

CORRESPONDENCE

Psychological Medicine, **35** (2005).
doi:10.1017/S0033291705005180

To the Editor:

In their comparison of the CIS-R and CIDI lay diagnostic interviews for anxiety and depressive disorders with reference to the clinician-administered SCAN in 105 attendees at one primary-care practice Jordanova and colleagues (2004) devote almost all of their discussion to comments on what, in our opinion, is our quite different comparison of these instruments in 205 geographically sampled community subjects (Brugha *et al.* 1999*a*, 2001); they argue that we may have been unduly pessimistic in our view that neither lay diagnostic interview was adequately valid for use in population-based research because the design of our study may have led to an underestimate of the true validity of the two lay survey measures. A particular design difference for which our study is criticized is that we erred by excluding 467 householders with low scores on the CIS-R (total scores from 0 to 7; Brugha *et al.* 1999*a*), a point also noted earlier by Regier (2000). Using their data restricted to patients with CIS-R scores of 8 or above the CIS-R was shown to have much poorer agreement with SCAN for 'any ICD-10 diagnosis'. They acknowledge that both comparison studies were designed to increase study efficiency, in their study by interviewing primary-care attendees and in ours by excluding low community scorers (CIS-R scores of 0–7); they suggest that both studies are, therefore, limited in their generalizability to the general population and that better studies need to be designed and carried out that include random samples of low and high scorers on other screening questionnaires. We concluded that bias due to the absence of data on low scorers in our study was almost certainly small, having carefully conducted sensitivity analyses and considered the literature on community comparisons almost all of which concurred with our

findings (Brugha *et al.* 1999*a, b*). We note that their criticism was conditional: 'for both the CIDI and the CIS-R, kappas for diagnoses and syndromes were generally similar for the whole sample and the subsample of 26 persons scoring 8 or over on the CIS-R psychiatric morbidity scale. However, where the level of agreement for the whole sample was in the moderate to good range, and the disorder was relatively prevalent, the kappa was dramatically lower for the subgroup of those scoring 8 or over on the CIS-R.' Thus, their criticism applied to one of many sets of analysis (the difference reported not being statistically significant). We further note that important information on cell frequencies, sensitivity and specificity was not reported for these subgroup analyses.

However, we would agree that more general population data ought to be collected and analysed across the full range of mental illness probability in order to further reduce uncertainty; such an opportunity has since arisen from the Second National Household Survey carried out in collaboration with ONS in 2000 (Singleton *et al.* 2001) in which CIS-R and SCAN data were collected on subsamples stratified with respect to severity of mental disorder (Taub *et al.* 2005). These community data do permit us to compare agreement within one study in low and high scorers on screening tests in subjects randomly sampled throughout Great Britain. Selection for a clinician interview was based on screening questions for psychosis and for personality disorder (Singleton *et al.* 2001). Of 612 pairs of community interviews 250 were collected on randomly sampled adults who screened negative and 362 were on screen positives on one or both screening instruments. Chance corrected agreement (Cohen, 1968) for the presence or absence of any anxiety or depressive disorder comparing SCAN and the CIS-R in the whole sample was 0.39 (95% CI 0.29–0.49); it was 0.31 (95% CI 0.04–0.58) in respondents who screened negative for psychosis and personality disorder and ranged from

0.09 (95% CI -0.19 to 0.37) to 0.53 (95% CI 0.30 – 0.75) in the five other screen positive strata (Taub *et al.* 2005). Interestingly the highest level of agreement (0.53) was in the stratum screening negative on personality disorder, but positively on psychosis, in other words in those most severely mentally ill and most likely to have contact with health services. Our view has always been that agreement between lay and clinician assessments is generally noticeably better in subjects presenting at and using health-care services (as in the Jordanova study) and is poorer in randomly sampled community subjects (Brugha *et al.* 1999*b*); all three studies discussed in this letter confirm that conclusion. Although this new larger study comparing the CIS-R and SCAN may be limited by a much longer interval of 3–6 months between lay and clinician interviews and by non-random interview ordering, the findings do not differ importantly from those reported in our original comparison study (Brugha *et al.* 1999*a*) in which we also found no difference in concordance comparing intervals between interviews of 1–7 days with those of 8–21 days. These new findings do not support the argument by Jordanova and colleagues (2004) in their small study that exclusion of screen negatives in our earlier research has led to biased underestimates of concordance.

Jordanova and colleagues (2004) also conclude that both comparison studies show the CIDI performing better than the CIS-R (each with reference to SCAN) in spite of the finding in our comparisons of no statistically significant difference in agreement with SCAN for ‘any ICD-10 study diagnosis’ [CISR comparison, $\kappa = 0.25$ (0.10 – 0.40); CIDI comparison, $\kappa = 0.43$ (0.29 – 0.57)]. But they appear to have overlooked the non-comparable time periods used in the CIDI comparisons in the two studies: the past month in our general population study, and apparently the past year in their primary-care attender study (they omit to explain what time period or periods were assessed using SCAN). This is a second substantive limitation on comparisons of the two projects. It is also worth noting that when clinical and lay diagnostic interviews are compared for longer time periods in community subjects concordance is better (Kessler *et al.* 1998), which suggests that their own comparison of the

CIS-R (covering the past month) was bound to generate different findings to the 1-year period covered by the CIDI in their study. A third design difference that they discuss deserves highlighting: their study patients consulted with their general practitioner and immediately afterwards completed the three diagnostic interviews; in our study randomly sampled householders completed two or three diagnostic interviews over a period of one to several weeks at a time when they would have been substantially less likely to be considering discussing their health with a doctor. We argue that their design is more likely to artificially boost consistency in replies to questions repeated in contrast to ours, which mimics far more closely the circumstances found in epidemiological field studies and in which we have found no effect on agreement of variation in the interval between interviews.

The CIDI family of diagnostic interviews and the lay version of the Clinical Interview Schedule (CIS-R) and their associated diagnostic algorithms were primarily designed to conduct epidemiological and public mental health studies in the general population; it is clear from the wealth of evidence collected in such samples that lay and clinician evaluations generate differing findings. This should not be so surprising because lay diagnostic interviews and clinical interviews measure different things: subjectively perceived symptoms and clinically judged psychopathology. Nevertheless, as researchers we continue to value and to make considerable but cautious use of *both* kinds of measures, taking into account the increased knowledge and understanding provided by carefully reported research on the strengths and limitations of both methods and we would urge other users to consider the value of this policy.

Declaration of Interest

None.

References

- Brugha, T. S., Bebbington, P., Jenkins, R., Meltzer, H., Taub, N. A., Janas, M. & Vernon, J. (1999*a*). Cross validation of a household population survey diagnostic interview: a comparison of CIS-R with SCAN ICD-10 diagnostic categories. *Psychological Medicine* **29**, 1029–1042.

- Brugha, T. S., Bebbington, P. E. & Jenkins, R. (1999*b*). A difference that matters: comparisons of structured and semi-structured diagnostic interviews of adults in the general population. *Psychological Medicine* **29**, 1013–1020.
- Brugha, T. S., Jenkins, R., Taub, N. A., Meltzer, H. & Bebbington, P. (2001). A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychological Medicine* **31**, 1001–1013.
- Cohen, J. (1968). Weighted Kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* **70**, 213–220.
- Jordanova, V., Wickramasinghe, C., Gerada, C. & Prince, M. (2004). Validation of two survey diagnostic interviews among primary care attendees: a comparison of CIS-R and CIDI with SCAN ICD-10 diagnostic categories. *Psychological Medicine* **34**, 1013–1024.
- Kessler, R. C., Wittchen, H. U., Abelson, J. M., McGonagle, K. A., Schwarz, N., Kendler, K. S., Knauper, B. & Zhao, S. (1998). Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey (NCS). *International Journal of Methods in Psychiatric Research* **7**, 33–55.
- Regier, D. A. (2000). Community diagnosis counts [Comment]. *Archives of General Psychiatry* **57**, 223–224.
- Singleton, N., Bumpstead, R., O'Brien, M., Lee, A. & Meltzer, H. (2001). National Statistics. *Psychiatric Morbidity Among Adults Living in Private Households*, p. 154. The Stationery Office: London.
- Taub, N. A., Morgan, Z., Brugha, T. S., Lambert, P. C., Bebbington, P. E., Jenkins, R., Kessler, R. C., Zaslavsky, A. M. & Hotz, T. (2005). Recalibration methods to enhance information on prevalence rates from large mental health surveys. *International Journal of Methods in Psychiatric Research* **14**, 3–13.

T. S. BRUGHA, H. MELTZER, R. JENKINS,
P. E. BEBBINGTON AND N. A. TAUB

Address correspondence to :
Professor T. S. Brugha,
University of Leicester,
Department of Health Sciences,
Academic Psychiatry, Brandon Mental Health Unit,
Leicester General Hospital, Gwendolen Road,
Leicester LE5 4PW, UK.
(tsb@le.ac.uk)

