Comparison of CATEGO-Derived ICD–8 and DSM–III Classifications using the Composite International Diagnostic Interview in Severely Ill Subjects

ANNE E. FARMER, PETER L. JENKINS, RANDY KATZ and LINDA RYDER

Two classifications, DSM–III and CATEGO4-derived ICD–8, included in the CIDI, are compared in 63 in-patients and out-patients with a broad range of psychiatric illness. Agreement for main diagnosis between these classifications was statistically significant for two time frames, the present state and lifetime. However, while diagnostic assignment by CATEGO4 remained fairly constant between time frames, there was a marked shift in DSM–III-assigned diagnosis, with cases changing from anxiety state diagnosis (present state) to affective disorder (lifetime). Thirty-nine subjects were assigned a DSM–III diagnosis of affective disorder for lifetime illness compared with 21 assigned to this diagnostic group by CATEGO4.

In this paper we compare two classifications of psychiatric disorder that can be derived from the Composite International Diagnostic Interview (CIDI; Robins *et al*, 1985) – DSM–III (American Psychiatric Association, 1980) and CATEGO(Wing *et al*, 1974; Wing & Sturt, 1978).

The CIDI has been developed from the Diagnostic Interview Schedule (DIS; Robins *et al*, 1981) for use in epidemiological research; the instrument's historical development and structure has been described in detail (Robins *et al*, 1988). Using a computer scoring program, diagnoses can be made according to DSM–III as well as other operational criteria (Feighner *et al*, 1972; Spitzer *et al*, 1975). In addition it is possible to hand-score CIDI. In developing CIDI from the DIS, the first 40 items and certain psychotic items from the Present State Examination (PSE; Wing *et al*, 1974; Robins *et al*, 1988) were included (although with different wording, scaling, and method of interview) so that certain CATEGO classes could be derived along with their tentative equivalents from ICD–8 (World Health Organization, 1974). In 1984 CIDI was approved by the World Health Organization (WHO) for international use in psychiatric research, and field trials of a "core version" of the instrument have recently been completed in 19 centres around the world.

Previously (Farmer *et al*, 1987 *a*), we evaluated the agreement between the PSE items in CIDI and the PSE given in separate interviews of the same subject. Although we found that item-by-item inter-rater agreement between the two interviews was variable and sometimes poor, when the PSE ratings from both interviews were scored by the CATEGO computer program, agreement for syndrome classification, diagnostic class and severity attained satisfactory statistical significance. In this study we assume that CATEGO scores from the CIDI approximate satisfactorily to those provided by the PSE and we therefore compare this classification with DSM–III.

The version of CATEGO that has been used to score the PSE items in CIDI is the fourth edition (CATEGO4). The CATEGO4 program groups together individual symptoms into 38 'syndromes'. These in turn are grouped into six subclasses. Finally, one CATEGO class is specified as the most likely diagnosis and a tentative ICD–8 category is suggested. This approach to classification is therefore somewhat different from the well known operational approach of DSM–III.

While the CATEGO4-derived ICD–8 classification may be considered more flexible, it provides less certainty of diagnosis and in particular gives weighting to certain symptoms or groups of symptoms. The program is probabilistic in its decision making. In contrast, operational criteria such as DSM–III produce greater certainty in assigning diagnosis but in general more cases fall outside the prescribed criteria into miscellaneous categories or categories of "not elsewhere specified".

In this paper comparisons are made across broad diagnostic groups only, for although highly comparable for three digits, ICD–8 and DSM–III deviate from each other at the four- and five-digit level. We report on the degree of agreement and disagreement between the two systems, and examine sources of disagreement to see if any trends emerge.

Method

Sixty-three subjects were chosen to represent a wide range of clinical diagnoses. Thirty subjects have been described

177

previously (Farmer *et al*, 1987 *a*), including ten patients with serious psychotic illness (six women, four men), selected from a consecutive series of patients admitted with a clinical diagnosis of psychosis, to the psychiatric ward of a district general hospital. Ten of these 30 patients (four men, six women) constituted a consecutive series of non-psychotic out-patients, seen by one psychiatrist at the Department of Psychological Medicine, King's College Hospital, London, and ten subjects were out-patients attending the Eating Disorders Unit at the Maudsley Hospital, London, nine of whom were women suffering from eating disorders and one of whom was a man who had a primary clinical diagnosis of neurotic depression. Lastly, there were 33 subjects (20 women, 13 men) who were in-patients and day patients at the Acute Psychiatric Unit at the University Hospital of Wales, Cardiff, and Whitchurch Psychiatric Hospital, Cardiff.

All patients were interviewed using the CIDI by psychiatrists (AEF and PLJ) or psychologists (RK and LR) trained in its use. Ratings for PSE-9 items from CIDI were then transcribed to PSE-9 coding sheets for past-months and lifetime frames by two graduate students in clinical psychology who had no knowledge of PSE-9 or CATEGO4 but who were given explicit transcription rules. The codes were then entered into the computer program CATEGO4 to obtain PSE syndrome frequencies, CATEGO class, and suggested ICD-8 diagnosis.

Broad ICD-8 diagnostic groups are compared with corresponding DSM-III categories, namely: 300 (anxiety states, phobic disorders, obsessive-compulsive disorders, neurotic depression), 295 (schizophrenic psychoses), 296 (affective psychoses), 297 (other psychoses) and, lastly, no diagnosis. As category 300.4 denotes separate disorders for DSM-III and ICD-8, the classifications are also compared for PSE syndrome profiles for 300.4 as well as 296 depression categories.

The DSM-III diagnosis was obtained blind to CATEGO4 data using a hand-scoring method (AEF) as the full computerised scoring program for CIDI was still being developed. As most subjects fulfilled several operational definitions for the same time frame, a clinical decision based on reference to hospital case records was made by AEF regarding the principal axis 1 DSM-III diagnosis. Both the present-state and lifetime diagnoses for DSM-III were produced for comparison with CATEGO4. For both time frames, the principal axis 1 diagnosis was employed in this comparison exercise.

Results

The interviewed sample consisted of 22 men and 41 women, mean age 42 (range 17–81 years). The full table of diagnostic assignments for DSM-III and CATEGO4 for each subject may be obtained upon request from the first author.

In order to apply tests of association between DSM-III and CATEGO4-derived ICD-8 classifications we must commence with the assumption that these are derived from two independent measures (Cicchetti, 1978). In practice this is not the case, since our results are derived from a single interview of each subject. In order to test whether we can

Table 1
Hierarchical log-linear analysis with case number (CN) and DSM-III and CATEGO4 diagnoses entered as factors

	d.f.	Partial χ^2	Probability
CN × DSM-III	248	82.061	1.000
CN × CATEGO4	248	81.344	1.000
DSM-III × CATEGO4	16	0.000	1.000
CN	62	28.933	0.999
DSM-III	4	29.183	0.000
CATEGO4	4	29.900	0.000

assume that diagnostic assignment is independent and not related to case identification, a hierarchical log-linear analysis was carried out (Frude, 1987) (Table 1). Case number (CN) and DSM-III and ICD-8 diagnoses were entered as factors, and only DSM-III and ICD-8 classification proved significant, there was no significant factor-by-factor interaction. Thus it is reasonable to conclude that association between DSM-III and ICD-8 classifications were independent of subject identity in the subsequent analyses.

Table 2 shows the cross-tabulations for DSM-III and ICD-8 present and lifetime diagnoses using the broad diagnostic groups described in the method. Chi-square and contingency coefficients for both time frames indicate associations which are comparatively strong and statistically significant (present state, $\chi^2 = 47.64$, d.f. = 9, $P = 0.000$; lifetime, $\chi^2 = 60.35$, d.f. = 9, $P = 0.000$). In addition, diagnostic assignment for DSM-III and CATEGO4 for each subject for each time frame was entered into a computer program (Frude, 1987) to assess diagnostic agreement via the kappa statistic, specifying the data as nominal (Cohen, 1960). For present state, percentage observed agreement was 57%, percentage expected agreement 26%, kappa = 0.418, P of kappa = 0.00001; and for lifetime, percentage observed agreement = 59%, percentage expected agreement = 31%, kappa = 0.399, P of kappa = 0.00001. Thus, on different tests of statistical association, agreement for diagnostic assignment between the two classifications is significantly better than chance. However, conventionally the findings for the kappa statistic are regarded as showing poor agreement for diagnostic assignment.

From the cross-tabulations in Table 2, the following observations can be made. For present state, 14% and 16% of cases (DSM-III and ICD-8 respectively) are not assigned to any diagnostic group. DSM-III does not assign any cases to the 297 (other psychoses) group although four cases are assigned to this category by CATEGO4. Both classifications assign approximately 30% of cases to the 300 (anxiety states) group. The main difference between the two classifications for present state diagnoses is shown by the percentage of cases assigned to the 295 (schizophrenic psychoses) and the 296 (affective psychoses) categories. DSM-III assigns the majority of cases (43%) to the affective psychoses category but only 13% to the schizophrenia category. In contrast, CATEGO4 assigns 25% of cases to 296 and 27% of cases to 295.

Table 2
Cross-tabulations of DSM-III and ICD-8 diagnoses from CIDI and kappa values

DSM-III	ICD-8 present state						ICD-8 lifetime					
	no diagnoses	300	296¹	295/297²	totals	kappa of agreement	no diagnoses	300	296	295/297³	totals	kappa of agreement
No diagnosis	5	4	0	0	9 (14%)	0.44	2	2	0	0	4 (6%)	0.65
300	4	9	3	3	19 (30%)	0.22	0	7	2	2	11 (17%)	0.29
296¹	1	7	13	6	27 (43%)	0.21	0	11	19	9	39 (62%)	0.35
295	0	0	0	8	8 (13%)	0.56	0	0	0	9	9 (14%)	0.53
Totals	10 (16%)	20 (32%)	16 (25%)	17 (27%)			2 (3%)	20 (32%)	21 (33%)	20 (32%)		

1. This includes 300.4/296.2.
2. Four cases assigned to 297 by CATEGO4 are included in the 295 group for the purpose of the analysis. No cases assigned to 297 by DSM-III.
3. Five cases assigned by CATEGO4 are included with 295 cases for the purposes of this analysis. No cases assigned to 297 by DSM-III.

For lifetime disorder (Table 2), fewer cases are assigned to the no-diagnosis group, which is to be expected since all subjects were receiving some type of psychiatric care at the time of interview. Thus subjects with only minor symptoms for the present state may have been under psychiatric review as they were still recovering from a severe episode of illness. However, 6% of cases for DSM-III and 3% of cases for CATEGO4 received no lifetime diagnosis. Once again, no cases were assigned to the 297 category by DSM-III while five cases were assigned to this group by the CATEGO4 program.

Fewer cases were assigned to the 300 (anxiety state) group for lifetime compared with present state by DSM-III. However, the same percentage (32%) were assigned to this group of diagnoses by CATEGO4. Once again, the main difference between the two classifications is found in the assignment to the 295 and 296 groups. Of DSM-III cases 62% were assigned to 296, while only 33% were by the CATEGO4 classification. Conversely, 14% of DSM-III cases were assigned to 295 compared with 32% categorised as schizophrenic psychoses according to CATEGO4.

It can also be seen that the broad diagnostic agreement for current illness for DSM-III and CATEGO4 is 56% and for lifetime illness is 59%. Conversely, there is 44% (28 cases) disagreement over present state, and 41% (26 cases) disagreement over lifetime illness. Three main sources of disagreement are examined: firstly, cases that are on the borderline of illness, i.e. where one classification assigns a category (usually 300 or 296) while the other classification fails to assign a diagnosis. Nine cases (32% of disagreements) are on the borderline between illness and not ill for present state and two cases (7%) for lifetime.
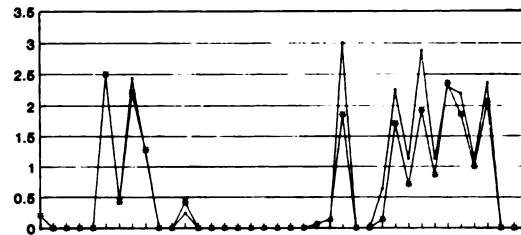

Fig. 1   The mean frequencies of 38 PSE syndromes for 296.2 and 296.3 categories (n = 63) (ICD-8, ▬▪▬ (296.2 only), DSM-III, ▬+▬(296.2 plus 296.3)).
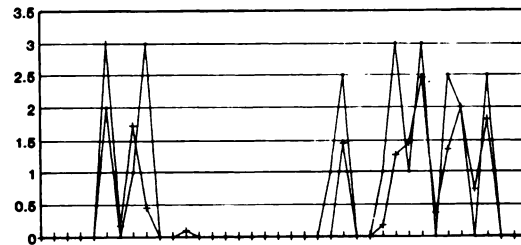

Fig. 2   The mean frequencies of 38 PSE syndromes for the 300.4 category (n = 63) (ICD-8, +, DSM-III, ▬▪▬).

The second source of disagreement is between a psychosis (295, 296 or 297 category) and a non-psychotic disorder (300 category). Here there is disagreement in 13 cases (46%) for present state and 15 cases (58%) for lifetime. Lastly, there is disagreement concerning type of psychotic illness, where one classification assigns a 296 (affective disorder) category while the other assigns to 295 or 297 (schizophrenic or other psychoses). Disagreement for present state occurred in six cases (21%) and for lifetime in nine cases (35%). As we have already noted, there is a marked tendency for DSM–III to assign to affective disorder subjects classed as schizophrenic by the CATEGO4 program.

The mean frequencies for 38 syndromes for 296.2, 296.3 and 300.4 categories are shown in Figs 1 (296) and 2 (300.4). CATEGO4 assignment for 11 cases was neurotic depression (N+; ICD–8 300.4) and for 17 cases was retarded depression (R+; ICD–8 296.2 or 300.4). Principal Axis 1 DSM–III diagnosis for two subjects was 300.4 and for 14 subjects was 296.22 or 296.32.

Discussion

While there have been previous comparisons of different classifications, none to our knowledge have used a single interview to derive two rather different nosologies such as DSM–III and CATEGO-derived ICD–8. Clearly there are a number of caveats which should be addressed before embarking on detailed consideration of our results.

Firstly, statistical tests of association assume independent measures which are clearly not the case in our study, since the same source has been used to derive both ICD–8 and DSM–III ratings. Log-linear analysis confirmed that the only significant interaction was between DSM–III and ICD–8 classifications. Thus we felt able to proceed with standard statistical tests of association the χ^2 and kappa statistics. However, the kappa agreement between the two classifications is modest, despite being significantly better than chance according to standard significance measures. Indeed, conventionally, kappa values less than 0.4 are considered poor and Table 2 shows that except for 295/297 and no diagnosis, agreement is clearly below this acceptable limit.

Secondly, the CATEGO4 ratings are not derived from the PSE but from a differently worded subset of CIDI items. While the PSE is a semi-structured interview devised for use by trained clinicians, CIDI is highly structured and specifically devised for use by lay interviewers. Both interviews however require a period of training in their use. While our previous study encouraged us to consider that CIDI-generated PSE ratings approximated well to PSE-generated ratings, they are clearly not the same and our results must be interpreted with this proviso.

In turn, the interpretation of ICD–8 diagnoses via the CATEGO4 program is tentative as ICD–8 contains no explicit rules to guide the construction of diagnostic algorithms. However, the CATEGO interpretation of ICD–8 has received widespread research use internationally and therefore general acceptance. Rather than just compare the end-point of the CATEGO process, namely the tentative ICD–8 assignment, we examined the PSE syndrome frequency for some of the categories of affective disorder. As mentioned in the introduction, 300.4 denotes rather different disorders in ICD–8 and DSM–III. In ICD–8 300.4 is a broad category covering an admixture of depressive and neurotic symptoms often developing acutely in response to a psychosocial stressor, while in DSM–III 300.4 is a dysthymic disorder without anxiety of at least two years' duration. Reference to the cross-tabulation in Table 2 shows that ICD–8 300.4 overlaps with DSM–III 300.0, 296.2, 296.3 and no diagnosis. Fig. 2 shows that the syndrome profile for 300.4 derived from DSM–III and ICD–8 is surprisingly congruent and also similar to Fig. 1 for 296 categories. Reference to the profile data from the International Pilot Study of Schizophrenia (Wing *et al*, 1974) shows that all four profiles resemble the R+ profile rather than that for anxiety (A+). However, Fig. 2 (for 300.4) must be interpreted with some caution since there are only two cases which are DSM–III 300.4.

The results show that the number of cases classed as having schizophrenic psychosis (295) by DSM–III and CATEGO4 remains fairly constant over time. This is probably related to patient selection, as all psychotic subjects included in the study had well established illness. The excess of cases assigned to 295 by CATEGO4 compared with DSM–III probably reflects differences in the coding algorithms for each classification. DSM–III may be considered to have somewhat stricter criteria for the diagnosis of schizophrenia. For instance CATEGO4 does not take symptom duration into account but is highly likely to assign a schizophrenia category solely on the presence of Schneiderian first-rank symptoms (Schneider, 1959), whereas DSM–III requires six months of illness or dysfunction before schizophrenia can be diagnosed. Thus only cross-sectional information is required for a CATEGO4 295 category whereas longitudinal data are necessary for DSM–III schizophrenia.

We have examined the sources of disagreement between the two classifications in order to see if any clear trends emerge. As discussed above, it might be predicted that differences between classifications for type of psychotic disorder might produce the greatest number of discordant cases. Indeed, as Table 2 shows, this source of disagreement produced 21% of

discordant cases for current illness rising to 35% for lifetime. Previous reliability studies (Robins *et al*, 1982) have shown that the area of greatest disagreement for inter-rater agreement lies on the border between mild illness and health. Much has been written and debated about the cut-off point that divides 'cases' from 'non-cases' (Wing *et al*, 1978). Thus it might be expected that the greatest area of discord between classifications might lie at the border between defined illness and relative health. While this source accounted for 32% disagreement for current illness, this fell to 8% for lifetime illness. Put more positively, the two classifications agree 97% of the time about what is a definite 'case'.

The greatest number of cases differed according to whether an anxiety state or a 29 diagnostic group was assigned (46% for present state and 58% for lifetime). The kappa analysis showed that bias was directed towards CATEGO4 assigned a 300 class, although this was not statistically significant (for present state bias = 27.0, $\chi^2 = 0.148$, $P =$ NS; for lifetime, bias = 26.0, $\chi^2 = 0.038$, $P =$ NS). Thus CATEGO4 tends to assign what might be considered a 'more severe or 29' diagnostic group compared with DSM-III, which tends towards the less severe or 300 grouping. This corresponds with the former distinction between 'neurotic' and 'psychotic' diagnostic categories which still remain in ICD-8 but which are omitted from DSM-III. The most likely explanation for this finding, as we have already pointed out, is that the coding algorithms for major psychiatric disorders in DSM-III are more stringent than for CATEGO4-derived ICD-8.

Approximately a third of cases are diagnosed as having an anxiety state by both classifications for present illness, and this proportion is maintained over lifetime illness by CATEGO4. However, in DSM-III the number of cases assigned to this group is reduced when lifetime diagnosis is examined, with cases apparently shifting to an affective disorder category. Indeed, for both time frames DSM-III assigns most cases to 296, but particularly for lifetime. In CATEGO4, however, there is a more even distribution of cases across 295, 296 and 300 diagnostic groups. Both these findings suggest that, compared with the CATEGO4 classification, DSM-III produces a diagnostic shift away from the schizophrenic psychoses (295) and anxiety states (300) towards affective disorder and that this is most marked for a lifetime diagnostic time frame.

These findings are at variance with an earlier study (Brink *et al*, 1989) comparing PSE/CATEGO and DSM-III in an out-patient group. These authors also found comparatively low agreement between classifications but showed a bias for PSE/CATEGO

diagnosing depression compared with DSM-III. However, a revised version of the PSE was used to make both CATEGO and DSM-III diagnoses, the inter-rater reliability for which is not reported in their study. In addition, these authors also suggest that in their out-patient group, PSE/CATEGO is relatively insensitive in diagnosing social phobias and obsessive–compulsive disorder. Our findings, the reverse of this, suggest that in a more severely ill sample, the DSM-III criteria for affective disorder are broader (Cooper *et al*, 1972; World Health Organization, 1973) than elsewhere. Such differences may have major implications in relation to patient management and statistics, and also for research in mental illness.

Both classifications appear to be somewhat sensitive to the type of subject group under scrutiny. In an out-patient setting, PSE/CATEGO categorises more individuals as suffering from affective disorder compared with DSM-III, while the reverse appears to be true for more severely ill in-patients. In addition, Spitzer & Williams (1988) have pointed out that the interpretation of seemingly quite explicit diagnostic criteria by individual clinicians can be quite different. While some consider schizophrenia as the top of a hierarchy of functional psychoses, other clinicians will place greater emphasis upon affective symptoms in making a diagnosis. Thus far, attempts to examine the relationship between different classifications have been somewhat limited and what research has been carried out indicates that different sets of criteria select cases according to different validating parameters (McGuffin *et al*, 1984; Farmer *et al*, 1987 *b*, *c*).

Although we can draw only limited conclusions from our study, we have attempted to highlight some of the nosological issues which can arise when classifications are compared. CIDI and the CATEGO4 program facilitate the reliable application of different classifications, which is an important gain. However, the absence of validating criteria means that all classifications remain working hypotheses and no single diagnostic schema can be regarded as correct.

References

AMERICAN PSYCHIATRIC ASSOCIATION (1980) *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn) (DSM-III). Washington, DC: APA.

BRINK, W., KOETER, W. J., ORMEL, J., *et al* (1989) Psychiatric diagnosis in an outpatient population. *Archives of General Psychiatry*, **46**, 369–372.

CICCHETTI, D. V. (1978) Assessing observer and method variability in medicine. *Connecticut Medicine*, **42**, 253–256.

COHEN, J. A. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.

COOPER, J. E., KENDELL, R. E., GURLAND, B. J., *et al* (1972) *Psychiatric Diagnosis in New York and London*. Maudsley Monograph. London: Oxford University Press.

FARMER, A. E., KATZ, R., MCGUFFIN, P., *et al* (1987a) A comparison between the Present State Examination and the Composite International Diagnostic Interview. *Archives of General Psychiatry*, **44**, 1064–1068.

——, MCGUFFIN, P. & GOTTESMAN, I. I. (1987b) Twin concordance for DSM III schizophrenia scrutinizing the validity of the definition. *Archives of General Psychiatry*, **44**, 634–644.

——, JACKSON, R., MCGUFFIN, P., *et al* (1987c) Cerebral ventricular enlargement in schizophrenia: consistencies and contradictions. *British Journal of Psychiatry*, **150**, 324–330.

FEIGHNER, J. P., ROBINS, E., GUZE, S. B., *et al* (1972) Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, **26**, 57–67.

FRUDE, N. (1987) *A Guide to SPSS/PC+*. London: Macmillan Education.

MCGUFFIN, P., FARMER, A. E., GOTTESMAN, I. I., *et al* (1984) Twin concordances for operational definitions of schizophrenia. *Archives of General Psychiatry*, **41**, 541–545.

ROBINS, L. N., HELZER, J. E., CROUGHAN, J., *et al* (1981) *NIMH Diagnostic Interview Schedule: Version III*. Bethesda: Department of Health and Human Services, Public Health Service, Alcohol Drug Abuse and Mental Health Administration, National Institute of Mental Health.

——, ——, RATCLIFF, K. S., *et al* (1982) Validity of the Diagnostic Interview Schedule, Version II: DSM–III diagnoses. *Psychological Medicine*, **12**, 855–870.

——, WING, J. K., WITTCHEN, H. V., *et al* (1988) The Composite International Diagnostic Interview. *Archives of General Psychiatry*, **45**, 1069–1078.

SCHNEIDER, K. (1959) *Clinical Psychopathology* (trans. M. Hamilton). New York: Grune and Stratton.

SPITZER, R. L., ENDICOTT, J. & ROBINS, E. (1975) *Research Diagnostic Criteria. Instrument No. 58*. New York: New York State Psychiatric Institute.

—— & WILLIAMS, J. B. W. (1988) Having a dream: a research strategy for DSM–IV. *Archives of General Psychiatry*, **45**, 871–874.

WING, J. K., COOPER, J. E. & SARTORIUS, N. (1974) *The Measurement and Classification of Psychiatric Symptoms*. London: Cambridge University Press.

——, MANN, S. A., LEFF, J. P., *et al* (1978) The concept of a 'case' in psychiatric population surveys. *Psychological Medicine*, **8**, 203–219.

—— & STURT, E. (1978) *The PSE–ID–CATEGO System. Supplementary Manual*. Mimeograph, MRC Social Psychiatry Unit, Institute of Psychiatry, London.

WORLD HEALTH ORGANIZATION (1973) *The International Pilot Study of Schizophrenia. Vol. 1*, Geneva: WHO.

—— (1974) *Mental Disorders, Glossary and Guide to their Classification in Accordance with the Eighth Revision of the International Classification of Diseases* (ICD–8). Geneva: WHO.

Anne E. Farmer, MD, MRCPsych, DPM; Peter L. Jenkins, MB, ChB, MRCPsych, Department of Psychological Medicine, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN; Randy Katz, PhD, Department of Psychology, Toronto General Hospital, 200 Elizabeth Street, Toronto, Ontario, Canada M5G 2C4; Linda Ryder, BSc, Department of Psychological Medicine, University of Wales College of Medicine

*Correspondence