Psychological Medicine, 1987, **17**, 275–282
Printed in Great Britain

EDITORIAL

Are the lifetime prevalence estimates in the ECA study accurate?[1]

The first basic goal of the NIMH Epidemiologic Catchment Area programme (ECA) was to estimate rates of prevalence and incidence of specific mental disorders (Eaton *et al.* 1984), and the preliminary papers in the October 1984 issue of the Archives of General Psychiatry are highlighted by an editorial (Freedman, 1984) which notes the assistance the ECA data bank can give to considerations of health service needs, health finances and manpower needs. Such assistance is predicated on one key issue – the accuracy of the 'head count'. In the ECA study, six-month and lifetime prevalence estimates of psychiatric morbidity have been determined. The ECA study design will not be reviewed as it has been comprehensively described elsewhere (Eaton *et al.* 1984; Regier *et al.* 1984). This review will principally consider the accuracy of the lifetime diagnoses. Lifetime prevalance is established by measuring at a single point in time the proportion of a population who have ever had the disorder under investigation. In relation to the ECA study, Robins *et al.* (1984) argued that lifetime prevalence estimates have several advantages: they are necessary for calculating rates of current diagnoses and annual first-episode incidence rates, and are useful for studying the aetiology of disorders.

There are three reasons to suspect the accuracy of the lifetime diagnostic estimates – the ratio of six-month and lifetime prevalence data, the discordance with previous estimates of lifetime morbidity, and the curvilinear association of lifetime prevalence estimates with increasing age.

First, the ratios of six-month and lifetime prevalence data will be considered. In terms of overall morbidity, the six-month : lifetime ratio (Myers *et al.* 1984) was 13·2 : 24·2 in New Haven, 12·6 ; 23·0 in Baltimore and 11·6 : 25·2 in St Louis. Thus the data, consistent across the three sites, suggest that the likelihood of being mentally ill in the preceding six months was approximately half the lifetime likelihood of being mentally ill, an extraordinarily high chance when it is kept in mind that the ECA reports consider some 9000 community adult residents. If true, such a striking ratio would indicate the marked recurring nature and/or chronicity of psychiatric disorder. On the other hand, such ratios could emerge as a consequence of response error, defined by Eaton *et al.* (1984)) as 'any source of error that originates in the respondent as he or she generates the information for the project'. In this instance, such ratios might be produced by inflation of the six-month data and/or by lowering of the lifetime prevalence reports by impaired recall and other factors. The ratios are not constant across diagnostic categories but do show some broad trends when calculation is made of the separately published six-month (Myers *et al.* 1984) and lifetime (Robins *et al.* 1984) prevalence data. Many would consider them counter-intuitive. For antisocial personality and substance use disorders the ratio is approximately 1 : 3, while for the remaining disorders (e.g. schizophrenia, schizophreniform disorder, manic episode, major depressive episode, phobia) the ratios are all less than 1 : 2, with consistency across sites being quite high. If valid, these findings, suggest a relatively good prognosis for antisocial personality and substance use disorders, and a far more chronic or recurring pattern for the functional psychoses and affective disorders.

As these findings suggest that establishing a lifetime diagnosis is likely to depend on the particular psychiatric condition, further consideration will be limited to depressive disorders, because they have been shown to be the most common diagnosis in studies of community groups (Weissman *et al.* 1978) and because their first occurrence may occur throughout adult life (Robins *et al.* 1984) and is not age-limited by DSM-III definition, such as occurs for schizophrenia. The six-month : lifetime prevalence ratio of dysthymia cannot be estimated (as a two-year period of depressed affect is

[1] Address for correspondence: Professor G. Parker, University of New South Wales, Psychiatry Unit, Prince of Wales Hospital, Randwick 2031, Australia.

https://doi.org/10.1017/S0033291700024806 Published online by Cambridge University Press

required for such a diagnosis, so preventing a six month estimate in the ECA study) but the ratios are approximately 1:1·5 for 'manic episode' and 1:1·8 for 'major depressive episode' across the three sites, again suggesting a chronic and/or recurrent pattern, or the possibility of a response error in this narrowed list of disorders.

The second reason for querying the accuracy of ECA lifetime diagnoses emerges from comparison of the findings with previous studies of depressive disorders. It would be inappropriate to make comparison with those early studies (e.g. Midtown Manhattan) where the data were not presented in terms of diagnosis. Additionally, it would be inappropriate to compare ECA data with surveys using depression inventories (which tend to suggest that up to 20% of community subjects have significant depressive symptoms) for the same reason. It is appropriate to consider only those studies which have been termed (Regier *et al.* 1984) the 'third generation of epidemiological research studies', in that case-finding techniques are used with pre-defined operational criteria to delineate discrete depressive disorders.

Before considering lifetime prevalence data, it would be useful to note findings from comparative studies of recent or current depression. In the ECA study, the six-month prevalence of a major depressive episode ranged from 2·2 to 3·5% in the three sites. A key comparison study, as it was conducted earlier in one of the ECA site communities, is that by Weissman *et al.* (1978) of 515 New Haven subjects interviewed by non-psychiatrist raters who administered the Schedule for Affective Disorders and Schizophrenia – Lifetime Version (or SADS-L). Subjects were classified on the basis of Research Diagnostic Criteria (RDC) to derive diagnoses corresponding to, but not necessarily identical with, DSM-III groupings, either in terms of duration (i.e. one week rather than two being specified) or number of symptom criteria. It must be noted, however, that there had been a considerable drop-out of respondents in the final third phase of this study, perhaps inflating morbidity estimates which were calculated as 3·7% (definite) to 4·3% (definite and probable). Nevertheless, the New Haven estimates correspond to independent studies in other regions. In England, Bebbington *et al.* (1981) administered the Present State Examinations (PSE) to a community sample of 800 adults, and assigned subjects above level 5 on the Index of Definition as being a psychiatric 'case'. In that study, the one-month prevalence of depressive disorder was 7·0%. That estimate is close to the point prevalence estimate (5·6% for depressive, 2·5% for 'mild mixed affective' disorders) reported by Murphy *et al.* (1984) in their Stirling County follow-up, where they used an algorithm which, they claimed, approximated to DSM-III and RDC criteria. In a major review, Boyd & Weissman (1981) considered 16 community surveys and estimated, at least for those studies using the new diagnostic techniques (e.g. SADS, RDC, PSE, DSM-III), that in industralized nations the point prevalence of non-bipolar depression was 3% for men and 4–9% for women. The ECA six-month prevalence data for major depressive episodes appear then to be slightly lower than comparable estimates, which might be expected when dysthymic disorder was excluded from the six-month data. If the lifetime diagnosis of dysthymia data are added to six-month prevalence data for major depression, as tabulated in one report (Myers *et al.* 1984), then the total ranges from 4·3–7·0% across the three sites, and suggests quite strong consistency with comparable surveys.

Turning to the lifetime risk of a depressive episode in the ECA study, then the estimates were 4–7% for major depression and 2–4% for dysthymia. These figures contrast with the study by Weissman & Myers (1978), where the lifetime rates were estimated at 18·0–20·0% for major depression, 8·6–9·2% for minor depression, and 24·7–26·7% for both types of depression. Such estimates correspond to findings in a multi-phase study by Reich *et al.* (1980) using similar measures and which found a strikingly similar lifetime prevalence estimate of 8–12% for men and 20–26% for women, although the sample was not a representative sample of the general population. Additionally, Bromet *et al.* (1986) had lay interviewers administer the SADS to a semi-rural community sample of Pennsylvanian women and calculated a lifetime prevalence rate of major depression (RDC) of 29% at baseline and 21% on retesting eighteen months later. Despite the limitations in comparing studies (and while recognising that lifetime rates are influenced by age distribution, educational levels, availability and effectiveness of treatment, migratory factors,

https://doi.org/10.1017/S0033291700024806 Published online by Cambridge University Press

differential mortality and measurement variables), as well as recognising that RDC criteria for major depression are less stringent than DSM-III criteria, their consistency suggests that the ECA estimate of lifetime depressive disorders is strikingly low.

Anomalies in the prevalence data when examined by age group (Robins *et al.* 1984) provide the third reason to suspect the accuracy of the ECA lifetime data. Unless depressive disease is associated with a markedly increased mortality rate or a most unusual period or cohort effect, the lifetime prevalence should increase with the age of the subjects. Recalculation of the ECA data published in Robins *et al.* (1984, Table 6) establishes that the lifetime risk for major depressive episode is 5% in those aged 18–24 years, 9% in those 25–44 years, 5% in those aged 45–64 years, and 2% in those aged more than 65 years. The lifetime prevalence data across the three centres for dysthymia are respectively 2%, 4%, 3% and 2% for those four age groupings. Such a phenomenon is consistent across all three sites. Robins *et al.* (1984) stated that it is 'not yet clear why the rates in the elderly are so low', with 'cognitive impairment' being the only one of the 15 diagnoses behaving as predicted. The curvilinear nature of the lifetime prevalence data for depressive disorders when examined by age suggests that recent and past episodes are differentially remembered, and that the reporting of lifetime prevalence data may be influenced by recent if not current episodes. The low prevalence in older age groups is not unique to the ECA study or to the diagnostic measuring instrument, being demonstrated in most previous epidemiological studies that have analysed separate age groups (e.g. Weissman & Myers, 1978). The fact that it has been shown in studies undertaken up to two decades apart in time argues against a period or cohort effect, and supports a recall or non-immediacy effect in older age subjects.

These concerns suggest a focus for the remainder of this paper – do the current case-finding techniques in psychiatry allow lifetime prevalence estimates to be made accurately in community studies? This may be best addressed first by examining the established properties of the history-taking schedules (such as the SADS-L and DIS) that have been used in such recent enquiries.

A number of representative studies will be reviewed, first considering the most commonly-cited lifetime measure, the SADS-L. Mazure & Gershon (1979) examined the reliability of the SADS-L measure with two interviewers assessing a heterogeneous group of patients, first-degree relatives and controls some 7 months apart. The test–retest reliability was high, with a kappa value of 0·79, although on the test occasion the interviewers were 'told of any previously held diagnosis'. The authors comment that most of those with major depression were consistently diagnosed, but that minor depression was an unstable diagnostic entity, with not one individual being consistently diagnosed and with reproducibility being 'essentially absent'. Andreasen *et al.* (1981) assessed the reliability of lifetime diagnosis of ill and well relatives of probands using the SADS-L and after making diagnoses on the basis of RDC criteria. Ratings compared against independent ratings made in the preceding 6 months by different interviewers established high agreement (the intraclass R statistic being used) in assessing the age at last episode (+0·78), moderate agreement (+0·59) in making the diagnosis of 'primary major depressive disorder', but no agreement in assessing the number of episodes of 'depressive syndrome' (−0·01). Even for interviews conducted on the same day (by different raters) there was no agreement in estimating the number of episodes of major depressive disorder (+0·18). Leckman *et al.* (1982) designed a more complex study. They interviewed 215 probands who either had an affective disorder or who were not mentally ill, interviewed a percentage of their spouses and first-degree relatives for a corroborative report, and consulted medical records for those in the patient sub-group. Using those sources they derived 'best estimate' lifetime diagnoses which were compared against SADS-L interview data and RDC-derived diagnoses. The sensitivity for major depression was 86%, suggesting that few 'non-cases' were misclassified. The level of agreement (Cohen's K coefficient) between 'best estimate' diagnoses and the clinicians' diagnoses was 0·86 for major depression and 0·53 for minor depression. Bromet *et al.* (1986) examined test–retest reliability of lifetime depression in a community population and over a relatively lengthy interval (18 months), having trained lay interviewers to administer the SADS. Poor temporal stability (with only 38% consistently reporting major depression at both interviews)

https://doi.org/10.1017/S0033291700024806 Published online by Cambridge University Press

was established and considered to be due to the semi-structured nature of the SADS, the clinical status of the subjects and the lengthy interval, rather than to fatigue effects at the second interview or to reflect the use of lay interviewers.

Turning to the Diagnostic Interview Schedule (DIS), the instrument developed by Robins *et al.* (1982) for use in the ECA studies, those authors noted problems in assessing validity. They stated that the preferred way of measuring validity would have been to compare the DIS against a different, validated instrument, but that they were unable to do so because 'no such validated instrument existed at the time'. Instead they examined 'procedural validity', a term advanced by Spitzer & Williams (1980) which, in effect, considers the concordance between diagnoses obtained using the instrument by lay interviewers against those obtained by psychiatrists independent of, and blind to, the lay interviews. This assessment, in essence an examination of reliability (although the stimulus situation is not held constant), established a moderate level of agreement for depressive episodes (kappa = 0·63), accuracy being lowered when the symptoms were not present in the previous year, with 'depressive episode' having the lowest kappa coefficient (0·40) of the 7 disorders examined. Incidentally, the kappa of 0·63 is very similar to the kappa coefficient of 0·68 for 'major affective disorders' obtained in phase one of the DSM-III field trials (American Psychiatric Association, 1980), where paired clinicians evaluated individuals at interview (raters being free either to interview together or separately) and with access to all clinical information. Kappa coefficients for 'other' specific affective disorders (0·49) and 'atypical' affective disorders (0·29) in that field trial were clearly lower and suggested moderate to low inter-rater reliability.

Helzer *et al.* (1985) have now examined a sub-sample of St Louis ECA respondents, selected so as to ensure adequate representation of those with and without diagnoses at initial assessment. The original DIS-derived diagnoses generated by lay interviewers were compared against reinterview diagnoses generated by psychiatrists, first using the DIS, and subsequently after a free-form examination. When the results of the two latter procedures were compared for major depressive episode the overall percentage agreement of 89% was high, but the lowest of the many different diagnostic groups. The specificity (of 99%) was impressive, suggesting that the DIS categorized few subjects as having major depression who had not been so diagnosed by the physician assessor. The sensitivity of 63% was less impressive, with more than one-third of those diagnosed by the physician as having a major depression not being so categorized by the physician-generated DIS diagnosis. When lay-generated and physician-generated DIS categories of major depression were compared, the overall agreement (kappa = 0·33) was weak. Only 42 of the 101 subjects diagnosed by the physician-derived DIS as having major depression were so diagnosed by the lay-derived DIS. The authors concluded that the lay interviewers significantly underdiagnosed the lifetime diagnosis of major depression, but the dissonance could reflect low consistency in reporting symptoms on the two occasions or difficulties with use of the DIS. The researchers established that threshold cases, where the number of symptoms is at or near the threshold of diagnostic definition, contributed considerably to inter-rater disagreement. Finally, they noted that the DIS underestimated depressive episodes by the same amount when given either by the lay interviewers or physicians, and drew attention to the greater disagreement between lay and physician raters than has been reported in previous studies when, generally, clinical groups have been studied. A similar finding of low agreement (kappa = 0·25) between lay interview DIS-derived diagnoses and standardized psychiatric diagnoses of major depression has also been described by Anthony *et al.* (1985), although they concentrated on the assessment of active mental disorder occurring within one month of the interview. In their study, however, the results were the reverse of those reported by Helzer *et al.* (1985). Of 61 Baltimore ECA respondents originally assessed as having major depressive disorder, only 13 were so diagnosed by a psychiatrist at the second assessment. Anthony *et al.* (1985) drew attention to several reasons for the discrepancies: invalid responses by the subject, lack of clarity in some DSM-III criteria to determine recency and activity of the disorder, the failure of the DIS to cover all DSM-III critera, overinclusive questions in the DIS, and the degree of reliance by the DIS on subjects' reports. Robins (1985) has acknowledged many of these and offered additional reasons, including clinical change in the respondents and failure to explain the purpose of the second interview.

https://doi.org/10.1017/S0033291700024806 Published online by Cambridge University Press

Bringing together these studies, some general conclusions may be drawn. Given the fact that the definition of psychiatric 'caseness' is difficult when the dimensional nature of much psychiatric disorder is recognized, the operational criteria of DSM-III have at least limited criterion variance to a manageable degree. Nevertheless, inter-rater reliability assessments, as undertaken in the DSM-III field trials (American Psychiatric Association, 1980), have established that clinicians show only moderate agreement in their diagnostic decisions about current psychiatric disorder. Evaluative studies have concentrated, in the main, on assessing the properties of measures of current morbidity. When lifetime diagnostic decisions are to be made, the capacity of case-finding techniques to evaluate the prevalence of disorder remains to be established. Klerman (1985) has written, in relation to the ECA study, that it would have been desirable if more extensive evidence as to the reliability and validity of the DIS and DSM-III criteria had been available, but that a decision to proceed had been forced by the financial and political climate, and that funding would probably not have been available if the project had been delayed. Helzer *et al.* (1985) accept that it would have been desirable to test the interview in the general population before embarking on the ECA study, but argued that obtaining a sufficiently large sample to generate enough cases of a particular diagnosis was a 'practical impossibility'. Nevertheless, at the time of ECA commencement, early studies of the properties of the DIS and of using lay interviewers had appeared encouraging, for reasons which may be articulated now in hindsight.

It is important to emphasize that the level of agreement between DIS diagnoses generated by lay interviewers and independent interviews by psychiatrists was examined in only one study (Robins *et al.* 1981, 1982) before the ECA survey commenced and suggested a moderate level of agreement, at least similar to that of the DSM-III field trials comparing psychiatrists' judgments, so suggesting that the use of lay interviewers was appropriate. However, as Anthony *et al.* (1985) note, the subjects in that study and 4 later studies were 'almost all psychiatric patients undergoing treatment or patients with a history of treatment'. The more recently reported studies of resampled ECA subjects (Anthony *et al.* 1985; Helzer *et al.* 1985) show very poor agreement between lay interviewer-generated DIS diagnoses and psychiatrist judgments, at least for major depressive episode. In fact, if the findings by Helzer *et al.* for lifetime prevalence of major depressive episode are extrapolated to the whole ECA group, the lifetime risk would rise from 4–7% to 10–17%. Helzer *et al.* suggest that greater diagnostic disagreement occurs in general population samples because of fewer symptoms, their lesser severity, and the fact that few cases come to treatment. It must be suspected that patients are more likely to remember an episode of depression and its features, not only because of its likely greater severity and duration, but because treatment (out-patient or in-patient) is an event that is unlikely to be forgotten, while conversations about symptoms with a physician 'may serve as rehearsal for responses to a survey instrument'. For such reasons, and because of their likely greater motivation to assist the interviewer, the testing of psychiatric patients will optimize reliability and validity examinations. Even if the DIS had been demonstrated to be a valid measure of psychiatric morbidity in the hands of lay interviewers in samples of psychiatric patients, such judgements about its validity could not be generalized to community samples without considerable risk. This point is worthy of re-phrasing: validity of measures can rarely be taken as demonstrated and must be assessed or reassessed, as is occurring now in the ECA study (Anthony *et al.* 1985; Helzer *et al.* 1985) in the target group or population.

As suggested above, the validity of the DIS as an accurate measure of lifetime psychiatric morbidity remains to be established. There has been no equivalent to the innovative study by Leckman *et al.* (1982) of the SADS-L which derived a criterion measure of 'best estimate' lifetime diagnosis and appeared to support strongly the validity of that measure. However, there was a large number of patients in that heterogeneous sample, which may have optimized results for the reasons noted above. As the authors did not report data for the sub-group of 'not mentally ill' subjects, who might correspond more closely with community subjects, we can only be suspicious that the overall validity findings were favoured by the sub-group of patients. Validity studies of the DIS will need to consider that possible limitation.

It is important to add a further challenge to the concept of 'procedural validity' (Spitzer &

https://doi.org/10.1017/S0033291700024806 Published online by Cambridge University Press

Williams, 1980), whereby data generated by the DIS are compared with data generated by another interview schedule or with DIS-data generated by another interviewer, and conclusions are drawn about the 'validity' of the DIS. Using such a strategy within a group of subjects who have experienced a major depressive episode it is quite possible that the same members might fail to remember or wish not to report any depressive episode, so that the consistency or reliability of the information may be perfect, and yet it will clearly be invalid. Validity estimates require quite different strategies, as considered shortly. While Robins *et al.* (1981) have admitted concerns about their 'validity' assessments, they argue that 'like many prior efforts to establish validity, our work is a bootstrapping operation'. In a recent paper, Robins (1985) has specified a number of the problems associated with attempting to validate the DIS, and the assembled list of problems is a comprehensive and weighty one. Robins states, that 'we wanted to know whether we could trust the prevalence figures the DIS produced'. Her claim that the goal of 'estimating population prevalences... can be successful even in the absence of strong evidence of validity' would appear over-optimistic when the ECA data are examined for lifetime diagnoses generated by the DIS.

In summary, the reviewed studies suggest that inter-rater reliability in establishing a current diagnosis of depressive illness is reasonably high in patient samples for major depressive episodes, providing that clinicians are used as raters, but that less impressive agreement has been determined for the diagnosis of minor or atypical depressive disorders. Reliability estimates appear weakened in non-clinical groups and when highly trained psychiatrists are compared with lay interviewers. The task of allocation to a specific diagnostic category has been shown to be somewhat more prone to error than merely attempting to determine whether a depressive episode is present or not. Nevertheless, 6-month prevalence data for depressive disorders derived by lay interviewers in the ECA study appear consistent with previous studies using case-finding techniques. For lifetime diagnosis of depressive episode there has only been one estimate (Leckman *et al.* 1982) of validity of a measure (the SADS-L) and no true estimate of the validity of the ECA measure, the DIS, so that the lifetime estimates derived in the ECA study appear more suspect than the 6-month estimates.

As only preliminary results of the ECA study have been published, it is appropriate to request clarification of some of these issues in the final reports. What might reasonably be requested? First, the properties of the DIS must be assessed within the target sample, as explored in the recent studies of Anthony *et al.* (1985) and Helzer *et al.* (1985). Secondly, test–retest reliability data for the two intervals (6-month and lifetime prevalence) should be assessed for the separate disorders as, in general terms, high reliability must be established before any estimate can be made of validity.

In this paper I have focused on depressive disorders and it would be unwise to generalize the observations to all disorders. It is conceivable that the DIS will be far more accurate in assessing other disorders (e.g. schizophrenia) where pathognomic features are more likely to be categorical than dimensional and when the salience of features may promote accurate and consistent recall by subjects.

If acceptable levels of test–retest reliability are established, be it for 6-month or lifetime prevalence data, then some estimate of the validity of the information is required. In this paper I have suggested that the ECA six-month prevalence data is comparable with findings from similar surveys. Such consistency, within and between the epidemiological surveys, supports but does not demonstrate validity. It may be that a significant percentage of the currently depressed deny symptoms, so that the possibility of response biases interfering with accurate measurement must be conceded. Again, the degree to which a DSM-III derived 'case' approximates to the real world of psychiatric morbidity, as against other systems attempting to define 'caseness', must be considered. How, then, might the validity of the DIS be best assessed? Corroborative reports, perhaps along the lines used by Leckman *et al.* (1982) for the SADS-L, could be useful. Reports by informants (e.g. relatives) would appear a particularly useful strategy, but some reservations must be entered. Informants' reports are not necessarily accurate, Platt (1980) noted that when informants' reports are compared with an independent and reliable external source of information, the response invalidity has been demonstrated to be as high as 42%. Platt also considered why low agreement does not necessarily mean poor validity and, conversely, why high across-interview agreement is not necessarily an

https://doi.org/10.1017/S0033291700024806 Published online by Cambridge University Press

adequate indicator of validity. He reviewed research noting that family members are not necessarily independent reporters, in that they may variably observe events as salient or not, that agreement may be made by consensus (particularly if the interviewee and informant have any opportunity to influence the other's responses), and that shared perceptions (as a consequence of family culture or other factors) may lead to similar patterns of over- or under-reporting, so producing spurious agreement. Thus, according to Platt, in certain circumstances agreement may not necessarily constitute validation if reports are not truly independent. But while corroborative reports from relatives or others have such potential limitations, many of those limitations may be pre-empted by design strategies, and such information is generally more readily obtained and more accurate than information obtained by reference to medical records. Lifetime prevalence data should also be examined against treatment data. If, for instance, it were to be established that 25% of those in a designated community sample have consulted their primary physician, 15% a non-medical therapist, and 10% a psychiatrist for depression over their lifetime, then a lifetime prevalence estimate of 2–5% for major depression generated by an interview schedule would require some consideration of a potential paradox.

This editorial is clearly sceptical in tone, raising doubts about the validity of the lifetime prevalence data generated in the early reports of the ECA project teams. Such doubts are expressed to reduce the chance of the data being regarded as firmly established, and health service needs, health finances and manpower needs then being predicated on the basis of the data published so far (Freedman, 1984). What is encouraging, if not inspiring about the ECA reports and the subsequent studies and commentaries, is the open debate about methodological issues by the involved researchers.

GORDON PARKER

Preparation of this paper was assisted by views of colleagues at the Mood Disordeis Unit, Prince Henry Hospital, Sydney, and Dr Robert Finlay-Jones.

REFERENCES

American Psychiatric Association. (1980). Committee on Nomenclature and Statistics: *Diagnostic and Statistical Manual of Mental Disorders*, Third edition. American Psychiatric Association: Washington, DC.

Andreasen, N. C., Grove, W. M., Shapiro, R. W., Keller, M. B., Hirschfeld, R. M. A. & McDonald Scott (1981). Reliability of lifetime diagnosis: a multicenter collaborative perspective. *Archives of General Psychiatry* **38**, 400–405.

Anthony, J. C., Folstein, M., Romanoski, A. J., Von Korff, M. R., Nestadt, G. R., Chahal, R., Merchant, A., Brown, C. H., Shapiro, S., Kramer, M. & Gruenberg, E. M. (1985). Comparison of the lay Diagnostic Interview Schedule and a standardized psychiatric diagnosis: experience in Eastern Baltimore. *Archives of General Psychiatry* **42**, 667–675.

Bebbington, P., Hurry, J., Tennant, C., Sturt, E. and Wing, J. K. (1981). Epidemiology of mental disorders in Camberwell. *Psychological Medicine* **11**, 561–579.

Boyd, J. H. & Weissman, M. M. (1981). Epidemiology of affective disorders: a reexamination and future directions. *Archives of General Psychiatry* **38**, 1039–1046.

Bromet, E. J., Dunn, L. O., Connell, M. M., Dew, M. A. & Schulberg, H. C. (1986). Long-term reliability of diagnosing lifetime major depression in a community sample. *Archives of General Psychiatry* **43**, 435–440.

Eaton, W. W., Holzer, C. E., Von Korff, M., Anthony, J. C., Helzer, J. E., George, L., Burnam, A., Boyd, J. H., Kessler, L. G. & Locke, B. Z. (1984). The design of the Epidemiologic Catchment Area surveys: the control and measurement of error. *Archives of General Psychiatry* **41**, 942–948.

Freedman, D. X. (1984). Psychiatric epidemiology counts. *Archives of General Psychiatry* **41**, 931–933.

Helzer, J. E., Robins, L. N., McEvoy, L. T., Spitznagel, E. L., Stoltzman, R. K., Farmer, A & Brockington, I. F. (1985). A comparison of clinical and Diagnostic Interview Schedule diagnoses: physician reexamination of lay-interviewed cases in the general population. *Archives of General Psychiatry* **42**, 657–666.

Klerman, G. (1985). Diagnosis of psychiatric disorders in epidemiologic field studies. *Archives of General Psychiatry* **42**, 723–724.

Leckman, J. F., Sholomskas, D., Thompson, D., Belanger, A. & Weissman, M. M. (1982). Best estimate of lifetime psychiatric diagnosis: a methodological study. *Archives of General Psychiatry* **39**, 879–883.

Mazure, C. & Gershon, E. S. (1979). Blindness and reliability in lifetime psychiatric diagnosis. *Archives of General Psychiatry* **36**, 521–525.

Myers, J. K., Weissman, M. M., Tischler, G. L., Holzer, C. E., Leaf, P. J., Orvaschel, H., Anthony, J. C., Boyd, J. H., Burke, J. D., Kramer, M. & Stoltzman, R (1984). Six-month prevalence of psychiatric disorders in three communities: 1980–1982. *Archives of General Psychiatry* **41**, 959–967.

Murphy, J. M., Sobol, A. M., Neff, R. K., Olivier, D. C. & Leighton, A. H. (1984). Stability of prevalence: depression and anxiety disorders. *Archives of General Psychiatry* **41**, 990–997.

Platt, S. (1980). On establishing the validity of 'objective' data: can we rely on cross-interview agreement? *Psychological Medicine* **10**, 573–581.

Regier, D. A., Myers, J. K., Dramer, M., Robins, L. H., Blazer, D. G., Hough, R. L., Eaton, W. W. & Locke, B. Z. (1984). The NIMH Epidemiologic Catchment Area program: historical context, major objectives, and study population characteristics. *Archives of General Psychiatry* **41**, 934–941.

Reich, T., Rice, J., Andreasen, N. & Clayton, P. (1980). A preliminary analysis of the segregation distribution of primary major depressive disorder. *Psychological Bulletin* **16**, 34–36.

https://doi.org/10.1017/S0033291700024806 Published online by Cambridge University Press

Robins, L. N. (1985). Epidemiology: reflections on testing the validity of psychiatric interviews. *Archives of General Psychiatry* **42**, 918–924.

Robins, L. N., Helzer, J. E., Croughan, J. & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. *Archives of General Psychiatry* **38**, 38–389.

Robins, L. N., Helzer, J. E., Ratcliff, K. S. & Seyfried, W. (1982). Validity of the Diagnostic Interview Schedule, version II. DSM-III diagnoses. *Psychological Medicine* **12**, 855–870.

Robıns, L. N., Helzer, J. E., Weissman, M. M., Orvaschel, H., Gruenberg, G., Burke, J. D. & Regier, D A (1984). Lifetime prevalence of specific psychıatric disorders in three sites. *Archives of General Psychiatry* **41**, 949–958.

Spitzer, R. L. & Williams, J. B. W. (1980). Procedural validity: validity of the diagnostic process. In *Comprehensive Textbook of Psychiatry, Third edition* (ed. H. I. Kaplan, A. M. Freedman and B. J. Sadock), Vol. 1, pp. 1039–1040. Williams & Wilkins Co: Baltimore.

Weissman, M. M. & Myers, J. K. (1978). Affective disorders in a US urban community. The use of Research Diagnostic Criteria in an epidemiological survey. *Archives of General Psychiatry* **35**, 1304–1311.

Weissman, M. M., Myers, J. K. & Harding, P. S (1978). Psychiatric disorders in a US urban community: 1975–1976. *American Journal of Psychiatry* **135**, 459–462.

https://doi.org/10.1017/S0033291700024806 Published online by Cambridge University Press