Definitions of Psychiatric Syndromes—Comparison in Hospital Patients and General Population

S. P. SASHIDHARAN

Four sets of commonly used 'operational definitions' of psychiatric disorders were applied to clinical information, obtained on the basis of semi-structured interviews with samples of hospital in-patients, out-patients, and general population. The agreement among them in defining 'cases' and in assigning specific diagnostic categories was examined. There was considerable variation among the diagnostic systems in distinguishing between 'cases' and 'non-cases' and in identifying sub-groups of cases as having specific diagnoses. Comparison of operationally defined syndrome categories with diagnoses assigned by clinicians also showed much disagreement.

In the last 10 to 15 years, a variety of operational definitions of psychiatric syndromes has become available for use in both clinical and epidemiological research. Such diagnostic criteria, usually allied to semi-structured interview schedules, are being used increasingly in defining cases and in categorising clinical syndromes. The original motives for introducing such operational definitions—to improve the reliability of psychiatric diagnoses and to make case identification and case definition methods more explicit, and hence comparable between various centres—have largely been achieved (Kendell, 1982).

Today, more than a dozen diagnostic interview schedules exist, each of which can generate operationally defined diagnostic categories on the basis of symptoms or clinical history or a combination of the two, but the existence of several alternative diagnostic systems means that a choice may have to be made among them. It has been suggested that two or three alternative definitions of whatever syndrome we are studying might be employed in all research, thus allowing examination of the relationship between the independent variable and each of the definitions in turn (Kendell, 1982). Whether or not we decide to choose one system or multiple criteria, it is necessary to understand how the various diagnostic systems deal with the same basis of clinical information. The overlap and contrast between them in identifying 'cases' and in assigning diagnostic labels in such contexts will help in understanding how the competing 'operational' rules relate to each other. This was one of the aims of a general population survey and a treatment inception study that were recently completed in Edinburgh. Our aim was to establish the agreement between four diagnostic systems commonly used in case identification and case definition in both the general population and among those referred to psychiatrists. An earlier paper was concerned with the general population survey results (Dean *et al*, 1983), and here we report a similar study on hospital treated individuals.

Method

The general population study was of a random sample of women aged 18–65, drawn from the electoral roll in a geographically defined area of North Edinburgh. Details are given in an earlier paper (Surtees *et al*, 1983); 576 women were interviewed and information on their psychiatric state ascertained. At the same time, all women between the ages 18 and 65 residing in the study area referred to psychiatric services over a period of six months were included in a treatment inception study. A stratified random sample of women referred to the out-patients clinic, who had not been in contact with psychiatric services in the previous six months, and all women admitted to hospital from the study area were interviewed by a research psychiatrist.

The psychiatric status of the community sample was assessed using the basic 40-item Present State Examination (PSE) (Wing *et al*, 1974), supplemented by additional questions from the Schedule for Affective Disorder and Schizophrenia (SADS) (Endicott & Spitzer, 1978) so that the symptom information obtained could be used to generate Catego Sub-classes (Wing & Sturt, 1978) as well as Research (RDC) diagnoses (Spitzer *et al*, 1978). In the hospital study, a full PSE and a full SADS (Part 1 and 2) were used, the schedules being completed at the same interview. All the patients in the hospital study were seen within a week of their admission or the day of attendance at the out-patient clinic.

The Bedford College checklist was applied (Finlay-Jones *et al*, 1980) to the information obtained by the PSE. Cases according to Feighner criteria (Feighner *et al*, 1972)

were identified, using information obtained on the SADS. Clinical diagnoses of the patients were obtained from their case records, and these were made independently by the clinical pyschiatrists who were responsible for the patient care.

Results

Detailed results of the general population study are reported elsewhere (Dean *et al*, 1983); in this paper, community cases are considered mainly for comparison with the hospital sample. The hospital sample consisted of 87 women between the ages of 18 and 65: 73 were chosen from a total of 307 new referrals to the psychiatric services in Edinburgh from the study area in a period of six months. Forty two of these were in-patients (out of 84 individual admissions) and 31 were out-patients at the time of their interview. In addition, 14 in-patients from a total of 44 patients who were considered as 'chronic' (i.e. had psychiatric contact in the previous six months) were added to this sample. Patients seen at special clinics and emergency out-of-hours referrals not admitted to hospital were excluded. Both the out-patient and in-patient samples were representative of all new referrals in terms of clinical diagnoses given at the time of their first contact. The Table shows the proportions of hospital attenders (both in-patients and out-patients) at the three levels (case, border-line case, non-case) of symptom severity, according to the different case definition systems used.

The Present State Examination provides an index of definition which assigns the PSE scores to increasing levels of certainty that a clinical diagnosis can be made (Wing 1976). Levels 1 to 4 are considered as 'non-cases', level 5 as 'threshold cases', and levels 6 to 8 are designated cases. The hospital sample can be assigned to these three levels, and the distribution of individuals will depend on whether the shortened (first 40 items) or the full version of the PSE is used. If symptom information obtained on the basis of a full PSE interview is used, then nearly 75% of hospital patients are seen to lie at or above the threshold level. The 40-item PSE, on the other hand, fails to pick up many of these individuals as cases or threshold cases, because they did not have sufficient of specific or severe affective symptoms, and this leads to a considerable increase (from 25% to nearly 45%) in the non-case category. In contrast, only 8.7% of the sample from the general population study achieved index of definition of 5 or above. The

percentage of individuals considered as non-cases in the hospital sample is much higher than in a previous report (Wing *et al*, 1981), where only 12% of out-patients and 5% of in-patients failed to reach the case-threshold level.

When the samples are compared, using the Research Diagnostic Criteria to distinguish between cases, probable cases and non-cases, the vast majority of individuals referred to hospital (98%) are designated as suffering from a current episode of probable or definite disorder. A larger proportion of individuals were identified as cases in the general population (79 out 576, i.e., 13.7%) according to the RDC, in comparison to 8.7% at ID level 5 or above.

While only 3.8% of the general population are designated as probable/definite cases according to Feighner's criteria, nearly a quarter (24%) of hospital referrals are given no diagnosis. The Bedford College Check list was applied to detect the proportions of individuals with affective disorder. Thirty seven percent of the hospital sample did not meet criteria for case or borderline case of depression and/or anxiety, and this can be contrasted with 80% of the general population who were similarly categorised as non-cases. The availability of only two diagnostic categories (depression and anxiety) on Bedford Check List may have contributed to the rather high proportion of individuals receiving no diagnostic labels in the hospital group.

The differences between diagnostic systems can be concealed or exaggerated by using a global measure like case/non-case distinction alone. A more rigorous test would be to examine the extent to which the different schemes pick up the same individuals as cases or non-cases, and to assign similar diagnoses to those identified as having sufficient symptoms.

Concordance between the systems in the general population was very poor, with only 54% agreement (33 out of 61) for cases of depression or affective personality and 17% (5 out of 30) for cases of anxiety or panic. Although there were similar numbers of panic/anxiety disorder cases according to the RDC and Catego A class (19 and 16 individuals respectively), in fact only five of these cases were picked up and designated as anxiety or panic by both RDC and Catego. If only the diagnostic labels ascribed by particular systems are reported, the one-month prevalence of anxiety states would be apparently similar under the two systems. Such spurious agreement would conceal the fact that different individuals are being given similar labels by the two systems.

TABLE
'Caseness' by alternative criteria hospital sample n = 87

Level of caseness	I.D % (Full PSE)	I.D % (40 item PSE)	RDC %	Feighner %	Bedford Check list %
Non-case	25.2	44.8	2.4	24.1	37.0
Threshold—borderline—probable	29.9	25.3	12.9	11.4	16.0
Definite case	44.8	29.9	84.7	64.5	47.0

The agreement between the systems remains poor when the hospital-referred cases are compared on the basis of diagnostic classes ascribed by the RDC and the Catego-ID. Under the Catego system, there are three descriptive categories of depression: Catego D requiring depressive delusions or hallucinations, Catego R requiring retardation, agitation or guilt, and Catego N for depression without any of these more severe, special features. There is a total of 44 individuals with RDC major/minor depressive disorder and/or Catego Class R, N, D, but the systems agree on only 33 of these, i.e. 75% agreement. This agreement falls to 42% (17 out of 41) if only the major depressive disorder and Catego R and D are compared. Of the remainder, one individual with an RDC diagnosis of major depressive disorder is identified as Catego class P and 2 as Catego A, and 3 from Catego Class D are all designated as either schizo-affective schizophrenia (depressed type) or depression superimposed on schizophrenia according to the RDC. While 2 of the Catego N class (out of 19) are designated as panic or generalised anxiety disorder and 1 as currently not mentally ill by the RDC, 2 individuals with RDC major depressive disorder are not assigned a Catego class. In fact, individuals considered not mentally ill or having an ID below 5 have an agreement of less than 20% (four out of 22) of being described as non-cases by both systems. The concordance between the systems remains equally poor for neurotic disorders other than depression. There are 17 individuals who meet either RDC anxiety or panic disorder criteria or Catego sub-class A requirements, but only five of them receive both diagnostic labels at the same time, an agreement of just over 29%. Similarly, three individuals with an RDC primary diagnosis of alcoholism or drug use disorder did not have sufficient symptoms to be allocated a Catego class. The agreement for non-organic, non-depressive psychosis is much more impressive. For schizophrenia (including schizo-affective), unspecified functional psychosis and Catego classes S and P, concordance was over 50%, and it rose to 80% for manic disorder and Catego class M.

A major problem in testing the validity of psychiatric diagnosis is the absence of external validating criteria against which syndromal descriptions can be compared. It is not possible to compare their validity by contrasting different diagnostic systems. The clinical relevance of the systems can, however, be looked at by comparing the diagnosis given according to each set of rules to the 'clinical diagnosis' arrived at by psychiatrists. It must be stressed, though, that such a clinical diagnosis, as given by psychiatrists, is usually based on total clinical information including history, while the diagnostic systems tested here rely almost exclusively on the present mental state for making a current diagnosis.

When the operational definitions are compared with clinical diagnoses ascribed to the hospital referred cases by the treating clinicians, out of the 25 individuals given a diagnosis of depressive psychosis, 17 are included under Catego sub-class R, N, or D. The majority of individuals with depressive neurosis (seven out of eight) are given Catego sub-classes R or N. However, only 24 individuals

out of the 39 who are given Catego R, N, D sub-classes are designated as having depressive psychosis or neurosis by the clinicians.

The index of agreement between Catego and clinical diagnosis shows much variation within individuals who are identified as suffering from similar conditions. Out of a total of 47 individuals given either a clinical diagnosis of depression and/or assigned Catego class N, D or R, only 24 (51%) are categorised under the similar label by both approaches. The agreement falls to 33% (11/33) for depressive psychosis and Catego D, R groups.

A similar comparison of RDC diagnosis with clinical diagnosis shows 17 out of the 25 individuals with depressive psychosis are operationally defined as major or minor depressive disorder. There are 43 individuals who are given a diagnosis of depression according to the RDC, but only 24 of them are given a similar diagnosis by the clinicians. The index of agreement for RDC major/minor depressions and all clinical diagnoses of depression is only 46% (24 out of 52), and this falls further when the depressive psychosis group is compared to the RDC major depression category (33%).

Discussion

The object of the various operational definitions is clearly not to diagnose, but to characterise and define for purposes of comparison between studies or samples. The major difference between the various systems is the way in which current mental state information is used to provide tentative diagnostic or syndromal categories. It is still discouraging, however, that the commonly used operational definitions shows such disagreements in case identification and in the categorisation of clinical symptoms. Although the agreement on case/non-case distinction according to all the definitions is better in the hospital sample than in the community, there is still considerable disagreement in assigning similar labels to the same individuals identified as cases. Two previous studies had indicated the extent of disagreement among different research diagnostic criteria, when applied to clinical information obtained through a single interview. In a study from the US (Zisook *et al*, 1980), four sets of criteria for depression were applied to 80 adult out-patients who participated in a clinical trial of anti-depressant medication. The agreement among four diagnostic systems varied from 58% to 74%, with Feighner/RDC agreement at 68%; the RDC diagnosis of depression (definite and probable) was given only to 57 of the 80 (71%) 'clinically depressed' patients.

Based on the data from the US/UK Diagnostic Project in 1966–1968, Brockington *et al* (1982) reported the concordance between Catego classes, RDC and DSM-III criteria for depression in a group of 125 patients admitted to a large London hospital.

Thirty-eight patients who received Catego class R + or N + , when compared with RDC major depressive disorder, had a concordance expressed as a kappa coefficient of 0.51, whilst 55 patients who received the RDC diagnosis showed a higher concordance (k = 0.86) with project psychiatrists' diagnosis of depressive psychosis.

In the present study, comparing the diagnostic criteria in the general population and the hospital groups, it is reasonable to expect greater agreement among the latter group, because they may well have more severe symptoms. This has proved to be the case to some extent, as shown by the agreement on depression in 54% in the community sample and in 75% in the hospital sample. For the anxiety/panic category, however, both samples show a concordance of around only 20%. Similarly, in the hospital sample, there is very good agreement between RDC and Catego in labelling schizophrenia and mania (40–80%), but for less severe and more common conditions like depression and anxiety, the agreement between the two systems is much poorer.

The extent of the disagreement between the RDC and the Catego/ID systems may be partly due to the differences in the time period covered by the SADS and PSE interviews. While the PSE is normally used to cover a period of one month prior to the interview, the SADS questions are directed to elict symptoms when they were at their worst in the current episode. This explanation may account for some of the discrepancy between the systems, although one would expect the hospital cases (most of them being 'new cases') to be at the worst period in their episode within a few weeks of hospital referral. Criteria for rating specific symptoms in the PSE and RDC are different. PSE symptoms are much more closely defined (e.g. for depressed mood to be rated as present, the symptom of lowered mood has to be of unvarying intensity, with no change in the symptom even if the patient tries to distract self from it), while the SADS symptoms lack such operational criteria. The presence of the symptom at any time during the episode, irrespective of how long it lasted or how severe it was in intensity, will fulfill the requirement according to the American system, but not the PSE.

Another source of disagreement between the two diagnostic systems could be the differences in the number of symptoms required to fulfill the criteria for a particular disorder or sub-class (as in Catego). Here, it is difficult to comment on the extent of the difference between the two systems, since the Catego classes are assigned on the basis of computer algorithms, which are not easily translatable in terms of clinical criteria as employed in the RDC. Similarly, the inclusion and exclusion criteria, e.g. definite major depressive disorder and Catego sub-classes R, N or D, cannot be matched against each other.

There could be other sources of disagreement, e.g. chronic dysthymic disorders as defined by the RDC, where the symptom is not of sufficient severity to obtain mood disturbance ratings on the PSE, and the absence of specific Catego categories to assign individuals with alcoholism or drug use disorder.

One of the problems in interpreting the findings of earlier studies in psychiatric nosology was the lack of rigour in applying well-defined criteria for diagnostic categories. We have now entered the age of structured interviews and operationally defined diagnostic systems of high reliability. However, it is still possible that by defining the dependent variable according to different diagnostic systems, we produce major disagreements between studies. Such discrepancies in diagnosis remain a major obstacle in interpreting differences in both clinical and epidemiological studies in psychiatry.

Acknowledgements

References

BROCKINGTON, I. F., HELZER, J. E., HILLIER, V. F., & FRANCIS, A. F. (1982) Definitions of depression: Concordance and prediction of outcome. *American Journal of Psychiatry*, **139**, 1022–1027.

DEAN, C., SURTEES, P. G., & SASHIDHARAN, S. P. (1983) Comparison of research diagnostic systems in an Edinburgh community sample. *British Journal of Psychiatry*, **142**, 247–256.

ENDICOTT, J. & SPITZER, R. L. (1978) A diagnostic interview: The Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry*, **35**, 837–844.

FEIGHNER, J. P., ROBINS, E., GUZE, S. B., WOODRUFF, R. A., WINOKUR, E., & MUNOZ, R. (1972) Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, **26**, 57–63.

FINLAY-JONES, R., BROWN, G. W., DUNCAN-JONES, P., HARRIS, T., MURPHY, E. & PRUDO, R. (1980) Depression and anxiety in the community: replicating the diagnosis of a case. *Psychological Medicine*, **10**, 445–454.

KENDELL, R. E. (1982) The choice of diagnostic criteria for biological research. *Archives of General Psychiatry*, **39**, 1334–1339.

SPITZER, R. L., ENDICOTT, J. & ROBINS, E. (1978) Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry*, 35, 773–782.

SURTEES, P. G., DEAN, C., INGHAM, J. G., KREITMAN, N. B., MILLER, P. MCC., & SASHIDHARAN, S. P. (1983) Psychiatric disorder in women from an Edinburgh community: association with demographic factors. *British Journal of Psychiatry*, 142, 238–246.

WING, J. K., COOPER, J. E. & SARTORIUS, N. (1974) The measurement and classification of psychiatric symptoms. London: Cambridge University Press.

—— (1976) A technique for studying psychiatric morbidity in in-patient and out-patient series and in general population samples. *Psychological Medicine*, 6, 665–671.

—— & STURT, E. (1978) The PSE-ID-CATEGO System Supplementary Manual. London: MRC Social Psychiatry Unit.

——, BEBBINGTON, P., HURRY, J. & TENNANT, C. (1981) The prevalence in the general population of disorders familiar to psychiatrists in hospital practice. In: *'What is a case?'* (eds.) J. K. Wing, P. Bebbington & L. N. Robins. London. Grant McIntyre.

ZISOOK, S., CLICK, M., JAFFE, K. & OVERALL, J. E. (1980) Research criteria for the diagnosis of depression. *Psychiatry Research*, 2, 13–23.

S. P. Sashidharan, MBBS, MPhil, MRCPsych, *MRC Unit for Epidemiological Studies in Psychiatry, University Department of Psychiatry, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF.*