Assessing psychiatric disorder with a human interviewer or a computer

Glyn Lewis

Abstract

Objective **– To compare a self administered computerised assessment of neurotic psychiatric disorder (psychiatric morbidity) with an identical assessment administered by a human interviewer. In particular, to discover whether a computerised assessment overestimates or underestimates the prevalence of psychiatric morbidity in relation to a human interviewer.**

Setting **– A health centre in south east London, UK.**

Subjects **– A non-consecutive series of health centre attenders. Complete data were available on 92 subjects.**

Design **– All subjects received both assessments on the same occasion but were randomised to receive either the computerised assessment first or the human interview first.**

Results **– The mean total score on the assessment was the same for both methods of administration; computer 8·77 *v* human 8·69 (95% confidence interval for difference −0·70, 0·87). The correlation between the human and interviewer assessments was 0·91.**

Conclusion **– Self administered computerised assessments are valid, unbiassed measures of psychiatric morbidity. In addition to their use as a research tool, they have potential uses in primary care including screening for psychiatric morbidity and in forming the basis for clinical guidelines.**

(*J Epidemiol Community Health* 1994;48:207–210)

Institute of Psychiatry, De Crespigny Park, London SE5 8AF and London School of Hygiene and Tropical Medicine, London WC1E 7HT
G Lewis

Correspondence to:
Dr G Lewis, Institute of Psychiatry.

Accepted for publication July 1993

British general practitioners are becoming increasingly familiar with computerised methods of assisting practice administration and recording the details of consultations. There is less awareness that computers provide an opportunity to extract information directly from patients by means of self administered questionnaires. Several self administered computerised assessments of psychiatric morbidity[1] have been developed, including one designed in the UK for use in primary care settings[2] which concentrates on rapidly assessing the common neurotic disorders of depression and anxiety (called psychiatric morbidity in this report). Many of the proponents of computerised assessments have been discouraged by the reluctance of clinicians to use such information technology in their work. Among the reasons for this, however, must be included the paucity of data on the validity of computerised assessments of psychiatric morbidity and lack of evidence supporting their clinical efficacy.[3]

There have been reports that in computerised assessments of alcohol consumption respondents admit to consuming larger quantities than they do in similar assessments administered by a human interviewer.[4] Evidence from alcohol purchases is used to suggest that the larger figure is also more accurate.[5] Though other researchers have not found this phenomenon,[6 7] it raises the issue of whether other questions about socially undesirable characteristics, perhaps including those about mental health, will be relatively over-reported when part of a computer administered questionnaire rather than a questionnaire administered by a human. Greist *et al*[8] used identical versions of the Diagnostic Interview Schedule,[9] administcred either by interviewer or computer, in a sample of psychiatric patients and did not find any bias in the ascertainment of psychiatric diagnoses. Previous studies in Britain have shown that a computerised assessment of psychiatric morbidity showed good agreement with a standardised psychiatric interview administered by a psychiatrist.[2] However, in that study the computerised and human assessments contained different wording and so it was not possible to investigate any possible bias in identifying psychiatric morbidity.

This study was designed to assess the size of any potential bias between an identical assessment of psychiatric morbidity, the revised Clinical Interview Schedule[10] (CIS-R), when it was administered either by a computer or a human interviewer.

Methods

A non-consecutive series of subjects who attended a health centre in Bermondsey, south east London were invited to take part. Subjects were selected by receptionist staff if they judged the person would have to wait for some time before seeing the doctor. They were given the computerised assessment and the human interview in random order in a quiet and private room in the health centre and were also asked to complete, by themselves, the "paper and pencil", 12 item General Health Questionnaire[11] (GHQ) and the Hospital Anxiety and Depression Scale.[12] Those who scored 2 or more on the GHQ were described as being above the threshold. The interviewer administered a short sociodemographic questionnaire that included items on sex, age, and social class, classified according to the Goldthorpe and Hope[13] criteria and then divided into

Table 1 The level of agreement for individual symptoms between the human (H) and computer (C) interview (n=92).

Correlation	Weighted kappa (SEM)	Interview		Score >2 (%)
		H	C	
Somatic	0·49 (0·03)	18	17	0·61
Fatigue	0·63 (0·04)	41	41	0·74
Concentration	0·72 (0·02)	11	16	0·87
Depression	0·65 (0·02)	16	18	0·81
Irritability	0·72 (0·03)	30	30	0·83
Sleep	0·60 (0·05)	53	37	0·65
Worry over physical health	0·64 (0·02)	13	16	0·75
Depressive ideas	0·70 (0·02)	16	20	0·85
Worry	0·66 (0·03)	14	20	0·75
Anxiety	0·53 (0·03)	21	14	0·71
Phobia	0·48 (0·02)	7	10	0·62
Panic	0·69 (0·07)	3	4	0·80
Compulsions	0·75 (0·01)	9	10	0·88
Obsessions	0·59 (0·02)	10	16	0·69

manual or non-manual categories. Subjects were asked about previous treatment for mental health problems and were told that the results would not be given to their GP.

The computerised questionnaire was administered using the program PROQSY[2] (PROgrammable Questionnaire SYstem) on a portable computer with a conventional keyboard. The interviewer was a nurse with psychiatric training who had been given one hour's tuition on using the CIS-R and had been observed by the author during two interviews in which the CIS-R had been used. The CIS-R, in both interviewer administered and computerised version, is divided into 14 sections (see table 1) each scoring between 0 and 4 (except the section on depressive ideas in which the score is up to 5). The total score is calculated by summing the scores of each section. A total score of 12 or more is used to define a "case".[10]

The main analysis compared the mean total scores for the interviewer and computer administered CIS-R. After subjects had been divided into cases and non-cases, the agreement between the two methods of assessment was measured using the kappa[14] statistic and the index of agreement on positives.[15] Each of the 14 sections of the CIS-R was then examined alone. The prevalence of scoring more than 2 on each section was compared for both methods of assessment and the possibility of bias was investigated by McNemar's test.[16] The weighted kappa calculated for each section of the CIS-R used quadratic weights.

The reliabilities calculated by confirmatory factor analysis employed the FACTOR procedure of SAS.[17] In essence, all four measures, the computerised and interviewer CIS-R and the two self completed questionnaires, were all assumed to be measuring the same contruct derived by maximum likelihood factor analysis from the scores on the four measures. The communalities of each of the measures can then be interpreted as a measure of the reliability of the individual assessments of psychiatric morbidity in measuring the factor derived contruct. Further details of the method can be found in Lewis *et al*[10] and Dunn.[18]

Results

Ninety seven subjects agreed to take part in the study and complete data was available for 92 subjects. The mean (SD) age of the subjects was 40 (17·7) years, and 81% were women. Twenty per cent were either divorced, separated, or widowed and 48% were in a manual social class. Altogether 12·5% were born outside the UK and 19% had previously consulted a doctor about a mental health problem.

There were no statistically significant differences in the characteristics of those who received the computerised assessment first and those who received it second (table 2). In particular, there was no difference in the proportion who scored above the threshold on the GHQ. The mean total scores on the CIS-R for the first and second assessments were compared and there was no evidence of any significant differences in scores. The order of presentation did not therefore have any influence on total scores.

The mean total score on the computerised assessment was 8·77 (95% confidence interval (CI) 6·89, 10·66) and on the interviewer assessment it was 8·69 (95% CI 6·86, 10·52) (paired *t* test; t=0·2, df=91, p=0·8; 95% CI for difference −0·70, 0·87). The correlation between the human and interviewer assessments was 0·91. Subjects were also divided into cases and non-cases on the basis of CIS-R scores (table 3). The index of agreement on positives was 0·66 and the kappa was 0·70 (SD 0·08). Agreement on the individual sections of the CIS-R was also examined (table 1). The mean kappa value across the sections of the CIS-R was 0·63. The possibility of bias within the sections was examined using McNemar's test. In only one of the 14 sections was there any indication of a statistically significant bias, the sleep section (χ^2=13·2; df=1; p<0·001), in which the computerised assessment resulted in higher scores.

The reliabilities of the computerised and human CIS-R were estimated using confirmatory factor analysis. The results in table 4 indicate that both measures were more reliable than the questionnaires and had reliabilities of around 0·90. Though computer administration had a higher reliability than the administration by a human interviewer, it is impossible, for technical reasons, to estimate the statistical significance of this difference.

Table 2 Comparison of respondents who were allocated to human and computerised administration for the first assessment.

Characteristic	Computer first	Computer second	Significance test
Female (%) (95% CI)	81·8 (73·9, 89·6)	81·8 (73·9, 89·6)	χ^2=0; df=1; p=1·0
Divorced, separated, or widowed (%) (95% CI)	22·4 (13·9, 30·9)	17·0 (9·3, 24·7)	χ^2=0·44; df=1; p=1·0
Born outside UK (%) (95% CI)	12·2 (5·9, 18·5)	12·8 (6·0, 19·6)	χ^2=0·006; df=1; p=0·94
In manual occupations (%) (95% CI)	55·1 (44·9, 65·3)	40·4 (30·4, 50·4)	χ^2=2·0; df=1; p=0·15
Previously treated for mental health problem (%) (95% CI)	18·4 (10·5, 26·3)	19·2 (11·2, 27·2)	χ^2=0·02; df=1; p=1·0
Mean (SD) age	39·9 (18·0)	40·2 (17·6)	t=0·09; df=91; p=0·93
Above GHQ threshold (%) (95% CI)	45·8 (35·6, 56·0)	42·6 (32·5, 52·7)	χ^2=0·1; df=1; p=0·75

Table 3 Agreement on case definition according to method of interviewing.

Human	Computer	
	Non-case	Case
Non-case	54	6
Case	5	21

Kappa = 0·70 (SD 0·084).

Discussion

No differences in the ascertainment of psychiatric morbidity were observed when an identical questionnaire was administered either by a human interviewer or by a computer. Though studies enquiring about alcohol intake have led to suggestions that people are more likely to divulge sensitive information to a computer than to another person, this effect was not seen here in assessing psychiatric morbidity. This is consistent with Greist's[8] findings in North America using a different assessment and investigating psychiatric patients rather than primary care attenders. Only one of the 14 sections of the CIS-R showed any evidence of bias, the section on sleep, and it is possible that this was due to chance. Overall, the computerised assessment gave very similar estimates to the human administered assessment. The level of agreement observed between the human and computerised assessments was also similar in magnitude to the results of the study of Lewis *et al*[10] of the agreement between two interviewers administering the CIS-R. This suggests that self administered computerised assessments of psychiatric morbidity, such as the one used here, are as valid as interviewer administered measures in general practice and community settings. It is important to emphasise that the assessment used was designed only to assess neurotic disorders and is therefore suitable for use in primary care and other settings where psychotic disorders are relatively uncommon. The conclusion concerning the validity of computerised assessments therefore applies only to these circumstances.

Only a single interviewer was used here, who was also a trained psychiatric nurse. It is possible that other interviewers, perhaps with a less sympathetic demeanour, may lead to a failure to disclose information about mental health. Indeed one of the attractions of using computerised assessments is the lack of observer bias and the consistency of the assessment in different situations. Even well trained interviewers will inevitably vary in interviewing styles. This might be reflected in the slightly higher reliability observed for the computerised assessment (table 4).

The sample of subjects was not representative of primary care attenders but they were randomised to the two groups after agreeing to take part. There is no empirical evidence to suggest that the unrepresentative sample would affect the level of agreement reported in this study. It is possible though that there are other situations which may affect the readiness of people to divulge details of their mental state. For example, subjects in a community survey, recruited outside the health service, may possibly be more reluctant.

Table 4 Reliabilities of the different assessments using a factor analysis measurement model.

Assessment	Score (mean (SD))	Reliability
CIS-R computer	8·7 (8·8)	0·96
CIS-R human	8·8 (9·1)	0·86
GHQ 12	12·3 (6·4)	0·61
HAD	9·9 (6·6)	0·48

CIS-R = Clinical Interval Schedule (revised); GHQ = General Health Questionnaire; HAD = Hospital Anxiety and Depression Scale.

Self administered computerised assessments for neurotic psychiatric disorders seem as valid and reliable as human assessments. It has already been observed that lay interviewers using the CIS-R show good agreement with psychiatrically trained interviewers.[10] For research, such computerised assessments have considerable advantages. They eliminate observer bias and reduce interviewer costs and the time involved in coding and data entry. The advantages for clinical use are fewer but these assessments might have potential value in primary care. Two such applications are of immediate interest.

Firstly, there is now considerable evidence that general practitioners do not identify a substantial proportion of attenders with a psychiatric disorder.[19] Though training in interviewing skills can improve this,[20] there is still a need to screen for undetected psychiatric morbidity and computerised assessments could perform this role. Screening for suicidal thoughts and intention may be particularly relevant in the light of the British Government's recent health strategy.[21]

A second potential use is in assisting the primary health care team to manage patients with psychiatric morbidity. The development of clinical guidelines for the management of neurotic disorder is still in its infancy and this may partly result from the absence of routinely used standardised assessments in primary care. However, computers could produce such a standardised assessment which might provide a basis for guidelines on the management and referral of patients with psychiatric morbidity. For example, determining the severity of depression above which antidepressant medication is indicated, could usefully be linked to a standardised computer assessment which could then be available for general practitioners. Computerised assessments could also be of value in providing general practitioners with additional information about the mental health of a patient. Clinicians could ask patients to return for a second consultation, perhaps focussed on mental health, after they had completed a computerised assessment. Such an approach might save time and allow the consultation to concentrate on the important issues flagged by the computerised assessment.

Information technology is very fashionable just now. Despite these encouraging results it is important that developments in computerised assessments are properly evaluated and that their use does not jeopardise the relationship between doctor and patient. Such relationships influence the non-specific aspects of

treatment as well as patient satisfaction and compliance. Computers cannot replace the clinical acumen of general practitioners but may prove to be a useful adjunct to the treatment of psychiatric morbidity by the primary care team.

This study and GL were funded by the Department of Health. My thanks to Gill Todd for carrying out the assessments and to Dr Richard Donmall for providing facilities and access to patients in the Albion Street Health Centre. I am grateful to Professor Anthony Mann for comments on an earlier draft.

1 Hedlund JL, Vieweg BW, Cho DW. Mental health computing in the 1980s: II. Clinical applications. *Computers in Human Services* 1985;**1**:1–31.

2 Lewis G, Pelosi AJ, Glover E, *et al*. The development of a computerized assessment for minor psychiatric disorder. *Psychol Med* 1988;**18**:737–45.

3 Pelosi AJ, Lewis G. The computer will see you now. *BMJ* 1989;**299**:138–9.

4 Waterton JJ, Duffy JC. A comparison of computer interviewing techniques and traditional methods in the collection of self-report alcohol consumption data in a field survey. *International Statistical Review* 1984;**52**:173–82.

5 Duffy JC, Waterton JJ. Under reporting of alcohol consumption in sample surveys: the effect of computer interviewing in fieldwork. *Br J Addict* 1984;**79**:303–8.

6 Bernadt MW, Daniels OJ, Blizard RA, Murray RM. Can a computer reliably elicit an alcohol history? *Br J Addict* 1989;**84**:405–11.

7 Skinner HA, Allen BA. Does the computer make a difference? Computerised versus face-to-face versus self-report assessment of alcohol, drug and tobacco use. *J Consult Clin Psychol* 1983;**51**:267–75.

8 Greist JH, Klein MH, Erdman HP, Bires JK, Bass SM, Machtinger PE, Kresge DG. Comparison of computer- and interviewer-administered versions of the Diagnostic Interview Schedule. *Hosp Community Psych* 1987; **38**:1304–11.

9 Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule; its history, characteristics and validity. *Arch Gen Psychiatry* 1981;**38**:381–9.

10 Lewis G, Pelosi AJ, Araya R, Dunn G. Measuring psychiatric disorder in the community: a standardised assessment for lay interviewers. *Psychol Med* 1992;**22**:465–86.

11 Goldberg DP. *The detection of psychiatric illness by questionnaire*. London: Oxford University Press, 1972.

12 Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiat Scand* 1983;**67**:361–70.

13 Goldthorpe J, Hope K. *The social grading of occupations*. Oxford: Oxford University Press, 1974.

14 Brennan P, Silman A. Statistical methods of assessing observer variability in clinical measures. *BMJ* 1992;**304**:1491–4.

15 Wing JK, Nixon JM, Mann SA, Leff JP. Reliability of the PSE (ninth edition) used in a population study. *Psychol Med* 1977;**7**:505–16.

16 Armitage P, Berry G. *Statistical methods in medical research*. 2nd ed. Oxford: Blackwell, 1987.

17 SAS Institute Inc. *SAS user's guide: basics*. Version 5. Cary, NC: SAS Institute Inc, 1985.

18 Dunn G. *The design and analysis of reliability studies*. Oxford: Edward Arnold, 1989.

19 Goldberg D, Huxley P. *Mental illness in the community*. London: Tavistock, 1980.

20 Gask L, Goldberg D, Lesser AL, *et al*. Improving the psychiatric skills of the general practice trainee: an evaluation of a group training course. *Med Educ* 1988;**22**:132–8.

21 Secretary of State for Health. *The health of the nation*. London: HMSO, 1992:126. Cm 1986.