Psychological Medicine, 2002, **32**, 585–594. © 2002 Cambridge University Press
DOI: 10.1017/S0033291702005457 Printed in the United Kingdom

Understanding cross-national differences in depression prevalence

G. E. SIMON,[1] D. P. GOLDBERG, M. VON KORFF AND T. B. ÜSTÜN

From the Center for Health Studies, Group Health Cooperative, Seattle, WA, USA: Institute of Psychiatry, London; and Assessment, Classification and Epidemiology and Division of Mental Health, World Health Organization, Geneva, Switzerland

ABSTRACT

Background. Previous epidemiological studies indicate large cross-national differences in prevalence of depression.

Methods. At 15 centres in 14 countries, 25916 primary care patients were screened for common mental disorders. A stratified random sample of 5447 primary care patients completed a baseline diagnostic assessment and 3197 completed a 12-month follow-up assessment. Psychiatric symptoms and diagnoses were assessed using the Composite International Diagnostic Interview (CIDI). Interviewer-rated disability was assessed using the Social Disability Schedule (SDS).

Results. Prevalence of current major depression varied 15-fold across centres. When centres were divided into three groups according to prevalence rates, the symptom pattern or latent structure of depressive illness was generally similar at low-, medium-, and high-prevalence centres. Depression was universally associated with disability, but this association varied significantly ($t = 3{\cdot}51$, $P = 0{\cdot}0005$) across centres. At higher-prevalence centres, depression was associated with lower levels of impairment. At 1 year follow-up, higher prevalence centres had both significantly higher rates of depression onset ($t = 3{\cdot}11$, $P = 0{\cdot}002$) and higher rates of persistence among those depressed at baseline ($t = 2{\cdot}49$, $P = 0{\cdot}013$).

Conclusions. Large cross-national variations in depression prevalence cannot be attributed to 'category fallacy' (cross-national differences in the nature or validity of depressive disorder). Use of identical measures and diagnostic criteria may actually identify different levels of depression severity in different countries or cultures. Cross-national differences in the onset and outcome of depression may reflect either true prevalence differences or differences in diagnostic threshold.

INTRODUCTION

Epidemiological studies consistently find large cross-national variation in the prevalence of depression. Summarizing findings from community surveys in 10 countries, Weissman *et al.* (1996) reported lifetime prevalence ranging from 1·5% in Taiwan and 2·9% in Korea up to 16% in Paris and 19% in Beirut. In a 15-site international primary care survey (Üstün & Sartorius, 1995), we found a 15-fold variation in prevalence of current depressive disorder – with highest rates in South American centres, lowest rates in Asia, and intermediate rates in the US and Western Europe. Other community and primary care surveys show a similar pattern (Araya *et al.* 1994, 2001; Lima *et al.* 1996; Almeida-Filho *et al.* 1997). Proposed explanations (Weissman *et al.* 1996) for this variability include ethnic differences in vulnerability to depression and community-level differences in exposure to stressors or traumatic events (e.g. political repression, rapid cultural shifts, socioeconomic deprivation and threat of violence).

Observed differences in depression rates may, however, represent problems with definition and measurement rather than true differences in prevalence. First, the depressive syndrome

[1] Address for correspondence: Dr Gregory E. Simon, Center for Health Studies, Group Health Cooperative, 1730 Minor Ave. No. 1600, Seattle, WA 98101-1448, USA.

https://doi.org/10.1017/S0033291702005457 Published online by Cambridge University Press

defined in the US and Western Europe may not exist – or exist in different forms – in other cultures. Cross-cultural application of criteria such as the American Psychiatric Association's Diagnostic and Statistical Manual or DSM-IV (American Psychiatric Association, 1994) may yield misleading or erroneous results (Littlewood, 1990; Mezzich *et al.* 1999), a problem Kleinman (1977) and others have called 'category fallacy'. Secondly, use of identical diagnostic methods may identify different levels of severity in different language or cultural groups. In this case, apparent cross-national differences in depression prevalence may reflect linguistic or cultural differences in the tendency to report distress (Dohrenwend, 1966; Phillips & Clancy, 1970; Krause & Carr, 1978). Simple translation of diagnostic measures developed in the US or Western Europe could significantly underestimate or overestimate prevalence in countries or cultures with a lesser or greater tendency to report distressing symptoms. Consistent with this hypothesis, we have reported that prevalence rates for a wide range of psychological and somatic symptoms show similar cross-national variation. In a cross-national primary care survey, sites with higher prevalence of depression also showed higher rates of anxiety and somatoform disorders (Üstün & Sartorius, 1995). Greater reporting of depression or anxiety symptoms was invariably associated with greater reporting of a wide range of somatic symptoms – whether or not those symptoms were considered medically unexplained (Simon *et al.* 1996, 1999).

The World Health Organization's Psychological Problems in General Health Care (PPGHC) study (Sartorius *et al.* 1993; Üstün & Sartorius, 1995) used identical methods and measures to examine the form, frequency, and course of common psychological disorders among primary care patients in 15 centres on five continents. We use these data to examine three questions: (1) Is the depressive syndrome defined by DSM-IV consistent across low-, medium- and high-prevalence centres?; (2) Does the use of identical diagnostic standards identify a different severity of illness across low-, medium- and high-prevalence centres?; and (3) Do low-, medium- and high-prevalence centres show differences in the course of depression consistent with true differences in prevalence?

METHOD

Participants

The PPGHC study examined the longitudinal epidemiology of common psychological disorders among primary care patients in Ankara, Turkey; Athens, Greece; Bangalore, India; Berlin, Germany; Groningen, Netherlands; Ibadan, Nigeria; Mainz, Germany; Manchester, UK; Nagasaki, Japan; Paris, France; Rio de Janeiro, Brazil; Santiago, Chile; Seattle, USA; Shanghai, China; and Verona, Italy. Study methods are described in detail elsewhere (Sartorius *et al.* 1993; Üstün & Sartorius, 1995) and will be summarized here. Each centre selected clinics typical of local primary health care and screened a consecutive or random sample of adult patients below the age of 65. Consenting patients ($N = 25916$, response rate 96%) completed the 12-item General Health Questionnaire or GHQ (Goldberg & Williams, 1988) prior to the physician visit. A stratified random sample (100% of those with GHQ scores above the 80th percentile, 35% of those scoring between the 60th and 80th percentiles, and 10% of those scoring below the 60th percentile for each site) was selected for diagnostic assessment. Sixty-two per cent of those selected ($N = 5447$) completed the diagnostic assessment, and non-participation was not significantly related to age, gender, or screening GHQ-12 score. Participation did vary significantly across centres, but sampling weights (see below) were used to adjust for this difference.

Assessment

The diagnostic assessment included a primary care version of the Composite International Diagnostic Interview or CIDI (Robins *et al.* 1988), the Social Disability Schedule or SDS (Wiersma *et al.* 1988), and self-report measures of disability and chronic medical morbidity. For each patient interviewed, the treating physician completed a five-level rating of severity of current medical and psychological illness. The CIDI is a structured psychiatric interview developed by the WHO for use in cross-national studies. For all respondents, the CIDI included 25 questions regarding depressive symptoms during the last month. Using standard CIDI scoring methods, these questions were combined into indicators of the nine depressive symptoms

https://doi.org/10.1017/S0033291702005457 Published online by Cambridge University Press

specified by DSM-IV. The SDS is a semi-structured disability assessment incorporating local norms and role expectations. Questions regarding observable events and behaviours (e.g. 'Were there complaints from your supervisor or colleagues about how much work you accomplished?') yield overall disability ratings on a four-point scale (None, Mild, Moderate and Severe). Each participating centre completed translation and back-translation of the CIDI and SDS followed by extensive field testing. Interviewer training and procedures for assuring reliability across centres are described elsewhere (Sartorius *et al.* 1993).

All patients with significant psychological symptoms and a 40% random sample of the remainder were asked to complete a diagnostic re-assessment 12 months later. Of those invited, 68·5% ($N = 3197$) did so, and non-participation was not significantly related to age, sex, or psychiatric diagnosis at baseline. Follow-up participation also varied significantly across centres, and sampling weights (see below) were again used to adjust for this difference.

Analysis

We examined our first question (Is the depressive syndrome defined by DSM-IV consistent across centres?) using item response analyses, a method examining the relationship between individual items and the construct measured by a complete scale (Bock & Zimowski, 1995). We considered the nine DSM-IV depression criteria as items in an overall test for depression. This method uses two parameters to describe the relationship between each symptom and overall depression severity. The first is the slope parameter, reflecting the correlation between the individual symptom and the overall severity level measured by all symptoms. The second parameter is the threshold value, the level of overall depression severity at which 50% of those interviewed report that particular symptom. A specific symptom (such as sleep disturbance) may show a relatively weak or strong correlation with depression severity. At the same time, that symptom may be common among patients with mild depression (i.e. low threshold value) or may not occur unless depression is severe (i.e. high threshold value). Significant variability across centres in either the pattern of correlations or the pattern of threshold values would undermine the uniform validity of the DSM-IV depressive syndrome and support the hypothesis that variations in depression prevalence represent 'category fallacy'.

We examined our second question (Does use of identical diagnostic standards identify a different severity of illness across centres?) by examining the relationship between depression and disability. In other words, SDS disability ratings were used as independent or proxy measures of depression severity. We specifically examined whether any given level of depression was associated with less disability at higher-prevalence centres, suggesting a lower effective diagnostic threshold at those centres.

We examined our third question (Do centres show difference in the course of depression consistent with true differences in prevalence?) by examining the 1-year course of depression – both the rate of onset among those without major depressive episode at baseline and the rate of persistence among those with major depression at the initial assessment.

While we report the actual (unweighted) number of observations in each analysis, all analyses incorporated sampling weights to correct for the stratified sampling procedure described above and for differences in participation rates across study centres (Cochran, 1977). Preliminary analyses were conducted using SPSS for Windows software (SPSS Inc., Chicago, IL). Final regression models were calculated using WesVar Complex Samples (SPSS Inc., Chicago, IL), a software package using jackknife replication to calculate variances and standard errors in complex survey samples. Item response analyses were conducted using the Bilog MG (Scientific Software International, Chicago, IL), a specialized package for evaluating differences in questionnaire item performance across groups. Item response analyses were based on two-parameter models with maximum likelihood estimation.

RESULTS

Variation in prevalence

As shown in Table 1, prevalence of current depressive disorder varied 15-fold across centres. In a logistic regression model including adjustment for age and sex, variation in depression prevalence far exceeded that expected by chance

https://doi.org/10.1017/S0033291702005457 Published online by Cambridge University Press

Table 1. *Prevalence of DSM-IV depressive disorder among consecutive primary care patients at* 15 *study centres. Adjusted odds ratio reflects odds of depression diagnosis relative to Bangalore after adjustment for age and sex* (N = 5447)

	Interviewed *N*	Cases *N*	Weighted prevalence %	95% CI	Adjusted OR
Nagasaki	336	16	1·6	0·7–2·5	0·18
Shanghai	576	39	2·5	1·6–3·4	0·27
Ibadan	269	25	4·0	2·6–5·5	0·39
Verona	250	32	4·6	2·9–6·3	0·47
Berlin	400	54	5·3	3·9–6·7	0·57
Seattle	373	60	6·4	4·6–8·2	0·64
Athens	196	38	7·3	4·6–9·9	0·82
Bangalore	398	70	8·6	6·1–11·1	1·0
Mainz	400	55	9·9	6·6–13·1	1·12
Ankara	400	94	10·7	7·7–13·7	1·18
Paris	405	128	13·5	10·9–16·0	1·66
Groningen	340	108	14·4	11·3–17·6	1·71
Manchester	428	159	17·1	14·4–19·8	2·02
Rio de Janeiro	393	151	18·3	14·2–22·3	2·19
Santiago	274	105	26·3	16·9–35·8	2·99

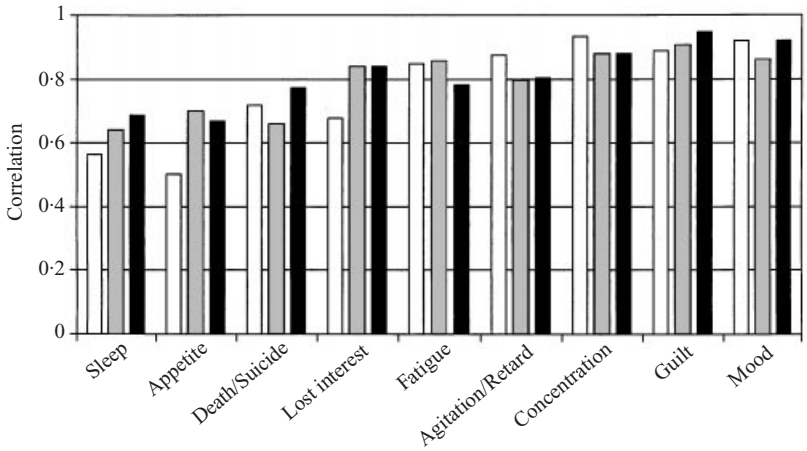


FIG. 1. Point-biserial correlation coefficients between individual depressive symptoms and overall depression symptom count with centres grouped according to depression prevalence. For each symptom: □, low-prevalence centres (Shanghai, Nagasaki); ▒, medium-prevalence centres; and ■, high-prevalence centres (Rio, Santiago). $N = 5447$ for all symptoms.

($t = 6{\cdot}84$, $P < 0{\cdot}001$). For all subsequent analyses, we consider centres in three categories: a low-prevalence group (Nagasaki, Shanghai; prevalence rate = 2·2%), a medium-prevalence group (Ibadan, Verona, Berlin, Seattle, Athens, Bangalore, Mainz, Ankara, Paris, Groningen, Manchester; prevalence rate = 9·7%), and a high-prevalence group (Rio de Janeiro, Santiago; prevalence rate = 21·6%).

Consistency of depressive syndrome

For the full item response model, chi-square goodness-of-fit statistic was 666·8 (df = 62, $P < 0{\cdot}001$) with goodness-of-fit statistics for individual items all > 40 ($P < 0{\cdot}001$ for all).

Fig. 1 displays correlations between individual symptoms and overall depression severity for low-, medium-, and high-prevalence centres. Across all three groups, correlations ranged from approximately 0·65 to approximately 0·95. The pattern (lowest correlation for sleep disturbance and changes in appetite, highest correlation for depressed mood and feelings of worthlessness) was quite consistent across the three groups.

Fig. 2 displays threshold values for individual symptoms in the three prevalence groups. Values ranged from approximately 1·3 to approximately 2·3 on a logarithmic scale indicating a 2·7-fold difference in symptom probability at the same

https://doi.org/10.1017/S0033291702005457 Published online by Cambridge University Press

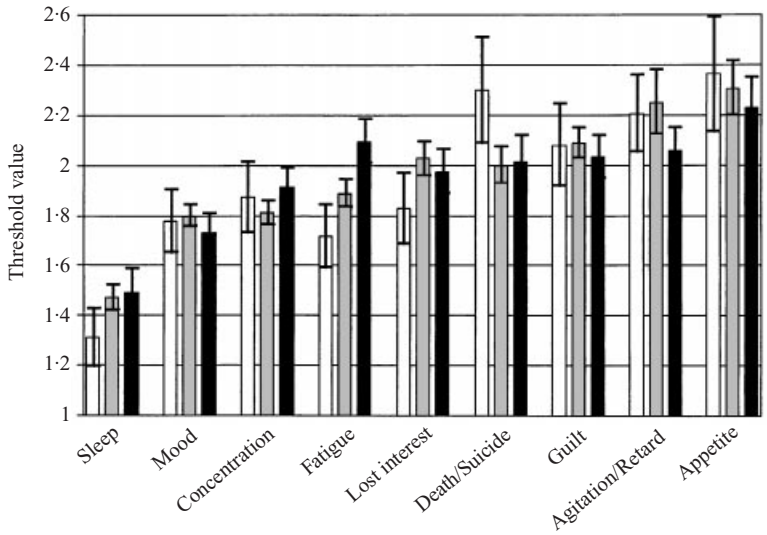


FIG. 2. Threshold values (from Item Response Theory analyses) with 95% confidence intervals for individual depressive symptoms as predictors of overall depression severity with centres grouped according to depression prevalence. For each symptom: □, low-prevalence centres (Shanghai, Nagasaki); ▨, medium-prevalence centres; and ■, high-prevalence centres (Rio, Santiago). $N = 5447$ for all symptoms.

level of depression severity. Some symptoms did show significant variation in threshold. Fatigue had a lower threshold (i.e. more common at lower levels of depression) in Rio and Santiago while thoughts of death or suicide had a higher threshold (i.e. rarely occurred until depression was more severe) at those same centres. Despite this variability, the pattern of symptom thresholds appeared generally consistent across the three groups: sleep disturbance, depressed mood and poor concentration were always among the low-threshold symptoms while feelings of guilt or worthlessness, psychomotor agitation or retardation, and changes in weight or appetite were always among the high threshold symptoms. This impression was confirmed by a formal test for differential item functioning. Goodness-of-fit for a differential item functioning model (allowing different symptom thresholds for the three prevalence groups) was not significantly better than that for a simpler model assuming identical symptom thresholds ($\chi^2 = 19{\cdot}4$, df = 18, $P = 0{\cdot}38$).

Two sets of secondary analyses used alternative methods to examine the consistency of the DSM-IV depression syndrome across prevalence groups. First, categorical factor analysis of the nine DSM-IV depression symptoms identified a first principle component accounting for over 60% of variance in all three groups. The pattern of symptom loadings on this component was remarkably similar in all three groups (highest loading for depressed mood and loss of interest and lowest for sleep disturbance and change in weight or appetite). Secondly, we examined the relative prevalence of DSM-IV Criterion A symptoms (depressed mood and loss of interest) *v.* other DSM-IV depressive symptoms. The relative prevalence rates were also remarkably similar across the three prevalence groups. Details of these secondary analyses are available on request.

Consistency of diagnostic threshold

As shown in Fig. 3, increasing number of DSM-IV depressive symptoms was uniformly associated with increasing SDS disability rating. The slope of this line, however, was less at high-prevalence centres (i.e. increasing depressive symptoms was associated with less increase in disability). A linear model including adjustment for age, sex, presence/absence of chronic medical illness, and treating physician's rating of medical illness severity confirmed a significant difference in slopes (t test for linear interaction between centre category and number of depressive symptoms = 3·51, $P < 0{\cdot}001$). To illustrate: patients with four DSM-IV depressive symptoms (i.e.

https://doi.org/10.1017/S0033291702005457 Published online by Cambridge University Press

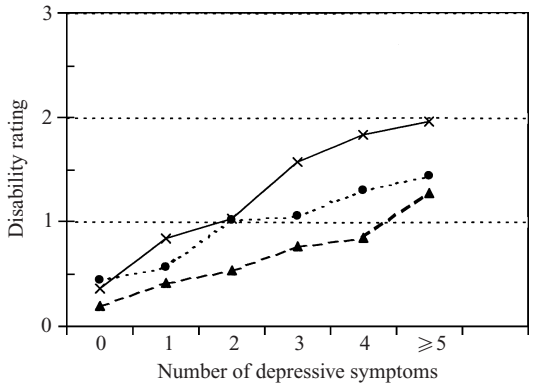


FIG. 3. Association between number of depressive symptoms and mean Social Disability Schedule global rating with centres grouped by baseline depression prevalence (×, Nagasaki, Shanghai; ▲, Rio, Santiago; ●, all others). For Social Disability Schedule ratings: 0, no disability; 1, mild disability; 2, moderate disability; and 3, severe disability. $N = 5420$.

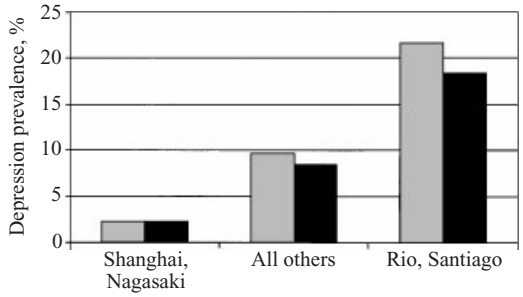


FIG. 4. Depression prevalence rates before (□) and after (■) restricting diagnosis to those with at least mild occupational role disability according to the Social Disability Schedule. $N = 5420$.

just below the threshold for diagnosis of major depression) had average disability ratings of 'Moderate' at low-prevalence centres, between 'Mild' and 'Moderate' at medium prevalence centres, and less than 'Mild' at high-prevalence centres. This pattern is consistent with the hypothesis that use of identical DSM-IV criteria identified patients with less severe illness and less impairment at high-prevalence centres.

To illustrate the potential magnitude of this effect, we re-calculated prevalence of current depressive disorder after requiring a rating of at least 'Mild' disability on the SDS. This restriction reduced the overall prevalence of major depression from 10·1% to 8·7%. As shown in Fig. 4, this restriction had no effect on rates of depression at low-prevalence centres, a slight effect at medium-prevalence centres, and a moderate effect at high-prevalence centres.

Course of depression

As shown in Table 2, both rate of onset and rate of persistence were higher at centres with higher prevalence rates. In a logistic regression model including adjustment for age and sex, centre category showed a significant linear association with both rate of onset ($t = 3·11$, $P = 0·002$) and rate of persistence ($t = 2·49$, $P = 0·013$).

DISCUSSION

We should acknowledge some significant limitations in the data used for these analyses. First, prevalence differences might reflect the specific primary care facilities studied rather than regional or national populations. For example, the highest-prevalence centres (Santiago, Rio, and Manchester) were public clinics serving socio-economically deprived areas. We certainly

Table 2. *One-year course of DSM-IV depressive disorder with study centres grouped by baseline depression prevalence rate* (*see text for details*)

Rate of onset ($N = 2776$)	Total N	Cases N	Weighted rate of onset %	95% CI
Nagasaki, Shanghai	410	12	0·8	0·1–1·5
All others	1963	142	3·6	2·8–4·5
Rio, Santiago	136	26	7·7	0·7–14·6
Rate of persistence ($N = 427$)	**Total N**	**Cases N**	**Weighted rate of persistence %**	**95% CI**
Nagasaki, Shanghai	34	3	9·3	0·0–19·8
All others	573	204	33·2	28·5–38·0
Rio, Santiago	87	40	44·4	31·0–59·8

https://doi.org/10.1017/S0033291702005457 Published online by Cambridge University Press

cannot claim that the Rio and Santiago centres accurately represent the population of South America or that the Nagasaki and Shanghai centres accurately represent all of East Asia. The pattern we observe, however (lower prevalence in Asian centres, higher in South American centres), is consistent with findings in several other community and primary care surveys (Araya *et al*. 1994, 2001; Lima *et al*. 1996; Weissman *et al*. 1996; Almeida-Filho *et al*. 1997). Secondly, interviewers were typically drawn from the same linguistic and cultural groups as the patients interviewed. Despite use of a completely structured diagnostic assessment and efforts to maintain cross-centre reliability, our data still reflect the interactions of patients and interviewers. Thirdly, we examine symptoms detected by systematic interviews rather than those presented to the physician. We have previously reported that presenting symptoms vary more across cultures than do symptoms elicited by systematic questioning. Finally, we did not examine differences between centres in the meaning or explanatory model attached to symptoms of depression. As we have reported elsewhere, patients' (Simon *et al*. 1999) and physicians' (Üstün & Sartorius, 1995) attributions of symptoms to physical or psychological causes varied widely across centres. Differences in attribution, however, do not explain prevalence differences detected by structured assessments. In fact, we have reported that use of identical diagnostic criteria led to higher prevalence of neurasthenia in Western Europe than in China or Japan (Üstün & Sartorius, 1995).

Any criteria or thresholds for dividing centres into low-, medium-, and high-prevalence groups will be arbitrary. We chose the scheme presented in this paper because the resulting low-prevalence group (Nagasaki and Shanghai) and high prevalence group (Rio and Santiago) had some geographic and cultural face validity. Furthermore, this grouping allowed our analyses to match results reported in other epidemiological studies (lower prevalence rates in Asia, higher in South America). We did, however, conduct a parallel set of analyses using an alternative grouping scheme: five centres with highest depression prevalence, five middle centres, and five with lowest prevalence. Results for this grouping were quite similar to those reported above. We believe that our primary findings are not sensitive to the thresholds used for classifying centres.

We do not find that prevalence differences are explained by cross-cultural differences in the form or validity of the depressive syndrome. When symptoms were assessed by systematic interview, the syndrome defined by DSM-IV showed a remarkably similar form at centres with low, medium, or high prevalence rates. Weissman *et al*. (1996) also found similar patterns of symptom reporting across community studies with widely different prevalence. Jacob *et al*. (1998) found that the underlying structure of common psychological symptoms varied little across cultures. We have previously reported (Furukawa & Goldberg, 1999) that the relationship between GHQ score and CIDI diagnosis (two independent measures of psychological disorder) was remarkably consistent across centres. Of course, no analysis of symptom patterns can definitively establish the clinical validity of the DSM depressive syndrome. We find no evidence, however, that the depressive syndrome defined by the DSM is more or less valid at centres with higher or lower depression prevalence. We conclude that apparent differences in depression prevalence cannot be attributed to 'category fallacy'.

When we use impact of depression on disability to compare diagnostic thresholds, we do find evidence that the same diagnostic criteria identified differing levels of depression severity. As we have reported previously (Simon *et al*. 1996, 1999), centres with higher prevalence of depressive symptoms generally had higher prevalence rates for the full range of psychological and somatic symptoms we assessed. These findings suggest that patients at the high-prevalence sites more often reported symptoms of all types – leading to higher apparent prevalence rates for depressive and anxiety disorders. Ortega & Rosenheck (2000) reported a similar finding in a study of post-traumatic stress disorder among US veterans. Prevalence was higher among Puerto Rican veterans, but diagnosis was associated with lower levels of disability in this high-prevalence group.

Prevalence differences are sometimes attributed to stigmatization of psychological distress in some cultures. In this and earlier reports (Simon *et al*. 1996, 1999), we found no evidence that patients at low-prevalence centres tended to

https://doi.org/10.1017/S0033291702005457 Published online by Cambridge University Press

deny psychological distress and report somatic symptoms. Instead, we found large and consistent differences in reporting across the full range of psychological and somatic symptoms. We could view this as a broad form of stigma affecting symptom reporting in general. Use of the term stigma in this context, however, implies that one level of reporting (presumably that in the US or Western Europe) is more accurate while another is artificially low.

Our use of disability to calibrate depression ratings presumes that interviewer-rated disability is less affected by language or culture than are self-reported depressive symptoms ratings. It is likely, however, that reporting of disability is influenced by the same factors that affect symptom reports. If so, cross-centre differences in symptom reporting (and the resulting variation in diagnostic threshold) could account for a larger portion of prevalence differences than indicated by our analyses. We did explore two alternative 'validators' of depression severity – number of self-reported disability days due to illness and recognition of psychological illness by the treating physician. Results for both measures were generally similar to those for interviewer-rated disability (i.e. increasing depression was associated with smaller increases in either disability days or physician recognition at high-prevalence centres). Definitive resolution of this issue would require a true 'gold standard' – a measure of depression severity proven invariant across language and culture.

One alternative explanation for our findings regarding depression and disability is that patients in high prevalence centres seek treatment for relatively mild depression while those in low-prevalence centres only seek treatment when depression causes significant impairment. Such a phenomenon could produce both the observed differences in prevalence and the observed variability in the relationship between depression and disability. Accepting such an explanation would lead to the same eventual conclusion: true prevalence differences are actually smaller than the 15-fold differences suggested by naïve analyses of our data.

We did observe differences in the course of depression consistent with true differences in prevalence; patients at high-prevalence sites had higher likelihood of both depression onset and persistence over 12 months. The observed differences (10-fold variation in onset rates, 3-fold variation in rate of persistence) are large enough to explain the overall 15-fold difference in prevalence rates. We must add, however, that a difference between centres in diagnostic threshold could also produce these differences in rates of onset and persistence. Our findings regarding course of depression are consistent with a true difference in depression prevalence, but they are also consistent with systematic differences in diagnostic threshold.

Weissman and colleagues (1996) have contrasted large cross-national differences in prevalence of depression with relatively uniform prevalence rates for bipolar disorder. Epidemiological studies of schizophrenia also show generally similar prevalence and symptom patterns across a wide range of culture and economic development (Sartorius *et al.* 1972, 1986). We see two quite different explanations for this contrast. First, social or environmental factors may have a stronger influence on depression than on more severe disorders such as bipolar disorder or schizophrenia. Alternatively, depression measures may be more difficult to apply across a range of cultures than are measures of bipolar or psychotic disorders. The defining characteristics of mania or psychosis (e.g. pressured speech, agitation, hallucinations) may be more easily distinguished from everyday experience than are the defining characteristics of depression (e.g. feeling down or depressed most of the day nearly every day).

Evidence for variability in diagnostic thresholds should not be misinterpreted as evidence that depressive illness is merely a social construction or a culture-bound phenomenon. The depressive syndrome defined in the US and Western Europe appeared in remarkably similar form across a wide range of language, culture and economic development. The strong association between depression and disability was universal. Furthermore, the absence of a crisp distinction between well and ill is not unique to depression. Numerous general medical conditions (such as hypertension or Type II diabetes) are defined by a severity continuum rather than a clear dichotomy. Standardizing a measure of depression severity across cultures may be more difficult only because depression is assessed through conversation rather than biochemical or mechanical measurement.

https://doi.org/10.1017/S0033291702005457 Published online by Cambridge University Press

Cross-national differences in the diagnostic threshold for major depression would have important implications for aetiological research. Results of the PPGHC study, along with earlier reports on cross-national differences in depression prevalence (Weissman *et al*. 1996), would suggest that risk factors for depression might be much more prevalent in South America and Europe and relatively uncommon in Asia. Our findings suggest that at least some of the observed differences in prevalence may instead reflect difficulties with cross-national measurement.

Cultural or linguistic variability in the threshold for depression diagnosis would also have important implications for treatment and treatment research. For conditions such as hypertension or Type II diabetes, the diagnostic threshold is often based on need for specific treatment – the severity level at which active treatment is superior to placebo or watchful waiting. Depression treatment studies in the US (Elkin *et al*. 1989) and UK (Paykel *et al*. 1988) find that antidepressant therapy is clearly superior to placebo only when depression is moderate or severe – roughly equivalent to the DSM-IV diagnostic threshold. Our findings suggest that the threshold for treatment and the criteria for defining successful treatment may show significant cross-national variability. An epidemiological study such as ours, however, cannot definitively determine whether cross-national variation in prevalence of DSM-IV major depression implies similar variation in prevalence of depression severe enough to warrant specific treatment. This question must be addressed by studying the effectiveness of specific depression treatments across a range of language, culture and economic development.

This study was supported by grants from the World Health Organization, the US National Institute of Mental Health (#MH47765 and #MH51338), Delagrange Laboratories and Synthelabo Pharmacie.

In addition to the authors, participating investigators in the Psychological Problems in General Health Care project include: Drs Orhan Ozturk and Murat Rezaki, Ankara, Turkey; Drs Costas Stefanis and Venetsanos Mavreas, Athens, Greece; Drs S. M. Channabasavanna and T. G. Sriram, Bangalore, India; Drs Hanfried Helmchen and Michael Linden, Berlin, Germany; Drs Wim van den Brink and Bea Tiemens, Groningen, The Netherlands; Drs Michael Olawatura and Oye Gureje, Ibadan, Nigeria; Drs Otto Benkert and Wolfgang Maier, Mainz, Germany; Drs Richard Gater and Steven Kisely, Manchester, United Kingdom; Drs Yoshibume Nakane and Shunichirou Michitsuji, Nagasaki, Japan; Drs Yves Lecrubier and Patrice Boyer, Paris, France; Drs Jorges-Alberto Costa e Silva and Louis Villano, Rio de Janeiro, Brazil; Drs Ramon Florenzano and Julia Acuna, Santiago de Chile; Drs Yan He-Quin and Xaio Shi Fu, Shanghai, People's Republic of China; Drs Michelle Tansella and Cesario Bellantuono, Verona, Italy. The study advisory group includes: Drs Jorges-Alberto Costa e Silva, David P. Goldberg, Yves Lecrubier and Hans-Ulrich Wittchen. Coordinating staff at WHO headquarters include: Dr Norman Sartorius and T. Bedirhan Üstün.

REFERENCES

Almeida-Filho, N., Mari, J. D. J., Coutinho, E., Franca, J. F., Fernandes, J., Andreoli, S. B. & Busnello, E. D. (1997). Brazilian multicentric study of psychiatric morbidity. *British Journal of Psychiatry* **171**, 525–529.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders, 4th edn*. American Psychiatric Association: Washington, DC.

Araya, R., Wynn, R., Leonard, R. & Lewis, G. (1994). Psychiatric morbidity in primary health care in Santiago, Chile. Preliminary findings. *British Journal of Psychiatry* **165**, 530–533.

Araya, R., Rojas, G., Fritsch, R., Acuna, J. & Lewis, G. (2001). Common mental disorders in Santiago, Chile: prevalence and socio-demographic correlates. *British Journal of Psychiatry* **178**, 228–233.

Bock, R. & Zimowski, M. (1995). Multiple group IRT. In *Handbook of Item Response Theory* (ed. W. van der Linden and R. Hambleton), pp. 115–122. Springer Verlag: New York.

Cochran, W. (1977). *Sampling Techniques*. Wiley: New York.

Dohrenwend, B. (1966). Social status and psychological disorder: an issue of substance and an issue of method. *American Sociology Reviews* **31**, 14–34.

Elkin, I., Shea, T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., Glass, D. R., Pilkonis, P. A., Leber, W. R., Docherty, J. P., Fiester, S. F. & Parloff, M. B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry* **46**, 971–982.

Furukawa, T. & Goldberg, D. P. (1999). Cultural invariance of likelihood ratios for the General Health Questionnaire. *Lancet* **353**, 561–562.

Goldberg, D. & Williams, P. (1988). *A User's Guide to the General Health Questionnaire: GHQ*. National Foundation for Educational Research: Windsor.

Jacob, K. S., Everitt, B. S., Patel, V., Weich, S., Araya, R. & Lewis, G. H. (1998). The comparison of latent variable models of non-psychotic psychiatric morbidity in four culturally diverse populations. *Psychological Medicine* **28**, 145–152.

Kleinman, A. (1977). Depression, somatization, and the new cross-cultural psychiatry. *Social Sciences and Medicine* **11**, 3–10.

Krause, N. & Carr, L. G. (1978). The effects of response bias in the survey assessment of the mental health of Puerto Rican migrants. *Social Psychiatry* **13**, 167–173.

Lima, M. S., Beria, J. U., Tomasi, E., Conceicao, A. T. & Mari, J. J. (1996). Stressful life events and minor psychiatric disorders: an estimate of the population attributable fraction in a Brazilian community-based study. *International Journal of Psychiatry and Medicine* **26**, 211–222.

Littlewood, R. (1990). From categories to contexts: a decade of the

https://doi.org/10.1017/S0033291702005457 Published online by Cambridge University Press

new 'cross-cultural psychiatry'. *British Journal of Psychiatry* **156**, 308–327.

Mezzich, J. E., Kirmayer, L. J., Kleinman, A., Fabrega, H., Jr., Parron, D. L., Good, B. J., Lin, K. M. & Manson, S. M. (1999). The place of culture in DSM-IV. *Journal of Nervous and Mental Disease* **187**, 457–464.

Ortega, A. N. & Rosenheck, R. (2000). Posttraumatic stress disorder among Hispanic Vietnam veterans. *American Journal of Psychiatry* **157**, 615–619.

Paykel, E. S., Hollyman, J. A., Freeling, P. & Sedgwick, P. (1988). Predictors of therapeutic benefit from amitriptyline in mild depression: a general practice placebo-controlled trial. *Journal of Affective Disorders* **14**, 83–95.

Phillips, D. & Clancy, K. (1970). Response biases in field studies of mental illness. *American Sociology Reviews* **35**, 503–514.

Robins, L. N., Wing, J., Wittchen, H.-U., Helzer, J. E., Babor, T. F., Burke, J., Farmer, A., Jablenski, A., Pickens, R., Regier, D. A., Sartorius, N. & Towle, L. H. (1988). The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems in different cultures. *Archives of General Psychiatry* **45**, 1069–1077.

Sartorius, N., Shapiro, R., Kimura, M. & Barrett, K. (1972). WHO international pilot study of schizophrenia. *Psychological Medicine* **2**, 422–425.

Sartorius, N., Jablensky, A., Korten, A., Ernberg, G., Anker, M., Cooper, J. E. & Day, R. (1986). Early manifestations and first-contact incidence of schizophrenia in different cultures. A preliminary report on the initial evaluation phase of the WHO Collaborative Study on determinants of outcome of severe mental disorders. *Psychological Medicine* **16**, 909–928.

Sartorius, N., Üstün, T. B., Costa e Silva, J. A., Goldberg, D. P., Lecrubier, Y., Ormel, J., VonKorff, M. & Wittchen, H.-U. (1993). An international study of psychological problems in primary care. *Archives of General Psychiatry* **50**, 819–824.

Simon, G. E., Gater, R., Kisely, S. & Piccinelli, M. (1996). Somatic symptoms of distress: an international primary care study. *Psychosomatic Medicine* **58**, 481–488.

Simon, G. E., VonKorff, M., Piccinelli, M., Fullerton, C. & Ormel, J. (1999). An international study of the relation between somatic symptoms and depression. *New England Journal of Medicine* **341**, 1329–1335.

Üstün, T. & Sartorius, N. (1995). *Mental Illness in General Health Care*. John Wiley and Sons: New York.

Weissman, M. M., Bland, R. C., Canino, G. J., Faravelli, C., Greenwald, S., Hwu, H. G., Joyce, P. R., Karam, E. G., Lee, C. K., Lellouch, J., Lepine, J. P., Newman, S. C., Rubio-Stipec, M., Wells, J. E., Wickramaratne, P. J., Wittchen, H. & Yeh, E. K. (1996). Cross-national epidemiology of major depression and bipolar disorder. *Journal of the American Medical Association* **276**, 293–299.

Wiersma, D., Jong, A. & Ormel, J. (1988). The Groningen Social Disabilities Schedule: development, relationship with ICIDH and psychometric properties. *International Journal of Rehabilitation Research* **11**, 213–224.

https://doi.org/10.1017/S0033291702005457 Published online by Cambridge University Press