

Validation of PSE / CATEGO Computer Programme on Arabic Population

A. Okasha, A. Sadek and M. Abdel-Mawgoud

ABSTRACT

The validity of the PSE / CATEGO computer programme is assessed on a sample of new 100 psychiatric Arabic patients. The CATEGO based ICD-9 diagnoses were compared to the consensus clinical ICD-9 diagnoses. A "good" overall agreement with Kappa = + 0.71 was achieved for the 3-digit ICD-9 diagnoses. Most disorders have a very good agreement except for paranoid disorder (Kappa = + 0.71) and neurotic disorders (Kappa = + 0.71). The Index of Definition (ID) suggested cut off point of 5 and above was found to be too high for our sample as it was likely to undetect 10% of cases. The results are discussed with the suggested changes in the CATEGO programme to suit Arabic patients.

INTRODUCTION

Computer technology is being applied to practically all aspects of psychiatry. Theories of personality and of psychotherapeutic changes are being subjected to computer simulation techniques, Computer interview, diagnosis and conduct therapy with real patients. Computers interpret and write psychological test reports. They compose reports of mental state examination (Spitzer and Endicott, 1975)

The well known limitations of the clinical methods for arriving at psychiatric diagnosis, especially its reliability, have led to several efforts to utilize computers for integrating clinical observations into psychiatric diagnoses.

There are several advantages to computer generated diagnoses. First of all, there is the value of necessarily perfect reliability. Secondly, the computer programme can utilize rules developed for a large and more diverse sample of actual patients than any single clinician can command. The rules by which the computer assigns diagnoses are explicit and public. Finally, empirically based rules constitute at least potential advances in our scientific understanding of the complex relationship between symptom characteristics and diagnosis.

Three different kinds of computer programmes have been used for generating diagnoses. The first is based on a logic decision tree such as that of Spitzer's DIAGNO (Spitzer and Endicott, 1968) or Wing's CATEGO (Wing *et al.* 1974). The second is the probabilistic method (Birnbbaum and Morrel, 1961; Overall and Gorham, 1963; Overall and Hollister, 1964; Smith, 1966). The third is based on the discriminant functional analysis (Fischer, 1963; Rao, 1984).

The three methods were tested together (Fleiss *et al.* 1972) and the authors concluded that the logic decision tree is preferable for the computer diagnoses.

Though based on the same principle as the DIAGNO, the PSE / CATEGO seems to have many advantages. First it is based on data collection by the PSE semi-structured psychiatric interview.

Secondly, instead of eliminating single diagnoses one by one as in DIAGNO, the CATEGO process the original 140 PSE items through a progressive series of condensations and all decisions about actual diagnoses are postponed until the final stages. The CATEGO prints out all symptoms, syndromes and descriptive categories exhibited by each individual patient, together with rough 3 point scale of certainty

(?+and++) for each. In addition, it prints the final professional diagnostic CATEGO subclass and CATEGO class (Wing *et al.* 1974; Kendell, 1975).

The potential of PSE / CATEGO system has been well illustrated in the large IPPS (World Health Organisation, 1973) and the US-UK (Cooper *et al.* 1972) trials. It has been used in many languages including Arabic (Leighton *et al.* 1963; Orley and Wing, 1977; Gatere, 1980; Okasha and Ashour, 1981; Gillis *et al.* 1982; Al-Khani *et al.* 1986).

In spite of its use in Arabic language, no attempts (according to the author's knowledge) have been made to validate its application on Arabic population.

METHODOLOGY

A series of the serial new 100 psychiatric referrals to the adult general psychiatric unit, both inpatients and outpatients, in Bahrain were assessed by one of the authors (M.A.M), who is trained for the use of PSE in UK. The English-Arabic version of the Present State Examination (Abdel-Mawgoud, 1986) was used for the assessment of patients.

The PSE scores for all patients were sent to the Institute of Psychiatry in London, for analysis using the CATEGO computer programme. The print out of the CATEGO analysis was used along with the Aetiology Schedule (AS) scores to allocate patients to a CATEGO based 3-digit ICD-9 diagnosis.

Meanwhile, an abstract form was designed to include all PSE scores, AS scores, duration of illness and patient's age and sex. The abstract forms, were then used by 3 raters to make an independent 3-digit clinical diagnosis according to the 9th revision of the International Classification of Diseases (ICD-9) (W.H.O., 1977). If the clinical diagnoses were identical for the 3 raters, the agreed upon diagnoses became the patient's "consensus clinical diagnosis". If the clinical diagnosis differs (this accounts for only 14 % of cases), a conference including all 3 raters

was held to arrive at the "consensus clinical diagnosis".

Using the "consensus clinical diagnosis" as a gold standard (Zheng *et al.* 1988), the CATEGO based diagnosis was compared to the "consensus clinical diagnosis" and Kappa values were calculated to measure the validity of the PSE / CATEGO system for each diagnostic category as well as the overall validity of the system (Cohen, 1960).

Kappa is the most frequently used measure of inter-class agreement because it adjusts for agreement that would occur by chance alone, given the distribution in diagnostic categories (Krippendorff, 1970; Fleiss, 1975; Fleiss, 1981). Its value ranges from -1 to +1, a higher value represents a better level of agreement. A Kappa score of 0.41-0.60 is considered to be a moderate degree of agreement, 0.61 - 0.80 is a good or substantial agreement and above 0.80 is a very good or almost perfect agreement (Landis and Kock, 1977). The only criticism for Kappa is the fact that it is base dependent, i.e. its accuracy is likely to be reduced in cases of rare syndromes; so interpretations of the results of disorders below 5 % in the sample should be taken cautiously (Spitznagel and Helzer, 1985).

The Index of Definition (ID) scores were also analysed to determine the appropriate cut off point for Arabic population.

RESULTS

Our 100 patients sample was found to represent 7 major categories of 3-digit ICD-9 diagnosis: 1) Dementia; 2) Schizophrenia; 3) Affective disorders; 4) Paranoid disorder; 5) Neurotic disorders; 6) Adjustment disorder and 7) Mental retardation. The number of patients allocated to each of these categories according to the consensus clinical diagnosis as well as the CATEGO based diagnosis are displayed in (table 1). Two more categories are also included, namely, "No" and "Other psychiatric diagnoses".

TABLE I
Validity of CATEGO Diagnosis (3 -Digit)

Code	Diagnosis	Clinical ICD-9 Diagnosis (No)	Catego Based Diagnosis (N0)	Kappa Correlation
0	No Diagnosis	1	1	+ 1.0
290	Dementias	2	2	+ 1.0
295	Schizophrenic D.	18	21	+ 0.90
296	Affective D.	56	52	+ 0.71
297	Paranoid D.	2	8	+0.17
300	Neurotic D.	4	3	+0.26
309	Adjustment D.	11	8	+0.70
319	Mental Retardation	2	2	+ 1.0
----	Others	4	3	+0.55
	Total	100	100	+ 0.71

TABLE II
Clinical 3-Digit ICD - 9 Diagnosis by ID SCORE

Code	Diagnosis	ID Score							Total
		2	3	4	5	6	7	8	
0	No Diagnosis	1	0	0	0	0	0	0	1
290	Dementias	0	0	1	0	1	0	0	2
295	Schizophrenic D.	0	0	0	0	1	1	16	18
296	Affective D.	0	0	0	0	10	29	17	56
297	Paranoid D.	0	0	0	0	1	0	1	2
300	Neurotic D.	1	0	0	1	2	0	0	4
309	Adjustment D.	1	1	2	4	2	1	0	11
319	Mental Retardation	1	1	0	0	0	0	0	2
----	Others	0	2	0	1	1	0	0	4
	Total	4	4	3	6	18	31	34	100

The CATEGO computer program was found to have an overall "good" validity for our sample with a Kappa of +0.71. Dementia, Mental Retardation and No diagnosis have a perfect Kappa of +1.0. Schizophrenic disorder has an excellent agreement with Kappa of +0.90 while Affective and Adjustment disorders have a good agreement with Kappa of +0.71 and +0.70 respectively. Only two categories, Paranoid and Neurotic disorders, have a very low Kappa of +0.17 and 0.26 respectively.

Table 2 shows the results of these 9 major categories along with their ID scores. Scores of 5 and above as suggested by Wing (Wing, 1983), would only allow the detection of 89 out of the 99 possible psychiatric cases. The cases that are likely to be undetected are those of Dementia, Mental retardation, Adjustment disorders as well as other psychiatric diagnoses.

DISCUSSION

The validity of instrument always relies on some acceptable "gold standard" against which the proposed test can be evaluated. In this study, our gold standard is the consensus clinical diagnosis of 3 senior psychiatrists. They all used the standard ICD-9 criteria of diagnosis to be able to compare it to the CATEGO based ICD-9 diagnosis (Zheng *et al.* 1988).

Using the consensus clinical diagnosis as a gold standard, the overall validity of the CATEGO programme for Arabic patients was found to be very good with a Kappa of +0.71. This is higher than the results achieved by Zheng and his colleagues, using the same gold standard, in their validation of Chinese computer programme when they achieved a Kappa of +0.47 (Zheng *et al.* 1988).

As Kappa is known to lose its accuracy for rare syndromes (Spitznagel and Helzer, 1985), interpretation of the results for the validity of disorders such as dementia, mental retardation, neurotic as well as other psychiatric disorders is rather

difficult. For schizophrenic disorder a Kappa of +0.90 was achieved which is comparable to those achieved by Wing and his colleagues (Wing *et al.* 1974). Affective and adjustment disorders achieved a "good" agreement with kappa of +0.71 and +0.70 respectively. The CATEGO was clearly not valid for the category of paranoid disorder with a Kappa of +0.17.

The results of our sample can improve further if the CATEGO programme is to consider slight change for Arabic population where patients with depressive illness (representing the majority of our sample) commonly expresses paranoid symptoms along with their depressive ones. This seems to be responsible for the marked discrepancy with 8 cases diagnosed as paranoid disorder in CATEGO compared to only 2 for the consensus clinical diagnosis.

As far as the ID score is concerned, our results suggest that the claimed very low cut of point of 5 and above (Wing, 1983) does not apply to our population. If applied it would lead to about 10% of cases passing undetected. The common categories to be undetected are the dementia, mental retardation, neurotic, adjustment and other psychiatric disorders. A possible explanation is the common clinical experience that Arabic patients tend to somatize their feelings which will not show on PSE score. A lower cut of point of 2 or a higher scoring for somatisation on PSE / CATEGO can help to solve this problem.

REFERENCES

- Abdel-Mawgoud, M. (1986) The English - Arabic version of PSE: An Arabic translation of the full version of Present State Examination (PSE). *Psychiatric Hospital*, Bahrain.
- Al-Khani, M. A. F., Bebbington, P. E., Watson, J. P. and House, F. (1986) Life events and schizophrenia: A Saudi Arabian study. *Brit. J. Psychiat.*, **148**: 12-22.

- Birnbaum, A. and Maxwell, A. E. (1961) Classification procedures based on Bayes formula. *Psychiatry*, **1**: 233-236.
- Cohen, J. (1960) A Co-efficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**: 37-46.
- Cooper, J. E., Kendell, R. E., Gurland, B. J., Sharpel, Copeland, J. R. M. and Simon, R. (1972) *Psychiatric Diagnoses in New York and London: A Comparative Study of Mental Hospital Admissions. Maudsley Monographs*, Institute of Psychiatry, London: Oxford University Press.
- Fischer, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**: 179-184.
- Fleiss, J. L., Spitzer, R. L., Cohen, J. and Endicott, J. (1972) Three computer diagnosis methods compared. *Arch. Gen. Psychiat.*, **27**: 643-649.
- Fleiss, J. L. (1975) Measuring agreement between two judges on the presence or absence of a trait. *Biometrics*, **31**: 651-659.
- Fleiss, J. L. (1981) *Statistical Methods for Rates and Proportions*. Chapter 13: 212-236.
- Gatere, M. (1980) M. Phil, Thesis, University of London.
- Gillis, L. S., Elk, R., Ben-Arie, O. and Taggin, A. (1982) The present state examinations: Experiences with Xhosa-Speaking psychiatric patients. *Brit. J. Psychiat.*, **141**: 143-147.
- Kendell, R. E. (1975) *The Role of Diagnosis in Psychiatry*. U. K.: Blackwell Scientific Publications.
- Krippendorff, K. (1970) Bivariate agreement coefficients for reliability of data. In Borgatta, E. F. (ed.), *Sociological Methodology*. San Francisco: Jossey-Bass.
- Landis, J. R. and Koch, G. G. (1977) The measurement of observer agreement for Categorical data. *Biometrics*, **33**: 159-174.
- Leighton, A. H., Lambo, T. A., Hughes, C. C., Leighton, D. C., Murphy, O. M. and Macklin, D. B. (1963) *Psychiatric Disorders among the Yoruba*. New York: Cornell University Press.
- Okasha, A. and Ashour, A. (1979) *Arabic Version of PSE*. (in arabic) Cairo: Ain Shams University.
- Orley, J. and Wing, J. K. (1979) Psychiatric disorders in two African villages. *Arch. Gen. Psychiat.*, **36**: 513-520.
- Overall, J. E. and Gorham, D. R. (1963) A pattern probability model for the classification of psychiatric patients. *Behavioral Science*, **8**: 108-116.
- Overall, J. E. and Hollister, L. E. (1964) Computer procedures for psychiatric classification. *J. of the Amer. Med. Assoc.*, **187**: 583-588.
- Rao, C. A. (1984) The utilization of multiple measurements in problems of biological classification. *J. of the Royal Statist. Soci. (Series B)* **10**: 159-193.
- Smith, W. G. (1966) A model of psychiatric diagnosis. *Arch. Gen. Psychiat.*, **14**: 521-529.
- Spitzer, R. L. and Endicott, J. (1968) Diagnostic-computer Program for psychiatric diagnosis utilizing the differential diagnostic procedure. *Arch. Gen. Psychiat.*, **18**: 746-756.
- Spitzer, R. L. and Endicott, J. (1975) *Amer. Handbook of Psychiatry*. 2nd edition, Chapter 36. Computer Applications in Psychiatry, edited by S. Arieti.
- Spitznagel, E. L. and Helzer, J. E. (1985) A Proposed solution to the base rate problem in the Kappa Statistic. *Arch. Gen. Psychiat.*, **42**: 725-728.
- Wing, J. K., Cooper, J. E. and Sartorius, N. (1974) The measurement and classification of psychiatric symptoms. *An Instruction Manual for PSE and Catego Program*. Cambridge University Press. U.K.
- Wing, J.K. (1983) Use and misuse of the PSE. *Brit. J. Psychiat.*, **143**: 111-117.
- World Health Organization (1973) *The International Pilot Study of Schizophrenia (IPPS)*. Volume I. Geneva: WHO.
- World Health Organization, (1977) *International Statistical Classification of Diseases, Injuries and Causes of death, (ICD- 9)*. Geneva: WHO.
- Zheng, Y., Yang, W., Phillips, M.R., Dai, C. and Zheng, H. (1988) Reliability and validity of a Chinese computerized diagnostic instrument. *Acta Psychiat. Scand.*, **77**: 32-37.

AUTHORS

A. Okasha, F.R.C.P., F.R.C. Psych., D.P.M., Professor and Head of Department of Psychiatry, Ain Shams University, Egypt.

A. Sadek, M.D., F.R.C. Psych., D.P.M., Professor of Psychiatry, Ain Shams University, Egypt.

M. Abdel-Mawgoud, M.R.C. Psych., M.D., Consultant and Clinical Tutor, Psychiatric Hospital, Bahrain.

ABSTRAIT

Validation de PSE / CATEGO Programme de L'ordinateur sur la Population Arabe

La validité de PSE / CATEGO programme d'ordinateur a été évaluée à travers un échantillon de 100 patients arabes nouveaux.

Les CATEGO basé sur le ICD - 9 diagnostics étaient comparés au consensus diagnostics cliniques de ICD - 9.

Une "belle" conversion avec kappa = + 0,71 été accomplie pour les 3-chiffres diagnostics de ICD - 9. La plupart des désordres avaient une très belle conversion sauf le désordre paranoïque (Kappa = + 0,71) et les désordres neurotiques (Kappa = + 0,71).

L'index de Définition (ID) suggéré l'idée de couper 5 et plus a été estimé très haut pour notre échantillon car il était probable de ne pas détecter 10% des cas. Les résultats ont été discutés avec les échanges suggérés dans le CATEGO programme pour approprier les patients arabes.

الموجز

تقييم درجة صدق "فحص الحالة الراهنة / برنامج كاتيجو الكمبيوترى" على عينة من العرب

الهدف من هذا البحث هو تقييم مدى صدق "فحص الحالة الراهنة / برنامج كاتيجو الكمبيوترى" عند تطبيقه على عينة تتكون من ١٠٠ عربيا يعانون من أمراض نفسية. ولمعرفة ذلك تمت مقارنة التشخيصات المؤسسة على التقسيم العالمى التاسع المبينة على الكاتيجو من ناحية، وتشخيصات نفس التقسيم المبينة على الاتفاق الإكلينيكى. ولقد أظهرت النتائج إجمالاً درجة جيدة من الاتفاق حسب إحصاء كابا (+ ٠,٧١) وذلك بالنسبة لتشخيصات التقسيم العالمى التاسع بتقريباتها الثلاث. وتفضيلاً فإن معظم الإضطرابات النفسية، فيما عدا الإضطرابات البارانونية والإضطرابات العصبية، أظهرت درجة جيدة جداً من الاتفاق. ولقد نوقشت النتائج من اقتراح التعديلات فى برنامج كاتيجو لكى تناسب المرضى العرب.