Using consensus OPCRIT diagnoses

An efficient procedure for best-estimate lifetime diagnoses*

M. H. AZEVEDO, M. J. SOARES, I. COELHO, A. DOURADO, J. VALENTE, A. MACEDO, M. PATO and C. PATO

Background The Operational Criteria Checklist (OPCRIT) generates diagnoses according to 12 operational diagnostic systems (e.g. DSM–III, DSM–III–R, Research Diagnostic Criteria, ICD–10).

Aims To examine the agreement between diagnoses generated by the OPCRIT, as completed by the interviewer, with a best-estimate lifetime procedure using the OPCRIT.

Method Subjects came from large multi-generational bipolar or schizophrenia pedigrees (n=100), and from a sample of unrelated subjects with schizophrenia (n=40). We analysed the diagnostic agreement between OPCRIT diagnoses generated by the interviewer and our best-estimate OPCRIT diagnoses, according to DSM–III–R and ICD–10, using Cohen kappa statistics.

Results Excellent agreement was found between interviewer OPCRIT diagnoses and OPCRIT diagnoses made by the best-estimate lifetime consensus procedure for DSM–III–R (κ=0.83) and ICD–10 (κ=0.81).

Conclusions Results suggest that this procedure for diagnostic assessment is an efficient alternative to classic best-estimate diagnosis procedures.

Declaration of interest Supported by grants from Junta Nacional de Investigacão Científica e Technológia and the US National Institute of Mental Health.

*Presented at the Sixth World Congress on Psychiatric Genetics, Bonn, Germany, 7–10 October 1998.

The standard procedure for genetic studies is to use a best-estimate lifetime diagnosis (Leckman *et al*, 1982). This method can be complex, time-consuming, expensive, and relatively unstructured. We have attempted to both structure and streamline the process using the Operational Criteria Checklist (OPCRIT) system, an operational criteria checklist for psychotic illness (McGuffin *et al*, 1991). It is a polydiagnostic instrument that generates diagnoses according to 12 operational diagnostic systems (e.g. DSM–III (American Psychiatric Association, 1980); DSM–III–R (American Psychiatric Association, 1997); Research Diagnostic Criteria (RDC; Spitzer *et al*, 1978); ICD–10 (World Health Organization, 1992)). OPCRIT-generated diagnoses have excellent agreement with the currently accepted 'gold standard' of consensus best-estimate diagnoses (Craddock *et al*, 1996), and the instrument has shown good reliability (Williams *et al*, 1996). We examined the agreement between diagnoses generated by the OPCRIT system, as completed by the interviewer including all known clinical data, with a best-estimate lifetime procedure using the OPCRIT, completed from ratings produced by blind independent diagnosticians using known clinical data from the Diagnostic Interview for Genetic Studies (DIGS) database (Nurnberger *et al*, 1994).

METHOD

Subjects

The sample included 140 subjects. One hundred subjects were members of multiply affected families (69 cases of affective disorder, 29 cases of schizophrenia, two cases unaffected). The other 40 subjects were members of a sample selected from an association study and affected by psychotic disorders, mostly schizophrenia.

Assessments

The DIGS was used to assess each subject. DIGS is a polydiagnostic interview covering the Feighner (Feighner *et al*, 1972), RDC (Spitzer *et al*, 1979), modified RDC (Mazure & Gershon, 1979), DSM–III, DSM–III–R and ICD–10 diagnostic systems, and DSM–IV (American Psychiatric Association, 1994) proposed guidelines for somatisation and schizophrenia (Nurnberger *et al*, 1994). The DIGS database is designed to bring together all the known clinical data on a subject, including a complete history of illness. Furthermore, the DIGS provides a structure for carefully defining both current and lifetime symptoms, as well as course-of-illness and comorbid states information.

The clinicians who administered the DIGS to probands and relatives in our genetic and epidemiological studies of schizophrenia and bipolar mood disorder are experienced psychiatrists with extensive training in the administration of the interview. The interrater reliability of the Portuguese version of the DIGS has been tested previously. In almost all cases the agreement level is at 100%. The only areas with some variability are somatisation (96% agreement), schizotypy (95% agreement) and suicidal behaviour (96.4% agreement) (Azevedo *et al*, 1993*a*,*b*). As a rule, once the interviewer completes a thorough review of the DIGS and all other data, he or she rates the operational criteria checklist on a 'lifetime–ever' basis. The data are then entered into the OPCRIT database program and the diagnoses are generated.

Procedures

A consensus lifetime OPCRIT diagnosis was made by a group of three blind raters using the DIGS database. The blind raters had extensive experience in both the DIGS and the OPCRIT. However, they did not participate in the subject's interview and were blind to the subject's diagnosis and to diagnoses within the subject's family. Diagnostic comparisons were made with computer-generated diagnoses, using DSM–III–R and ICD–10. For all cases where there was complete agreement, the diagnoses were accepted as the final diagnoses. If disagreement was present, the two OPCRITs were compared item-by-item, and DIGS information was reviewed and discussed by all raters, in order to reach a consensus OPCRIT, which was then used to produce the final diagnosis (Fig. 1).

https://doi.org/10.1192/bjp.175.2.154 Published online by Cambridge University Press

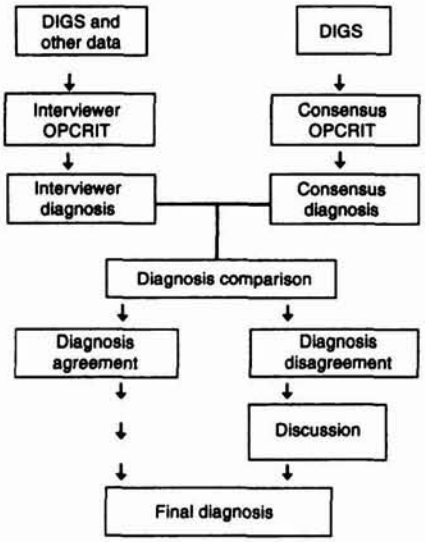


Fig. 1 Procedure used to produce the best-estimate diagnosis (DIGS, Diagnostic Interview for Genetic Studies).

The index of agreement used to measure diagnostic concordance was Cohen's kappa (κ) coefficient, expressed as a proportion of agreement corrected for chance. We divide concordance into six levels of agreement for interpreting positive κ values, following guidelines adapted from Altman (1994) and Mannuzza *et al* (1989). The concordance was poor when $\kappa \leqslant 0.20$, fair if $0.21 \leqslant \kappa < 0.40$, moderate if $0.40 \leqslant \kappa < 0.60$, good if $0.60 \leqslant \kappa < 0.75$, very good if $0.75 \leqslant \kappa < 0.81$ and excellent when $\kappa \geqslant 0.81$.

DSM–III–R and ICD–10 diagnoses were organised into seven or eight categories to facilitate analysis and to increase comparability of the two diagnostic criteria systems. This procedure was used by others (Williams *et al*, 1996) and by us in another study (Soares *et al*, 1997).

An item-by-item analysis was made in order to identify the sources of OPCRIT coding discrepancies. All diagnoses were produced by the OPCRIT computer program, version 3.31. The distribution by DSM–III–R and ICD–10 final diagnoses are shown in Table 1.

RESULTS

Agreement between interviewers' diagnoses and consensus diagnoses showed a kappa value of 0.83 for DSM–III–R and a kappa value of 0.81 for ICD–10. We also found excellent agreement between interviewer's diagnoses and final diagnoses (DSM–III–R: $\kappa=0.88$; ICD–10: $\kappa=0.88$), and when we compared final with consensus diagnoses (DSM–III–R: $\kappa=0.89$; ICD–10: $\kappa=0.92$). A marginally better concordance between consensus and final diagnosis is unlikely to be significant; it appears to be related to rare errors in coding in the OPCRIT by the interviewer (Table 2).

Table 1 Distribution of final DSM–III–R and ICD–10 diagnoses ($n=140$)

Diagnosis	*n* (DSM–III–R)	*n* (ICD–10)
Schizophrenia	66	65
Schizoaffective disorder	1	2
Other non-organic psychotic disorders	1	8
Probable schizoaffective or affective disorder with psychosis	1	–
Mild, moderate and severe depression	12	13
Mania	1	1
Bipolar disorder	24	43
Bipolar disorder with psychosis	25	–
Other affective disorders with psychosis	7	7
No diagnosis	2	1

Table 2 Measures of diagnostic agreement for DSM–III–R and ICD–10 ($n=140$)

Diagnoses compared	Diagnostic systems	Concordance observed	κ
Interviewers	DSM–III–R	123	0.83
v. consensus diagnoses	ICD–10	122	0.81
Interviewers	DSM–III–R	128	0.88
v. final diagnoses	ICD–10	129	0.88
Consensus	DSM–III–R	129	0.89
v. final diagnoses	ICD–10	132	0.92

DISCUSSION

The results indicate that it is possible to make reliable diagnoses by independent raters using the DIGS database and using the OPCRIT system rated on a lifetime basis to provide a more structured approach to the best-estimate consensus process. Moreover, excellent levels of agreement were achieved between lifetime diagnoses generated by OPCRIT on the basis of rating by one interviewer and diagnoses produced by the consensus procedure (κ coefficients vary between 0.88 and 0.89 for DSM–III–R and between 0.88 and 0.92 for ICD–10).

Limitations

However, as with all diagnostic methods in psychiatry, discrepant diagnoses may arise from the instrument, the interviewers, the clinical material rated and the method used for analysis. In this case, methodological limitations related to the kappa coefficient can be present, because the kappa coefficient is highly dependent on the prevalence of the disorder in the sample and does not permit the identification of the source of disagreement (Roy *et al*, 1997). Another limitation is the quality of input data on the OPCRIT 3.31 computer program, which can include some unexpected errors leading to inappropriate diagnoses. Kappa limitations were reduced by using diagnosis prevalence as a guide to interpreting the kappa coefficients (Table 1).

We also undertook an item-by-item analysis to identify the factors associated with OPCRIT rating disagreements. Data-entry errors were rare but represent a possible source of diagnostic disagreement. The interviewer's subjective sense of the subject and the variability in the clinical material collected can also play a role. For example, a patient may in an interview deny symptoms that are clearly documented and corroborated from other clinical sources.

A source of differences between interviewer diagnoses and consensus diagnoses may have been related to *when* the

https://doi.org/10.1192/bjp.175.2.154 Published online by Cambridge University Press

interviewer completed the OPCRIT. Initially, the interviewer would complete the OPCRIT immediately after completing the interview. This would lead to the interviewer relying on information and impressions from memory rather than a systematic review of all available data included in the DIGS. In the process of developing these consensus procedures, we identified this potential problem and corrected it. Interviewers systematically add all relevant data, as marginal comments, to the DIGS and deliberately review the finalised DIGS database in filling out the OPCRIT.

The DIGS is a semi-structured instrument designed to allow for some flexibility in the collection of data and the ability to expand beyond the strict data points. The database allows for elaborating – both in a clinical summary and in marginal comments. A detailed family history is expected to be collected, and is not limited to strict questions such as, "Do you have a family history of schizophrenia?". It is clear that, in many cases, no single data point from the DIGS maps directly to an OPCRIT rating. The interviewer is expected to note these items as a synthesis of all available data and register this information.

Related diagnoses

The ability to differentiate between a diagnosis of schizophrenia and one of schizoaffective disorder lies entirely in the assessment of the relationship between the duration of mood symptoms and the duration of psychotic symptoms. As has been noted previously (Craddock *et al*, 1996; Williams *et al*, 1996), OPCRIT item 52 becomes a single factor in defining the relationship between these two diagnoses. The limitations of the OPCRIT merely reflect the fundamental difficulty in distinguishing these two entities.

Best-estimate diagnosis

Our best-estimate diagnosis process includes both the blind consensus diagnosticians and the interviewers. We find this leads to an ongoing improvement in the quality of our interviews and in our inter-rater reliability. The knowledge of the operant relationships between DIGS questions and OPCRIT items in the final diagnostic process is also very helpful in improving the quality of the process.

CLINICAL IMPLICATIONS

- OPCRIT provides a reliable method for comparison between studies.
- OPCRIT helps structure the diagnostic process.
- OPCRIT can increase the efficiency of the best-estimate process.

LIMITATIONS

- A limited number of items is used to determine diagnoses.
- A limited number of diagnoses is included.
- Single items may determine a diagnosis.

M. H. AZEVEDO, PhD, M. J. SOARES, I. COELHO, MD, A. DOURADO, MD, J. VALENTE, MD, A. MACEDO, PhD, Department of Psychiatry, College of Medicine and Centre for Neuroscience, Coimbra University, Coimbra, Portugal; M. PATO, MD, C. PATO, MD, Department of Psychiatry, College of Medicine and Centre for Neuroscience, Coimbra University, Coimbra, Portugal, and Department of Psychiatry, State University of New York at Buffalo, Buffalo, New York, USA

Correspondence: Professor Maria Helena Azevedo, Psicologia Médica, Faculdade de Medecina, Universidade de Coimbra, Rua Larga, 3049 Coimbra Codex, Portugal. e-mail: helenamh@ci.uc.pt

(First received 30 November 1998, final revision 8 March 1999, accepted 11 March 1999)

An efficient procedure for best-estimate diagnosis

The results of this study demonstrate that the OPCRIT-based best-estimate procedure is an efficient alternative to classic lifetime best-estimate diagnostic procedures. Developed for molecular genetic research, the OPCRIT provides a reliable, rapid and valid approach to polydiagnostic assessment of psychotic and affective disorders diagnoses and has been adopted by several groups for clinical, epidemiological and biological research (Craddock *et al*, 1996). The best-estimate diagnosis process attempts to review all the available clinical and related data on a subject in order to arrive at a diagnosis. For research purposes this needs to be done in a blind fashion in order not to bias the process. However, most procedures require a structure to the diagnostic algorithm that each diagnostician follows. The OPCRIT provides a structure for this process and has been adopted by a number of groups focused on genetic research to provide a minimum set of data with which to compare different studies. The reliability of the diagnostic algorithm used by the computer package in the OPCRIT adds an additional unbiased check to the diagnostic process. We feel this represents an important addition to our research tools that in some ways helps to both structure the process and render it more efficient.

ACKNOWLEDGEMENTS

This work was supported by grants from Junta Nacional de Investigacão Científica e Technológia (Proj. No. PBIC/C/SAU/16117/92, PECS/C/SAU/222/95) and the US National Institute of Mental Health (Grant No. R01-MH52618).

REFERENCES

Altman, D. G. (ed.) (1994) *Practical Statistics for Medical Research*. London: Chapman & Hall.

American Psychiatric Association (1980) *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn) (DSM–III). Washington, DC: APA.

— **(1987)** *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn, revised) (DSM–III–R). Washington, DC: APA.

— **(1994)** *Diagnostic and Statistical Manual of Mental Disorders* (4th edn) (DSM–IV). Washington, DC: APA.

Azevedo, M. H. P., Dourado, A., Valente, J., *et al* (1993*a*) The Portuguese version of the Diagnostic Interview for Genetic Studies. *Neuropsychiatric Genetics*, **3**, 189.

https://doi.org/10.1192/bjp.175.2.154 Published online by Cambridge University Press

—, Valente, J., Macedo, A., *et al* (1993*b*) Versão Portuguesa da "Entrevista Diagnóstica para Estudos Genéticos". *Psiquiatria Clínica*, **14**, 213–217.

Craddock, N., Asherson, P., Owen, M. J., *et al* (1996) Concurrent validity of the OPCRIT diagnostic system. Comparison of OPCRIT diagnoses with consensus best-estimate lifetime diagnoses. *British Journal of Psychiatry*, **169**, 58–63.

Feighner, J., Robins, E., Guze, S., *et al* (1972) Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, **26**, 57–63.

Leckman, J. F., Sholomskas, D., Thompson, D., *et al* (1982) Best estimate of psychiatric diagnoses. *Archives of General Psychiatry*, **39**, 879–883.

Mannuzza, S., Fyer, A. J., Martin, L. Y., *et al* (1989) Reliability of anxiety assessment. *Archives of General Psychiatry*, **49**, 1093–2001.

Mazure, C. & Gershon, E. S. (1979) Blindness and reliability in lifetime psychiatric diagnosis. *Archives of General Psychiatry*, **36**, 521–525.

McGuffin, P., Farmer, A., Harvey, I., *et al* (1991) A polydiagnostic application of operational criteria in studies of psychotic illness. *Archives of General Psychiatry*, **48**, 764–770.

Nurnberger, J. I., Blehar, M. C., Kaufmann, C. A., *et al* (1994) Diagnostic Interview for Genetic Studies: rationale, unique features, and training. *Archives of General Psychiatry*, **51**, 849–859.

Roy, M.-A., Lanctôt, G., Mérette, C., *et al* (1997) Clinical and methodological factors related to reliability of the best-estimate diagnostic procedure. *American Journal of Psychiatry*, **154**, 1726–1733.

Soares, M. J., Dourado, A., Macedo, A., *et al* (1997) Estudo de Fidelidade da Lista de Critérios Operacionais Para Doenças Psicóticas. *Psiquiatria Clínica*, **18**, 11–24.

Spitzer, R. L., Endicott, J. & Robins, E. (1978) Research Diagnostic Criteria: Rationale and reliability. *Archives of General Psychiatry*, **35**, 773–782.

Williams, J., Farmer, A. E., Ackenhell, M., *et al* (1996) A multicentre inter-rater reliability study using the OPCRIT computerised diagnostic system. *Psychological Medicine*, **26**, 775–783.

World Health Organization (1992) *The Tenth Revison of the International Classification of Diseases and Related Health Problems* (ICD–10). Geneva: WHO.

https://doi.org/10.1192/bjp.175.2.154 Published online by Cambridge University Press