

## EDITORIAL

### The end of the beginning: a requiem for the categorization of mental disorder?<sup>1,2</sup>

The well regarded Section of the World Psychiatric Association, on Epidemiology and Public Health, held its biennial meeting at Johns Hopkins University, Baltimore in June 2001. A new feature of the meeting was the ‘work group’, an opportunity to debate topical but unresolved issues. There was standing room only for a discussion on why the differences between self-report and clinician-rated measures of psychopathology matter (Brugha *et al.* 1999*a*). This is one of several topics increasingly debated in the context of the seemingly unremitting disparities between epidemiological estimates of the prevalence of psychiatric disorders (Regier *et al.* 1998). Recent comparisons of clinician and lay (self-report) measures suggest that dichotomous diagnostic categories cannot be measured reliably in large scale community surveys against a clinical standard measure (Brugha *et al.* 1999*b*, 2001; Eaton *et al.* 2000). In these three comparisons of a Present State Examination based clinical interview with fully structured survey diagnostic interviews, agreement for a diagnostic category, depressive disorder, according to the kappa statistic was typically < 0.2. Agreement was substantially improved when depression was expressed (using the same data) as a continuum of severity with correlations at between 0.4 and 0.5, although still falling short of adequate as an indicator of depressive symptoms in the community. Thus, although it is likely that the problem is primarily due to disparate measurement constructs, it may be partly overcome by shifting the measurement target from such dichotomous scales to continua.

Given the effect on concordance of the scale of measurement used in community surveys, it was inevitable that the work group discussion would focus its attention on the defining nature of the phenomena to be measured. The argument discussed here against dichotomizing a population into two groups, the healthy and the non-healthy, is that it is unnatural (does not reflect reality), it is difficult to measure reliably in the general population, it is generally unnecessary and it is wasteful. The argument for retaining such dichotomies is that they may have some utility in clearly specified contexts. Thus, re-opened ground was explored in a similar meeting held two decades previously (Wing *et al.* 1981*a*).

#### ARE THERE NATURAL JOINTS?

The human perceptual system has a ‘natural’ propensity to dichotomize as propounded in assimilation-contrast theory and as exemplified in stereotyping. One of two original papers presented at the work group (Melzer *et al.* 2002) begins by arguing, however, that in nature, which includes physiology, physical medicine and psychiatry, there are hardly ever true dichotomies (Khaw, 2001). Geoffrey Rose (1993) credited Sir George Pickering with the argument that the whole of a population should be studied as an entity that exists as a continuity and not just by comparing individuals at its extremes with the rest. Natural dichotomies are rare, as in the example of severe mental retardation at the lower tail of the normal distribution of intelligence. In attempting to define the concept of ‘a disease’ Scadding, a physician, acknowledged the vague and imprecise nature of

<sup>1</sup> Professor Brugha is a co-author of the paper by Melzer *et al.* published on pp. 1195–1201 of this issue to which this editorial refers.

<sup>2</sup> Address for correspondence: Professor Traolach S. Brugha, Section of Social and Epidemiological Psychiatry, University of Leicester, Department of Psychiatry, Brandon Mental Health Unit, Leicester General Hospital, Gwendolen Road, Leicester LE5 4PW.

the required term 'biological disadvantage'. He argued that disease definitions should rely instead on normal standards for relevant populations, which should be established on a statistical basis (Scadding, 1967).

The hypothesized discontinuity between depression and health should be studied in the general population. Melzer re-analysed CIS-R data from the British National Survey of psychiatric morbidity (Jenkins *et al.* 1997). As in the example of blood pressures in populations, and using a single exponential model, this analysis suggests that counts of common mental symptoms in general populations fall within a single distribution, with no sign of distinct case and normal subpopulations. The same model held for symptoms of depression, and for both low risk and high risk subpopulations, with the population means alone providing a good prediction of the numbers of actual cases scoring above the conventional cut-off on the CIS-R. Melzer's analysis builds on empirical work begun by Rose & Huppert, who showed that the frequency distributions of General Health Questionnaire scores in different populations move up and down as a whole: like other physiological and behavioural attributes, these mental health outcomes in individuals (including 'who is a case' and therefore 'prevalence') are associated with characteristics of the whole population in which they arise, including the larger subpopulation conventionally considered to be 'non cases' (Anderson *et al.* 1993). Huppert & Whittington (1995) showed that both disorders and symptoms predicted survival 7 years later, but importantly that mortality tends to rise with higher GHQ threshold cut points. Thus, the entire distribution of scores predicts survival, pointing to the arbitrariness and more limited value of a conventionally chosen single cut-point.

A similar case has recently been made for depression by Kendler and colleagues, who found little empirical support for the DSM-IV requirements for 2 weeks' duration, five symptoms, or clinically significant distress or impairment (Kendler & Gardner, Jr., 1998). Most functions appeared to be continuous and they concluded that major depression, as articulated by DSM-IV, may be a diagnostic convention imposed on a continuum of depressive symptoms of varying severity and duration. Latent class analysis, a technique similar to factor analysis but with the assumption that the groups to be derived are categorical, has been used to study data on symptoms measured in community surveys. Distinguishable syndromes of depression both below and above conventional diagnostic criterion levels of severity emerged (Eaton *et al.* 1989; Kendler *et al.* 1996; Sullivan *et al.* 1998). These studies complement earlier work investigating hierarchical patterns in psychiatric symptoms, showing that the rarer symptoms found in severe disorders are strongly predictive of commoner symptoms found in the general population, each of which is highly correlated with overall severity (Surtees & Kendell, 1979; Sturt, 1981). Having shown that symptoms of anxiety, depression and somatic symptoms are all strongly correlated, Goldberg (1996) argued that the failure to demonstrate that these represent distinct disorders undermines a basic assumption in the term 'co-morbidity', and provides a more parsimonious explanation for their so called co-occurrence.

A possible criticism of all of these studies comes from the realisation that the process of measurement inevitably requires some element of categorization. Thus, 'an element of arbitrariness is always implicit in measurement' (Wing *et al.* 1981*b*). The instruments used in these studies selectively impose a restricted (disease) perspective on the measurement process; there is hardly ever an attempt to measure the full, i.e. normal, range of emotions, forms of experiencing and thinking (Kovess & Beaudet, 2001) that must surely be distributed across the population as a whole (compare again with blood pressure). Thus, in spite of setting up these studies in order to confirm the existence of discrete disease entities, the failure to do so argues persuasively against an exclusively split view of the universe, as embodied in the accepted systems of classification.

### **ARE DICHOTOMIES NECESSARY OR JUST WASTEFUL?**

A frequently cited argument against a shift from categories to dimensions is that it would be too complicated, particularly in clinical practice (Andreasen, 1982; Kendell, 1983). Dichotomies can make things look easier. For example, the use of odds ratios helps compare the magnitude of

different associations in comparisons of persons with and without disease (phenotype). But discarding variance wastefully diminishes statistical power to explain and predict. A potential concern with diagnostic interviews is the incorporation of single item 'cut-off' or 'skip' points applied if full diagnostic criteria are unlikely to be met (Kessler *et al.* 1998). Phenotypes can be studied as continua and, arguably, should be if the underlying genetic model is thought to be complex rather than simple (Gottesman & Carey, 1981). In most instances, exposure to environmental risk also varies according to a continuum. Even death itself is now viewed as a process (Wade, 2002) most precisely modelled as survival time, albeit a limited continuum.

In relation to prevention, Rose acknowledged that confining intervention to high risk groups could have some useful advantages: intervention is appropriate and acceptable to the individual; it avoids interference with those not especially at risk and it offers a more cost effective use of resources (Rose, 1993). But are dichotomous clinical data essential for all (binary) clinical decisions such as treatment initiation or referral? Risk curves can now be used in decision making in physical medicine more accurately employing the full range of data available. The *British National Formulary* began to include Coronary Risk Prediction Charts with its September 2000, 40th edition (*British National Formulary*, 2000). To calculate the percentage risk (on a continuum) of coronary heart disease (CHD) the clinician measures the patient's systolic blood pressure (another continuum) and looks up tables for the patient's gender. Because the *BNF*, a desk-top book, is limited to presenting information, on printed paper, i.e. in no more than two dimensions, the other predictors of risk must be simplified to a forced dichotomy (diagnosis of diabetes) and a scale, age in 10 year groupings. But such limitations could be handled instantly by an electronic digital assistant or office computer.

#### **SHOULD WE AVOID DICHOTOMIES EXCEPT WHEN USEFUL OR UNAVOIDABLE?**

Gottesman credited Stromgren (Stromgren, 1950) with the realization that where a threshold is drawn may often depend on the purpose of the researcher (Gottesman & Carey, 1981). Henderson has since reminded us of '... the quest for criteria to dichotomize a population crudely into cases and non-cases ... should be complemented by a dimensional strategy in which both symptoms and disability are measured as continua' (Henderson *et al.* 2001). Copeland warned against trying to find one definition of a case for many different purposes in the (vain) hope of a reliable diagnosis (Copeland, 1981). In spite of claims of methodological progress, marked variability in prevalence reporting has persisted between studies using similar self-report measurement methods (Regier *et al.* 1998; Henderson *et al.* 2001) with identical classification rules. But mental health policy advisers want to count people who need (can benefit from) health care (Regier *et al.* 1998), not people who are otherwise unhappy, distressed but not 'ill' (although subthreshold states may substantially elevate the risk of future 'diagnosable' and potentially preventable disorders (Eaton *et al.* 1995)). Evidence based medicine seeks to assist the clinician in applying valid evidence about treatment effectiveness to decisions that lead to a better prognosis (Sackett *et al.* 1997). The selection of clearly defined homogeneous groups of patients for inclusion in treatment trials is, perhaps, almost unavoidable, but the boundary chosen should reflect the clinical and scientific problem to be solved and not a threshold set in a different context. The requirement to (randomly) assign interventions to two or more discrete groups is possibly the most compelling argument for dichotomization. Unfortunately, extrapolating such knowledge to the public at large, in order to measure unmet need, has put enormous demands on the comparability of the selection and measurement methods used in surveys. Furthermore, this poses a considerable challenge on journal editors and authors to describe in sufficient detail those studied. A possible way forward for deciding about those who do require treatment has been suggested by David Melzer: treatment trials are needed, stratified by proximity to the conventionally chosen case definitions for mental disorders, to establish an empirical evidence base for cut-points (Melzer *et al.* 2002). Thus, the odds of benefiting from a treatment (or the relative clinical benefit risk ratio) need not be fixed (yes or no) but would vary

according to more than one threshold on a disease continuum measure, analogous to the *BNF* coronary risk prediction equations. By taking this richer and more useful evidence base into account, public mental health surveys would then have a more informative basis for determining unmet need for care in the population that makes us of severity and benefit risk ratios.

### **HOW DID WE GET INTO THIS PARADIGM IN THE FIRST PLACE?**

The earlier meeting of the same WPA section ('What is a Case'), and the literature on psychiatric classification 'pre DSM-III', makes interesting reading in the light of recent events (Wing *et al.* 1981*a*). The benefit of an agreed consensus on case definition was heralded without any discussion of the theory of dichotomies (Feighner *et al.* 1972). A cut-point on the Index of Definition corresponding to the assembling of sufficient criteria to make a categorization according to the ICD was developed and tested with similar 'reliable case definition' for the purposes of comparison (Wing *et al.* 1978). We need to take stock of the modest scope of these first useful steps, which soon took hold and replaced a former era of clinical diagnostic anarchy (Skodol & Spitzer, 1982). 'Unfortunately, although we have learnt how to make reliable diagnoses we still have no adequate criterion of their validity and the achievement focuses attention on the failure' (Kendell, 1989). Not for the first time we have failed to learn from our history and to move on.

### **WHAT ABOUT THE BEDSIDE?**

It would be wholly misleading to suppose that psychiatric clinicians slavishly follow a uni-axial categorical approach; if anything, multi-modal approaches that include dimensions and individual personal narratives (Nurcombe & Gallagher, 1986) seem to be preferred. Yet in training them we seem to expect our doctors to make 'correct' diagnoses (or fail in the attempt). A powerful but erroneous message implicit in accredited classification rules may have pervasively undermined critical enquiry. Paradoxically, a high proportion of medical school entrants may well be more mathematically and statistically able to adopt dimensional (statistical) approaches than are we their teachers. Our failure may have grown from an uncritical adoption of taxonomical principles more suited to studying discrete organisms (plants, animals and micro-organism) instead of concepts for defining and measuring variations between individuals. Imagine for a moment a committee endorsed classification of categories of human height and weight based on the rationale that categories work fine for eye and hair colour?

### **BUT WHAT ABOUT THE REST OF US?**

Arguably, presenting the results of clinical predictions on continua could facilitate a more honest sharing of difficult decisions with interested patients; unreasonable expectations (for which we doctors are greatly to blame) of 100% and 0% answers to worrying questions would no longer demand commensurately distorted replies (Elwyn *et al.* 2001). If the population should no longer be divided into the 'healthy' and the 'mentally disordered' then all of us are to some extent, varying over time and amongst ourselves, both mentally well and mentally unwell, in much the same way that all of us have varying levels of blood pressure or body weight. Stigmatization depends on separating off the stigmatized. The phrase 'nervous breakdown' is the understandable lay person's attempt to delimit 'us from them'. Splitting off the mentally ill doesn't help us to learn to lead less depressing, anxiety-laden lives. Positive attention to the mental health of all of us in the whole population, if Rose and Melzer are correct, could serve to shift the population mean to the left, benefiting the population as a whole (Weich, 1997). Sadly, none of these arguments are new or in any way original (Kendell, 1989). Meanwhile, public health and physical medicine have moved on

to implement dimensional clinical tools, neglected in psychiatry in the mistaken belief that physical medical disorders are precise discrete entities.

### WHAT ABOUT DSM-V AND ICD-11?

Quoting Goldberg: 'Despite these warnings recent advances in taxonomy have applied a categorical model across the whole range of mental disorders' (Goldberg, 1996). Is the formerly understandable argument that dimensional approaches would be equally complex and no more useful (Andreasen, 1982) tenable? Have we over extended the potentially useful concept of co-morbidity although dimensional population data point to a more parsimonious taxonomy than our current 400 plus discrete nosological categories imply? Can we defend the absence from one of our current classification systems of, perhaps, the commonest and most widespread form of all, 'mixed anxiety and depression', possibly because even fewer, if any taxonomists could discern a line of demarcation between it and how many of us feel this morning?

The APA and WHO have been lauded for the remarkable achievement of a close agreement on diagnostic thresholds that we know to be essentially arbitrary. When the diagnostic rules for major depressive episode (MDE) and for depressive episode (F32.0 in ICD-10) were applied to a PSE-9 patient symptom data set, the level of agreement (kappa) was 0.55 (Brugha & Bebbington, 1992). However, the correlation between the total number of criterion symptoms listed in each of the two classification rules exceeded 0.9. Surely, what we should value most is the remarkable achievement of an international consensus on the choice of an essentially common pool of items that comprises the spectrum of human psychopathology. The reported intention of these bodies to make greater use of research data and evidence in approaching possible future taxonomic revisions is to be commended also.

TRAOLACH S. BRUGHA

Comments on earlier drafts of this paper from Jeremy Coid, Bill Eaton, Scott Henderson, Felicia Huppert, Rachel Jenkins, Howard Meltzer and David Melzer are gratefully acknowledged. The author takes sole responsibility for the views expressed here.

### REFERENCES

- Anderson, J., Huppert, F. & Rose, G. (1993). Normality, deviance and minor psychiatric morbidity in the community. A population-based approach to General Health Questionnaire data in the Health and Lifestyle Survey. *Psychological Medicine* **23**, 475–485.
- Andreasen, N. C. (1982). Concepts diagnosis and classification. In *Handbook of Affective Disorders* (ed. E. S. Paykel), pp. 24–44. Churchill Livingstone: Edinburgh.
- British National Formulary* (2000). British Medical Association and the Royal Pharmaceutical Society of Great Britain: London.
- Brugha, T. S. & Bebbington, P. E. (1992). A comparison of DSM-III-R Major Depression and ICD-10 DCR Depressive Episode diagnostic criteria using PSE-9 clinical data. *International Journal of Methods in Psychiatric Research* **2**, 1–14.
- Brugha, T. S., Bebbington, P. E. & Jenkins, R. (1999a). A difference that matters: comparisons of structured and semi-structured diagnostic interviews of adults in the general population. *Psychological Medicine* **29**, 1013–1020.
- Brugha, T. S., Bebbington, P., Jenkins, R., Meltzer, H., Taub, N. A., Janas, M. & Vernon, J. (1999b). Cross validation of a household population survey diagnostic interview: a comparison of CIS-R with SCAN ICD-10 diagnostic categories. *Psychological Medicine* **29**, 1029–1042.
- Brugha, T. S., Jenkins, R., Taub, N. A., Meltzer, H. & Bebbington, P. (2001). A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychological Medicine* **31**, 1001–1013.
- Copeland, J. (1981). What is a case? A case for what? In *What is a Case: The Problem of Definition in Psychiatric Community Surveys* (ed. J. K. Wing, P. E. Bebbington and L. Robins), pp. 9–11. Grant McIntyre: London.
- Eaton, W. W., Dryman, A., Sorenson, A. & McCutcheon, A. (1989). DSM-III major depressive disorder in the community: a latent class analysis of data from the NIMH Epidemiologic Catchment Area Program. *British Journal of Psychiatry* **155**, 48–54.
- Eaton, W. W., Badawi, M. & Melton, B. (1995). Prodromes and precursors: epidemiologic data for primary prevention of disorders with slow onset. *American Journal of Psychiatry* **152**, 967–972.
- Eaton, W. W., Neufeld, K., Chen, L. S. & Cai, G. (2000). A comparison of self-report and clinical diagnostic interviews for depression: diagnostic interview schedule and schedules for clinical assessment in neuropsychiatry in the Baltimore epidemiologic catchment area follow-up. *Archives of General Psychiatry* **57**, 217–222.
- Elwyn, G., Edwards, A., Eccles, M. & Rovner, D. (2001). Decision analysis in patient care. *Lancet* **358**, 571–574.
- Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Winokur, G. & Muñoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry* **26**, 1–63.
- Goldberg, D. (1996). A dimensional model for common mental disorders. *British Journal of Psychiatry* **168**, 44–49.
- Gottesman, I. & Carey, G. (1981). Defining cases by genetic criteria. In *What is a Case: The Problem of Definition in Psychiatric Community Surveys* (ed. J. K. Wing, P. E. Bebbington and L. Robins), pp. 29–41. Grant McIntyre: London.
- Henderson, S., Korten, A. & Medway, J. (2001). Non-disabled cases in a national survey. *Psychological Medicine* **31**, 769–777.



- Huppert, F. A. & Whittington, J. E. (1995). Symptoms of psychological distress predict 7-year mortality. *Psychological Medicine* **25**, 1073–1086.
- Jenkins, R., Lewis, G., Bebbington, P., Brugha, T., Farrell, M., Gill, B. & Meltzer, H. (1997). The national psychiatric morbidity surveys of Great Britain – initial findings from the household survey. *Psychological Medicine* **27**, 775–789.
- Kendell, R. E. (1983). The principles of classification in relation to mental disease. In *Handbook of Psychiatry 1: General Psychopathology* (ed. M. Shepherd and O. L. Zangwill), pp. 191–198. Cambridge University Press: Cambridge.
- Kendell, R. E. (1989). Clinical validity. *Psychological Medicine* **19**, 1–55.
- Kendler, K. S. & Gardner, C. O., Jr. (1998). Boundaries of major depression: an evaluation of DSM-IV criteria. *American Journal of Psychiatry* **155**, 172–177.
- Kendler, K. S., Eaves, L. J., Walters, E. E., Neale, M. C., Heath, A. C. & Kessler, R. C. (1996). The identification and validation of distinct depressive syndromes in a population-based sample of female twins. *Archives of General Psychiatry* **53**, 391–399.
- Kessler, R. C., Wittchen, H. U., Abelson, J. M., McGonagle, K. A., Schwarz, N., Kendler, K. S., Knauper, B. & Zhao, S. (1998). Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey (NCS). *International Journal of Methods in Psychiatric Research* **7**, 33–55.
- Khaw, K. (2001). Health and disease are a continuum. *Drug and Therapeutics Bulletin* **39**, 47–47.
- Kovess, V. & Beaudet, M. P. (2001). Concepts and measurement of positive mental health. *Psychiatria Fennica* **32**, 14–34.
- Melzer, D., Tom, B. D. M., Brugha, T. S., Fryers, T. & Meltzer, H. (2002). Common mental disorder symptom counts in populations: are there distinct case groups above epidemiological cut-offs? *Psychological Medicine* **32**, 1195–1201.
- Nurcombe, B. & Gallagher, R. M. (1986). The clinical process in psychiatry: diagnosis and management planning. In *The Clinical Process in Psychiatry* (ed. B. Nurcombe and R. M. Gallagher), pp. 225–242. Cambridge University Press: Cambridge.
- Regier, D. A., Kaelber, C. T., Rae, D. S., Farmer, M. E., Knauper, B., Kessler, R. C. & Norquist, G. S. (1998). Limitations of diagnostic criteria and assessment instruments for mental disorders. *Archives of General Psychiatry* **55**, 109–115.
- Rose, G. (1993). Mental disorder and the strategies of prevention. *Psychological Medicine* **23**, 553–555.
- Sackett, D. L., Richardson, W. S., Rosenberg, W. & Haynes, R. B. (1997). *Evidence Based Medicine: How to Practice and Teach EBM*. Churchill Livingstone: Edinburgh.
- Scadding, J. G. (1967). Diagnosis: the clinician and the computer. *Lancet* **ii**, 877–882.
- Skodol, A. E. & Spitzer, R. L. (1982). The development of reliable diagnostic criteria in medicine. *Annual Review of Medicine* **33**, 317–326.
- Stromgren, E. (1950). Statistical and genetical population studies within psychiatry: methods and principle results. In *Congress International de Psychiatrie*. Heinmann: Paris.
- Sturt, E. (1981). Hierarchical patterns in the distribution of psychiatric symptoms. *Psychological Medicine* **11**, 783–794.
- Sullivan, P. F., Kessler, R. C. & Kendler, K. S. (1998). Latent class analysis of lifetime depressive symptoms in the national comorbidity survey. *American Journal of Psychiatry* **155**, 1398–1406.
- Surtees, P. G. & Kendell, R. E. (1979). The hierarchy model of psychiatric symptomatology: an investigation based in Present State Examination ratings. *British Journal of Psychiatry* **135**, 438–443.
- Wade, D. T. (2002). The dis-integration of death. *Lancet* **360**, 425–426.
- Weich, S. (1997). Prevention of the common mental disorders: a public health perspective. *Psychological Medicine* **27**, 757–764.
- Wing, J. K., Mann, S. A., Leff, J. P. & Nixon, J. M. (1978). The concept of a case in psychiatric population surveys. *Psychological Medicine* **8**, 203–218.
- Wing, J. K., Bebbington, P. E. & Robins, L. N. (1981a). *What is a case? The Problem of Definition in Community Surveys*. Grant McIntyre: London.
- Wing, J. K., Bebbington, P. E. & Robins, L. N. (1981b). Theory-testing in psychiatric epidemiology. In *What is a Case: The Problem of Definition in Psychiatric Community Surveys* (ed. J. K. Wing, P. E. Bebbington and L. Robins), pp. 3–8. Grant McIntyre: London.