supervision by measuring the reliability, sensitivity and specificity of psychiatric diagnosis elaborated through indirect supervision, in different psychiatric assistance services, as compared to the diagnosis obtained using semi-structured interviews.

Methods

1. Subjects

Patients were selected from three academic services: a psychiatric outpatient clinic, a psychiatric ward in general hospital and a psychiatric emergency setting.

Thirty-five subjects were recruited among the patients being followed-up in the outpatient service during a period of six months. Patients were selected by a psychiatrist not enrolled in the performance of the semi-structured interviews and with no previous contact with them. Sixty-one patients who were consecutively admitted along a period of six months composed the sample of patients from the infirmary. In the psychiatric emergency setting, the collection data started three years later than in the others services. The sample was made of 40 patients attended along a period of 21 months.

The local ethics committee approved this study and a written informed consent was obtained from each volunteer and his/her relative.

Supervision diagnosis

Medical residents with little clinical experience (first or second year students) elaborated a psychiatric clinical history, from information obtained through non-structured interviews. The diagnostic impressions came up under indirect supervision, in discussions or clinical meetings with the academic staff from the HCFMRP-USP. Medical residents and supervisors were aware of this study.

In the outpatient service, the supervisor had no systematic training whatsoever in the employment of operational diagnostic criteria and in rating instruments, although his/her theoretical basis was the diagnostic classification proposed by the American Psychiatric Association.[7-8] In the inpatient service, the supervisors made their diagnoses in a more systematic way, using the diagnostic classification proposed by DSMs,[7-8] but only one supervisor out of four, was trained in the usage of the semi-structured interview. In the emergency service, all of the four supervisors involved in this study were trained in the application of the structured diagnostic interview and used the diagnostic criteria proposed by the DSM[7-8] for the elaboration of the supervision diagnostic.

Semi-structured interview

The supervision diagnosis was compared with the diagnosis obtained through the "Structured Clinical Interview for the DSM-III-R" patient's version (SCID-P),[9] translated and adapted to Portuguese.[10] In the three services, the application of SCID-P was done by a group of five psychiatrists who were familiar with the DSM-III-R and trained in the instrument application, with good agreement indexes among themselves in the use of the SCID-P.[10] In the three situations, the raters had no information about the patient before the interview application.

Data analysis

The agreement between the diagnosis made under supervision and that by the SCID-P was calculated through the Kappa coefficient.[11-12] Values of $p < 0.05$ were considered significant. It was also assessed an overall index of reliability, through weighted Kappa (Kw),[13] where weight is distributed for the disagreements. When the specific diagnostic categories obtained by the indirect supervision and by the SCID-P were different, and did not belong to the same major diagnostic category, it was considered as a major disagreement (weight 2); when they belonged to the same major category, as a minor disagreement (weight 1). For a qualitative analysis, Kappa's value equal or higher than 0.75 were considered excellent reliability, Kappa's value between 0.40 and 0.74, good reliability and Kappa's value equal or lower than 0.39, poor reliability.[11]

The accuracy of the supervision diagnosis was assessed, by estimating its sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) as compared to the diagnosis obtained with the SCID-P, considered the gold standard.

Results

The sample was composed by 136 patients with an average of 35.7 (SD = 14.2) years of age, 64 of them (47.1%) being male.

Table 1 presents the reliability between the diagnosis elaborated under indirect supervision and the one obtained through semi-structured interviews, for specific diagnostic categories and the reliability of global analysis (weighted Kappa). The global reliability was good at discharge and in the emergency service, but poor in the hospital admission and in the outpatient service.

In the outpatient service the reliability was significant only for Bipolar Disorder. In the psychiatric ward, the reliability indexes were significant in most of the categories

Table 1 – Agreement (Kappa) between the diagnosis elaborated under indirect supervision and the main diagnosis obtained through the application of SCID-P, for specific diagnostic categories (n = 136)

	Outpatients (N = 35)		Inpatients (N = 61)				Emergency (N = 40)	
			Admission		Discharge			
	K	P	K	p	K	p	K	P
Major Depression	0.13	NS	0.83	<0.001	0.65	<0.001	0.65	<0.001
Bipolar Disorder	0.69	<0.001	0.54	<0.001	0.61	<0.001	0.64	<0.001
Brief Psychotic Disorder	–		-0.04	NS	0.27	0.001	–	
Schizophrenia	0.34	NS	0.48	<0.001	0.65	<0.001	0.73	<0.001
Schizoaffective Disorder	-0.10	NS	0.35	0.003	0.52	<0.001	–	
Substance related disorders	–		–		–		0.53	<0.05
Weighted Kappa	0.18	NS	0.38	<0.001	0.55	<0.001	0.63	<0.001

K = Kappa coefficient; NS = non significant

assessed, except for the Brief Psychotic Disorder diagnosis, elaborated at the patient's admission. In the qualitative analysis, only the Major Depression diagnosis made at admission showed an excellent agreement, while the diagnoses of Schizoaffective Disorder at admission and the Brief Psychotic Disorder, both at admission and discharge presented a poor reliability.

In the emergency service, the reliability was significant and good for Major Depression, Bipolar Disorder, Schizophrenia and Psychoactive Substance Related Disorders.

The frequency of specific diagnostic categories in the three services under study are shown in the Table 2 and the sensitivity, specificity, positive predictive value (PPV) and negative predictive values (NPV) of the diagnosis performed under indirect supervision against a SCID-P diagnosis are shown in Table 3. The sensitivity of Bipolar Disorder diagnosis was moderate in the emergency service, but lower than in the others services, with a high specificity in all services.

Table 2 – Frequency of the diagnosis obtained through the application of SCID-P (n = 136) in three different services (outpatient, inpatient and emergency)

	Outpatients (N = 35) %	Inpatients discharge (N = 61) %	Emergency patients (N = 40) %
Major Depression	28.6	16.4	25.0
Bipolar Disorder	11.4	23.0	10.0
Schizophrenia	28.6	29.5	25.0

On the other hand, the PPV of the diagnosis of Bipolar Disorder was moderate in the outpatients and hospital unit, but high in the emergency setting. The Major Depression diagnosis had a low sensitivity, combined with a high specificity in the outpatient service, moderate sensitivity and high specificity, at the hospital unit and higher levels of both measures in the emergency service. The emergency diagnosis of Major Depression showed the lowest PPV. Schizophrenia diagnosis held moderate sensitivity and high specificity in outpatient and in the psychiatric ward. The highest sensitivity and negative predictive value associated with a high specificity for Schizophrenia diagnosis was taken in the emergency room.

Discussion

The results obtained in this study show that there was a good reliability of the diagnosis performed under indirect supervision at discharge and in the emergency service, whereas at admission to the psychiatric ward and in the outpatient service the supervision diagnosis held a poor reliability.

Among inpatients, the longitudinal follow-up during a hospitalization period of around 45 days may have contributed for the improvement in the diagnosis accuracy. The discharge diagnosis reflects, in fact, the process of various successive supervisions, which allows the attainment of a large volume of information, from different sources, as well as the observation of the patient's evolution. This should be the ideal condition for the establishment of a "gold standard" for the psychiatric diagnosis validity.[14]

The psychiatric diagnosis elaborated in an emergency service presents a series of shortcomings inherent to the situation. In general, time is brief, frequently without additional information from relatives, and in most of the cases, there is a need of immediate intervention.[15-16] In

Table 3 – Sensitivity, specificity, Positive Predictive Value (PPV) and negative Predictive Value (NPV) of the psychiatric diagnosis elaborated under indirect supervision compared with the main diagnosis obtained through the application of SCID-P (n = 136)

		Major Depression	Bipolar Disorder	Schizophrenia
Outpatients (n = 35)	Sensitivity (%)	10.00	75.00	44.44
	Specificity (%)	100.00	96.77	88.46
	PPV (%)	100.00	75.00	57.14
	NPV (%)	73.53	96.77	82.14
Inpatients discharge (n = 61)	Sensitivity (%)	60.00	64.29	55.00
	Specificity (%)	96.08	91.50	97.56
	PPV (%)	81.82	69.23	75.00
	NPV (%)	98.00	87.50	81.63
Emergency patients (n = 40)	Sensitivity (%)	90.00	50.00	80.00
	Specificity (%)	83.33	100.00	93.33
	PPV (%)	64.29	100.00	80.00
	NPV (%)	96.15	94.74	93.33

PPV = positive predictive value; NPV = negative predictive value

spite of these restrictions, in this study the levels of reliability, sensitivity, specificity and predictive values obtained in the emergency service was mostly good and comparable to the ratios obtained at discharge.

A reason that could justify the lowest validity indexes obtained in the outpatient service may be the supervisors' training. As mentioned before, the supervisors' team of this service did not have, at the time of the data collection, systematic training in the application of operational diagnostic criteria, whereas in the other services, though in different ways, the supervisors were acquainted with these diagnostic criteria.

Another fact that may have contributed for the diagnostic accuracy improvement is the training given to the residents in the application of the diagnostic criteria proposed by the American Psychiatric Association. The data concerning the emergency diagnosis reliability were collected some years after the data collection in the outpatient and inpatient services. Along this interval there were some changes in the aims of the medical residence program in Psychiatry at the institution, with the adoption of descriptive diagnosis based on current diagnostic classifications as a routine in the process of diagnosis elaboration.

Our data suggest that systematic training in operational diagnostic for both supervisors and medical residents could improve reliability and validity of psychiatric diagnosis elaborated under indirect supervision. Nevertheless, it has not been enough for the attainment of excellent agreement indexes, what may be related to drawbacks inherent to the indirect supervision, as discussed before.[1,3-4,17] The use of recorded interviews made by experienced professionals, observation of interviews through unidirectional mirror, and the practice of joint interviews, for example, may minimize the discrepancies between the diagnosis elaborated by the professional-to-be and the experienced professional.

Acknowledgements

The authors are very grateful to Dr Serdar Dursun, Dr Luciana Del Ben and Dr Frederico Graeff for their comments and suggestions in the revision of the manuscript.

References

1. Spitzer RL, Skodol AE, Williams JB, Gibbon M, Kass F. Supervising intake diagnosis. A psychiatric "Rashomon". Arch Gen Psychiatry. 1982;39(11):1299-305.
2. Skodol AE, Williams JB, Spitzer RL, Gibbon M, Kass F. Identifying common errors in the use of DSM-III through diagnostic supervision. Hosp Community Psychiatry. 1984;35(3):251-5.
3. Kendell RE, Cooper JE, Gourlay AJ, Copeland JR, Sharpe H, Gurland BJ. Diagnostic criteria of American and British psychiatrists. Arch Gen Psychiatry. 1971;25(2):123-30.
4. Hyler SE, Williams JBW, Spitzer RL. Reliability in the DSM-III field trials: interview v case summary. Arch Gen Psychiatry. 1982;39(11):1275-8.
5. Schuster DB, Freeman EN. Supervision of the resident's initial interview. Arch Gen Psychiatry. 1970;23(6):516-23.
6. Jaynes S, Charles E, Kass F, Holzman S. Clinical supervision of the initial interview: effects on patient care. Am J Psychiatry. 1979;136(11):1454-7.
7. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd ed. revised, Washington, DC: APA; 1987.
8. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed. Washington, DC: APA; 1994.
9. Sptizer RL, Williams JR, Gibbon M, First MB. Structured clinical interview for DSM-III-R - patient edition (SCID-P, version 1.0). Washington, DC:American Psychiatric Press; 1990.
10. Del-Ben CM, Rodrigues CRC, Zuardi AW. Reliability of the Portuguese version of the structured clinical interview for DSM-III-R (SCID) in a Brazilian sample of psychiatric outpatients. Braz J Med Biol Res. 1996;29(12):1675-82.
11. Fleiss JL. The measurement of interrater agreement. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley & Sons; 1981.
12. Bartko JJ, Carpenter WT Jr. On the methods and theory of reliability. J Nerv Ment Dis. 1976;163(5):307-17.
13. Cohen J. Weighted Kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bul. 1968;70(3):213-20.
14. Spitzer RL, Williams JR, Gibbon M, First MB. Instruction manual for the structured clinical interview for DSM-III-R (SCID,5/1/89 Revision). New York, New York State Psychiatric Institute; 1989.

15. Oldham JM, Lin A, Breslin L. Comprehensive psychiatric emergency services. Psychiatr Q. 1990;61(1):57-67.
16. Segal SP, Egley L, Watson MA, Miller L, Goldfinger SM. Factors in the quality of patient evaluations in general hospital psychiatric emergency services. Psychiatr Serv. 1995;46(11):1144-8.
17. Stein PS, Karasu TB, Charles ES, Buckley PJ. Supervision of the initial interview. A study of two methods. Arch Gen Psychiatry. 1975;32(2):265-8.