continue, because rates for such categorical diagnoses seem to have an administrative impact.

The scientific value of this categorical approach is less certain. Much greater scientific value will come from the analysis of dimensional scales such as the Kessler Psychological Distress Scale (K–10; Kessler *et al*, 2002) and 12-item General Health Questionnaire (GHQ–12; Goldberg & Williams, 1988). Unlike the diagnostic interview, which can take over an hour, these take only a few minutes to administer, even for persons with limited literacy. The scales cover the full range of symptoms of epidemiological interest and they detect the sub-clinical level of morbidity that is of such relevance to disease burden. Their continuous scores avoid the loss of statistical information that comes with categorical diagnoses. They are also less open to the biases that make international comparisons of prevalence rates so difficult to interpret. As Rose (1993) emphasised, the distribution of scores can be used to characterise whole populations. The K-10 has been included in many of the surveys, but has not yet been used to test aetiological hypotheses.

Further large-scale epidemiology needs to contribute to knowledge of aetiology. For example, in addition to obtaining psychiatric measures and exposures to adversity, obtaining genetic data is now feasible. This would allow the study of interactions between genotype and environment in very large samples. National surveys of mental health are major epidemiological undertakings. Their yield can now be expanded.

References

American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders* (4th edn) (DSM–IV). American Psychiatric Press.

Andrews, G., Henderson, S. & Hall, W. (2001) Prevalence, comorbidity disability and service utilisation. Overview of the Australian National Mental Health Survey. *British Journal of Psychiatry*, 17, 145–153.

Compton, W. M., Conway, K. P., Stinson, F. S., *et al* (2006) Changes in the prevalence of major depression and comorbid substance use disorders in the United States between 1991–1992 and 2001–2002. *American Journal of Psychiatry*, 163, 2141–2147.

Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., *et al* (World Mental Health Survey Consortium) (2004) Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA*, 291, 2581–2590.

Goldberg, D. P. & Williams, P. A. (1988) *User's Guide to the GHQ*. NFER Nelson.

Kessel, N. (1965) Are international comparisons timely? *Milbank Memorial Fund Quarterly*, 2, 199–212.

Kessler, R. C., Andrews, G., Colpe, L. J., *et al* (2002) Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32, 959–976.

Mohammadi, M–R., Davidian, A., Noobala, A. A., *et al* (2005) An epidemiological survey of psychiatric disorders in Iran. *Clinical Practice and Epidemiology in Mental Health*, 1, 16–23.

Rose, G. (1993) Mental disorders and the strategies of prevention. *Psychological Medicine*, 23, 553–555.

World Health Organization (1992) *The ICD–10 Classification of Mental and Behavioural Disorders. Diagnostic Criteria for Research*. WHO.

ORIGINAL PAPER

A re-evaluation of the yield from national surveys of mental health

Ronald C. Kessler PhD[1] and T. Bedirhan Üstün MD[2]

[1]Department of Health Care Policy, Harvard Medical School, email kessler@hcp.med.harvard.edu

[2]Classifications, World Health Organization

Having spent a considerable amount of time thinking about the uses of large-scale descriptive psychiatric epidemiological needs assessment surveys in our capacity as co-directors of the World Health Organization's World Mental Health Survey Initiative, we agree with many of the conclusions of Henderson and Andrews. Most importantly, we agree:

❍ that among the most important benefits of these surveys have been their political value in documenting high prevalence and high disability

❍ that the time has come to expand the focus to study causes.

However, we also disagree with Henderson and Andrews on several points.

The first paragraph of their article raises a concern that the resources used to carry out psychiatric epidemiological surveys might be better used to address the problem of unmet need for treatment. This criticism is ill-conceived. These surveys make it clear that an increase in healthcare resources is required to address the enormous problem of unmet need for treatment of mental disorders. Needs assessment surveys must be carried out to document this unmet need, as well as to track changes in unmet need as mental healthcare policies change. It is a mistake to set the cost of carrying out these surveys in opposition to the need for increased treatment resources. This is especially so in light of the fact that the resources needed to carry out a typical large-scale psychiatric epidemiological survey (e.g. face-to-face interviews with a representative sample of 5000 respondents) are trivial in relation the resources needed to provide treatment. The former resources amount to about 25 person-years. In a country of 40 million people with a 15% prevalence of mental illness, this comes to roughly 1 minute per person with a mental illness. Not much good

can be done in 1 minute of treatment. Much more good could be done by investing that minute in a needs assessment survey that documents the magnitude of unmet need for treatment, obtains information about misallocation of treatment resources, and provides a firm foundation for seeking expanded treatment resources and implementing more efficient outreach and treatment initiatives.

We also take issue with the characterisation by Henderson and Andrews of recent psychiatric epidemiological surveys being based exclusively on categorical diagnostic assessments and excluding environmental risk factors. The World Mental Health Surveys include structured versions of disorder-specific symptom severity scales, such as the Quick Inventory of Depressive Symptomatology (Rush *et al*, 2003) for major depression and the Panic Disorder Severity Scale (Houck *et al*, 2002) for panic disorder. In addition, all respondents are administered the dimensional Kessler Psychological Distress Scale (K–10) (Kessler *et al*, 2002), as well as separate dimensional assessments of disability. With regard to environmental risk factors, the surveys include a detailed assessment of childhood adversity, which is a focus of ongoing causal analysis.

We agree with Henderson and Andrews that differences in fundamental survey conditions make it most useful to study comparative, within-country correlates. We also agree that socio-demographic correlates, such as the higher prevalence of depression among women than men, are by now so well known that we need to investigate causes rather than merely replicate these associations in new surveys. The survey analyses are doing just that. For example, we are examining the role of cross-national differences in age at menarche and in gender roles in accounting for time–space variation in the association between gender and depression. Preliminary results suggest that the gender gap in depression is becoming significantly smaller in numerous countries because of changes in the social roles of women (as indicated by trends in such things as rates of female labour force participation, access to birth control and delays in age at first child-bearing).

We agree with the spirit of Henderson and Andrews' conclusion that future surveys might profitably use brief screening scales rather than diagnostic interviews. However, the use of screening scales needs to be coupled with diagnostic interviews in targeted sub-samples to realise their full potential. At least three different types of integration are important in this way. We briefly discuss each of these three below.

There is a key question in this regard. What do we gain and what do we lose by using short screening scales rather than long diagnostic interviews to assess mental illness in community epidemiological surveys? The first part of the answer is very simple: the brevity of screening scales allows them to be included in ongoing omnibus government health-tracking surveys, whereas diagnostic interviews are too lengthy to be included in omnibus surveys. By the term 'omnibus survey' we mean multi-purpose surveys that ask respondents about the presence, severity and course of physical and mental disorders, often along with questions about health service utilisation, health protective behaviours and health risks.

Screening scales are commonly included in omnibus health surveys, making it possible to assess many more people than in psychiatric epidemiological surveys. For example, the K–6 screening scale of non-specific psychological distress (Kessler *et al*, 2002) is included in all three major omnibus government health-tracking surveys in the USA:

❍ the Behavioral Risk Factor Surveillance Survey (BRFSS; http://www.cdc.gov/brfss; 350 000 respondents per year interviewed by telephone in monthly replicates representative of each city, county and state in the USA)

❍ the National Household Survey of Drug Use and Health (NHSDUH; http://www.oas.samhsa.gov/nhsda.htm; 70 000 respondents per year interviewed face-to-face in quarterly replicates representative of the largest states as well as of each census division in the USA)

❍ the National Health Interview Survey (NHIS; http://www.cdc.gov/nchs/nhis.htm; 50 000 respondents per year interviewed face-to-face in monthly replicates representative of each census division in the USA).

Clinical calibration studies show that the K–6 is an excellent screening scale for the global assessment of broadly defined DSM–IV anxiety and mood disorders (Kessler *et al*, 2003). Simple transformation rules consequently can be applied to K–6 scores to estimate the prevalence and correlates of recently occurring anxiety and mood disorders. To do this, though, we need to have faith in the calibration rules that link K–6 scores with DSM–IV diagnoses. This requires *ongoing* calibration in the omnibus surveys in which the K–6 is administered, because we need to know if any significant time trends in estimated prevalence are due to changes in true prevalence or to changes in the concordance of the K–6 with true prevalence. Recognising the importance of ongoing calibration, an initiative is currently under way in the USA to administer follow-up diagnostic interviews to a representative sub-sample of respondents in government omnibus health surveys, to refine K–6 calibration rules over time. This is the first of the three important ways in which clinical interviews can be integrated with screening scales.

Screening scales are limited because they include only broad-gauged information. When we discovered, for example, that the K–6 scores in the NHIS increased dramatically after the terrorist attacks of 11 September 2001, we had no way to know the extent to which this was owing to rises in adjustment disorder, post-traumatic stress disorder, major depression, other disorders, or various combinations of disorders. Nor could we tell the extent to which the K–6 increases were owing to first onsets versus recurrences of mental disorders. It would have been fairly simple to obtain this information with in-depth follow-up interviews. We might even have followed such cases over time to study the persistence and severity of post-disaster mental illness, patterns of help-seeking and modifiable barriers to treatment. Such information could have been of great value to the authorities. This is the second important way in which clinical interviews could be integrated with screening scales. An initiative of this sort is currently under way in the USA to embed follow-up psychiatric diagnostic risk factor interviews into the BRFSS in the wake of future large-scale natural and human-made disasters (Kessler *et al*, in press).

The same limitations of screening scales also exist in interpreting cross-sectional associations. Furthermore, typical omnibus surveys include only superficial questions about treatment of emotional problems, and no questions about barriers to such treatment or about disablements caused

by emotional problems. More in-depth assessments are needed to collect such data. An argument consequently could be made that a probability sub-sample of respondents in omnibus health surveys should be administered in-depth follow-up psychiatric interviews to collect this additional information. The second-stage sub-samples would over-sample respondents with high K–6 scores so as to enrich the sample for cases, making this the third important way in which clinical interviews could be integrated with screening scales.

We are currently investigating the possibility of using this design in the USA. Rather than continue to replicate the National Comorbidity Survey (NCS) in point-in-time surveys each decade (Kessler et al, 1994, 2005), we would administer an NCS interview to a probability sample of omnibus survey respondents every month of every year, accumulating a sample of 12 000 cases over each decade. This approach has several advantages over the stand-alone survey method:

❍ increased statistical power to study respondents with mental illness by over-sampling those who are 'positive' according to the screening scales
❍ expansion of information on respondents from the omnibus survey
❍ fine-grained time trend information
❍ the ability to modify assessments quickly, rather than once a decade, when modifications would be useful.

Finally, as omnibus surveys are typically very large, we could target selection of follow-up sub-samples in ever-changing ways over time for purposes of refining causal analyses in a case–control framework. Respondent burden would increase and further elaboration would be needed to launch longitudinal studies, but this general approach holds much promise to realise the greater potential of general-purpose psychiatric epidemiological surveys along the lines so rightly urged by Henderson and Andrews.

References

Houck, P. R., Spiegel, D. A., Shear, M. K., et al (2002) Reliability of the self-report version of the Panic Disorder Severity Scale. Depression and Anxiety, 15, 183–185.

Kessler, R. C., McGonagle, K. A., Zhao, S., et al (1994) Lifetime and 12-month prevalence of DSM–III–R psychiatric disorders in the United States. Archives of General Psychiatry, 51, 8–19.

Kessler, R. C., Andrews, G., Colpe, L. J., et al (2002) Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychological Medicine, 32, 959–976.

Kessler, R. C., Barker, P. R., Colpe, L. J., et al (2003) Screening for serious mental illness in the general population. Archives of General Psychiatry, 60, 184–189.

Kessler, R. C., Berglund, P., Demler, O., et al (2005) Lifetime prevalence and age-of-onset distributions of DSM–IV disorders in the National Comorbidity Survey replication. Archives of General Psychiatry, 62, 593–602.

Kessler, R. C., Keane, T. M., Ursano, R. J., et al (in press) Sample and design consideration in post-disaster mental health needs assessment tracking surveys. International Journal of Methods in Psychiatric Research.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., et al (2003) The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS–C), and self-report (QIDS–SR): a psychometric evaluation in patients with chronic major depression. Biological Psychiatry, 54, 573–583.

ORIGINAL PAPER

A comment on the yield from national surveys of mental health

Rachel Jenkins,[1] Howard Meltzer,[2] Terry Brugha[3] and Paul Bebbington[4]

[1]Director, WHO Collaborating Centre, Health Services and Population Research Department, King's College London, Institute of Psychiatry, London, UK, email R.Jenkins@iop.kcl.ac.uk

[2]Professor of Mental Health and Disability, Section of Social and Epidemiological Psychiatry, University of Leicester, UK

[3]Brandon Mental Health Unit, Section of Social and Epidemiological Psychiatry, Leicester General Hospital, UK

[4]Professor of Social and Community Psychiatry, Department of Mental Health Sciences, University College London, UK

Henderson and Andrews have written a timely paper to examine the yield from national surveys, and pose questions of value for money, survey methods, delineation of pathology by categorical boundaries rather than dimensions, and breadth of risk factors examined. We would like to address the points they raise, exemplified by the purposes (Jenkins et al, 1997) and yield of the British survey programme.

First, Henderson and Andrews ask whether national surveys give good value, when considered against the unmet need for services. The sheer scale of unmet need has in fact been delineated only by national surveys – without them we would not know the scale of need, and how far it is met and unmet by clinical services (Bebbington et al, 2000). Moreover, without the use of repeat surveys, we would not have known about the considerable changes in Britain in the use of services and the delivery of treatment (Brugha et al, 2004).

Henderson and Andrews focus on surveys using the DSM diagnostic criteria, standardised lay interviews and data derived solely from self-report. However, the World Mental Health Survey (WMHS) includes a clinical component in some countries, although psychosis, developmental disorders and personality disorder are not covered throughout the WMHS. The British adult survey programme does use lay interviewers to administer the Clinical Interview Schedule – Revised (CIS–R) to cover anxiety and depressive symptoms, but it is augmented by clinical assessments using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) and the Structured Clinical Interview for DSM–IV Axis II Personality Disorders (SCID–II), in the case of adults (e.g. Brugha et al,