Brit. J. Psychiat. (1983), **142**, 66–70

Accuracy of Behavioural Assessment by Computer

ANTHONY C. CARR and ANINDYA GHOSH

Summary: An automated assessment interview was given by a microcomputer to 26 randomly selected patients, referred for treatment of phobias. The results were compared with those of conventional clinical assessment by experienced behaviour therapists. Ratings of overall severity and intensity of specific types of agoraphobia and social phobia were derived from the computer interview, and correlated very closely with global ratings by the clinician and also with an independent structured clinical assessment. The computer also elicited behavioural targets to serve as a basis for exposure treatment. Blind assessors rated these targets as highly as those arrived at by clinicians in respect of their practicability, precision and appropriateness for treatment. The automated technique is inexpensive, saves clinicians' time and can be made widely available for screening, assessment and progress monitoring. It may also provide a basis for automated exposure treatment.

It is estimated that over 13,000 adults in Britain suffer from phobias severe enough to handicap them markedly in daily activities (Agras *et al*, 1969). Without specialized treatment most will remain handicapped throughout life (Lewis, 1936). The great majority, if given behavioural treatment, could return to normal work and social life (Gelder *et al*, 1973; Marks *et al*, 1977). However, for most of these patients treatment is not available for two reasons.

One is that of identification. Although phobic symptoms are easy to elicit, diagnosis rests upon a careful discrimination between true phobias and phobic symptoms accompanying anxiety or depressive neuroses which must be managed differently. The distinction requires a lengthy interview, preferably with a clinician experienced in this field. Unfortunately, such expertise is located in specialist behavioural units in teaching centres, out of reach of many sufferers, particularly the many whose illness makes travel difficult. Some psychiatric questionnaires contain a section relating to phobias, but little is known of the accuracy and specificity of such scales.

A second problem is the shortage of staff experienced in behaviour therapy. In addition to identifying the phobic patient, management requires a two to three hour behavioural assessment to establish the types of disability present, the appropriateness of behaviour therapy and the definition of treatment goals suitable for him. This is followed by two to three months of exposure treatment, the success of which depends on the accuracy of the initial assessment. During treatment, further assessments are needed to monitor progress. Current manpower is quite insufficient to deal with the problem, even if the patients could be satisfactorily identified.

Marks *et al* (1977) have developed an itemized scheme to enable para-medical personnel to carry out behaviour therapy with considerable success, but resources are still inadequate. Attempts have been made to automate the treatment (Lang, 1970), but the need for expert assessment remains.

Automated psychiatric assessment at a computer console has been shown to assess symptoms accurately (Greist *et al*, 1973; Carr *et al*, 1981, 1982). Patients reply via Yes/No or multiple-choice buttons, and find the procedure acceptable (Lucas, 1977). If screening and behavioural assessment of phobic patients could be performed effectively by computer it could be made widely available, and would save skilled therapists' time enabling more patients to be treated.

We have designed an automated behavioural assessment for phobias which has proved highly acceptable to most patients (Carr and Ghosh, 1982). The accuracy of that technique is compared in this study with that of a written inventory, the Hopkins symptom checklist 90 (SCL90) (Derogatis, 1977); a standardized structured assessment, the fear questionnaire (FQ) (Marks and Mathews, 1979); and clinical ratings by an experienced therapist.

Method

Automated interview

A small computer was built and programmed by the authors to conduct interviews with patients. Questions

and comments phrased in conversational style were displayed on a television screen, and the patient typed his replies in English on a normal typewriter keyboard. The system has been fully described elsewhere (Carr *et al*, 1982). Each patient was left alone with the computer with little prior instruction: the machine itself explained how to proceed, and shaped his responses progressively until he has mastered the technique.

The automated assessment then commenced. First the patient was screened for other psychiatric conditions such as generalized anxiety, depression and drug abuse, and for severe physical disease likely to make behavioural treatment difficult or hazardous. Unsuitable patients would be rejected automatically at this stage, and referred back to the clinician. Otherwise the computer proceeded to ascertain the areas in which the patient's troubles occurred, for example, open spaces, crowds, heights, travel or animals, and explained the principles of behavioural assessment and treatment, broadly following the procedure outlined by Marks (1978). The patient typed in descriptions of his problems. As his understanding of behavioural principles grew, these problems were re-displayed to him at successive points during the interview so that he might redefine them in more specific and practical terms, until finally they constituted target behaviours forming a suitable basis for exposure treatment.

The computer also administered three questionnaires: the depression, anxiety and phobia sections of the symptom checklist 90; a computer depression inventory of known validity, the depression rating scale (Carr *et al*, 1981); and the fear questionnaire (FQ), which also contains a short section on anxiety and depression. This last is a well-validated instrument for detecting change in severity of phobias (Marks and Mathews, 1979), but in its standard form requires the presence of a clinician. Before completing the normal paper and pencil FQ, the patient receives advice and explanations from the therapist, who will read and discuss the completed form with him, on occasions allowing the patient to modify his ratings as a result of their discussion.

Procedure

A pilot study was performed on 11 patients, whose comments allowed us to refine and clarify the automated interview displays.

The main sample consisted of 26 consenting patients from a series of 32 consecutive referrals to our unit for treatment of phobias. Of these, four patients were considered unsuitable for behaviour therapy; three suffered from depression and one from schizophrenia. Of the 28 entering the trial, two failed to attend for the automated assessment interview: one had recovered while on our waiting list, and the other decided that the distance to our hospital was too great to permit regular attendance. All patients were referred directly from home by their family doctor for treatment of phobias. No normal control subjects were studied.

Each patient received a general psychiatric assessment from one of us, who rated him on various scales for phobia, anxiety, depression and motivation. He then sat through the automated assessment, and also received a standard behavioural assessment which included the FQ from an independent blind assessor, a behaviour therapist unaware that the patient was being assessed by computer. All three assessments were carried out within three weeks, taking a total of four hours.

To assess the quality of the treatment target behaviours elicited by the computer, a random selection of 12 targets typed by the sample patients were mixed with 12 targets defined in their blind assessment interviews. This list of 24 targets was rated by five other experienced therapists for treatment suitability, practical feasibility and preciseness of definition.

Results

The computerized fear questionnaire (CFQ) scores for total severity of phobia (all symptoms) correlated very highly with the corresponding clinician's total for each patient. Individual subsections dealing with different types of phobia also corresponded closely (Table I). In contrast, the SCL90 phobia scale showed poor correlation with the clinician's total, although it was still statistically significant. The computer estimates of depression (DRS and SCL90 depression scales) correlated less well with the clinician's estimates, though significantly. The SCL90 general anxiety scale agreed moderately with the clinician (Table I), although the patients' self-ratings were much higher than those of the clinician.

CFQ estimates of phobic severity correlated very closely indeed with those made at the blind assessment interview on the same questionnaire. The computer estimates tended to be slightly higher, but not significantly so (Table II). The computerized SCL90 estimates correlated fairly well with blind assessments of phobia.

Each target was rated by five experienced therapists unaware whether it had been derived from the computer or a blind assessor: their ratings are given in Table III. Computer and assessors' targets did not differ significantly on any of the qualities rated. Taking minimum acceptable ratings of 62 per cent (practicability and suitability) and 50 per cent (precision), three-quarters of the targets were acceptable for use in treatment. Of the six targets unacceptable by this

https://doi.org/10.1192/bjp.142.1.66 Published online by Cambridge University Press

TABLE I

Comparison of computer's and clinician's assessments

Symptom	Computer scale	Product-moment correlation (n = 26)	Significance of correlation
All phobias	FQ	r = +.88	P < .01
Agoraphobia	FQ	r = +.77	P < .01
Social phobia	FQ	r = +.84	P < .01
Specific phobias	Too few cases	—	—
All phobias	SCL 90	r = +.44	P < .05
Depression	DRS	r = +.47	P < .02
Depression	SCL 90	r = +.60	P < .01
Anxiety (generalized)	SCL 90	r = +.64	P < .01

TABLE II

Comparison of computer's and blind assessor's ratings

Symptom	Computer scale	Mean severity** (0 = absent, 100 = severe)		Product-moment correlation (n = 24)*	Significance of correlation
		Computer	Assessor		
All phobias	FQ	46	39	r = +.88	P < .01
Agoraphobia	FQ	53	44	r = +.88	P < .01
Social phobias	FQ	47	39	r = +.86	P < .01
Specific phobias	Too few cases	—	—	—	—
All phobias	SCL 90	45	39	r = +.69	P < .01
Anxiety	SCL 90	76	37	r = +.59	P < .01

* Two of the 26 patients failed to attend for blind assessment.
** None of the means differed significantly.

criterion, three had been elicited by the computer and three by the blind assessors.

Discussion

It is clear that the computerized fear questionnaire (CFQ) provides an accurate estimate of the severity of phobic symptoms, whether compared with the standard structured assessment or with an experienced clinician's global ratings, both performed in a specialist behaviour therapy unit. The CFQ gave accurate indications of the presence of the two main types of phobia, and may serve as an effective screening instrument.

Unfortunately we were not able to provide a control group for this study. Normal subjects would not form a useful control group. Although they would score zero on almost all items and be easily distinguished from phobics by the CFQ, this would give no information on its screening effectiveness in real life. The problem is to differentiate phobic neurosis from such conditions as depression, anxiety states and severe personality disorders, in which similar symptoms may occur. Patients entering our trial had already been screened for other major psychiatric illness before entering the trial, so that all our sample patients had a principal diagnosis of phobic neurosis. In view of this fact, the high correlations we obtained are particularly impressive; correlation coefficients are artificially increased by the inclusion of controls with disparate scores. As a rough guide to screening efficiency, we divided our sample into moderate/severe and mild phobias on the basis of the blind assessors' assessments, and again on the CFQ results, taking a cut-off of 25 per cent on each scale to represent phobias severe enough to warrant treatment. It was found that 22 patients were correctly classified by the CFQ (Table III).

It is interesting that patients recorded slightly

https://doi.org/10.1192/bjp.142.1.66 Published online by Cambridge University Press

TABLE III

Efficiency of computer assessment for screening phobic patients

Patients rated by blind assessors	Patients scoring on CFQ	
	Above 25%	Below 25%
Above 25%	15	0
Below 25%	2	7

Agreement highly significant, $P < .01$, Fischer's exact test.

TABLE IV

Quality of proposed target behaviours for treatment

Mean ratings by experienced therapists of:	Targets elicited by computer (%)	Targets defined by blind assessors (%)
Practical feasibility	88	85
Suitability for exposure treatment	80	80
Definition within set limits	56	63

All computer-assessor differences non-significant (t-test).

higher scores on the CFQ than when answering the same questions in the presence of the blind assessor. This was true on each sub-section also (see Table II), and may be more than a chance finding. Each patient completed the computer interview first. By the time the therapist asked him to complete the FQ, increasing understanding of phobias and their treatment may have encouraged the patient to re-evaluate his own problems more modestly. Alternatively, the immediacy of the clinician may have deterred some patients from rating their symptoms as highly as they did in the privacy of the computer interview, as occurs with alcoholics for example (Lucas *et al*, 1977).

The fear questionnaire and the earlier fear survey schedule were developed to measure changes in severity accurately, rather than absolute values. In this study, FQ ratings from the blind assessment interviews matched the clinician's global ratings of phobias ($r = +.76$ for all phobias, $+.65$ for agoraphobia and $+.85$ for social phobics), suggesting that the standard FQ also provides a useful absolute rating of severity.

Neither of the depression inventories correlated well with the clinician's ratings for depression, either in the main study or the pilot study. This is surprising in view of the known accuracy of such inventories; we obtained a correlation of $+.78$ in an earlier study of depressed patients (Carr *et al*, 1981). Several factors may account for this. All but three of our patients were rated as mildly or not depressed and this narrow range of scores would tend to lower the correlations observed. The clinician did not enquire exhaustively for depressive symptoms unless the interview suggested that depression was present, and may perhaps have missed minor instances. Or, patients anxious to concentrate on describing their phobic symptoms may have given abbreviated accounts of their depression in the clinical interviews. The hypothesis that inaccuracies arose in the clinical interview is supported by the close agreement between the two computer depression ratings for our patients (correlation between DRS and SCL90 depression ratings $= +.81$).

Alternatively, it may be difficult to assess depression accurately in phobic patients, often with obsessional personalities, by means of questionnaires. This would have serious implications in view of the importance of detecting underlying depression before administering behaviour therapy. The problem is not simply one of over-rating physical symptoms, as patients were rated equally depressed on average by the clinician and the DRS.

Generalized anxiety assessment by computer did not closely match clinical assessments, probably because of the difficulty of distinguishing anxiety from phobic symptoms in a questionnaire. Thus, patients' scores on the SCL90 anxiety scale were much higher than the clinician's assessments, suggesting that phobic symptoms were boosting the anxiety scores. There was an appreciable relation between phobic symptoms and anxiety ratings, and better results might be obtained using a questionnaire expressly designed to detect non-phobic anxiety.

The computer-derived behaviour targets were considered as suitable for treatment as those formulated by our blind assessors. This remarkable finding suggests that computers could perhaps play an extensive part in the treatment of phobics, as well as in their assessment. Our target rating scales were not a standardized instrument. However, the agreement between raters as to the relative merits of the targets, despite some variation in overall standards between raters, would have sufficed to detect any important differences in quality. For example, of the six worst targets at least four received low ratings from each rater.

Poorly defined targets tended to come from male patients with social phobias rather than agoraphobia, although the numbers are too small for statistical comparison. Age and severity of handicap appeared irrelevant, but the worst targets all came from the minority of our patients who had had no previous

treatment for phobias at all, even from their general practitioner. Medical interviews would appear partially successful in educating patients about this condition.

Automated assessment of phobias is therefore an accurate and acceptable technique for estimating the severity of the condition and deriving suitable target behaviours on which treatment may be based. It is a cheap, simple procedure which could easily be provided locally for patients unable to travel, and yields results comparable with those of experts in this field. Its usefulness as a screening test depends on the computer's ability to distinguish phobic neuroses from other conditions causing phobic symptoms. As an instrument to monitor a patient's progress it offers a considerable saving in clinician's time and provides an objective, quantitative record, leaving the way open for further automation of the management of phobias.

Acknowledgement

This work was supported by a Medical Research Council grant. We are grateful to Professor Isaac Marks and Dr John Greist for their support and encouragement.

References

AGRAS, W. S., SYLVESTER, D. & OLIVEAU, D. C. (1969) The epidemiology of common fears and phobias. *Comprehensive Psychiatry*, **10**, 151–6.

CARR, A. C., ANCILL, R. J., GHOSH, A. & MARGO, A. (1981) Direct assessment of depression by microcomputer: a feasibility study. *Acta Psychiatrica Scandinavica*, **64**, 415–22.

—— GHOSH, A. & ANCILL, R. J. (1983) Can a computer take a psychiatric history. *Psychological Medicine*, (in press).

—— —— (1983) Response of phobic patients to direct computer assessment. *British Journal of Psychiatry*, **142**, 60–5.

DEROGATIS, L. R. (1977) *The SCL-90 Manual I: Scoring, Administration and Procedures for the SCL-90*. Baltimore: Johns Hopkins University School of Medicine, Clinical Psychometrics Unit.

GELDER, M. G., BANCROFT, J. H. S., GATH, D. H., JOHNSTON, D. W., MATHEWS, A. M. & SHAW, P. M. (1973) Specific and non-specific factors in behaviour therapy. *British Journal of Psychiatry*, **123**, 445–62.

GREIST, J. H., GUSTAFSON, D. H., STAUSS, F. F., ROWSE, G. L., LAUGHREN, T. P. & CHILES, J. A. (1973) A computer interview for suicide risk prediction. *American Journal of Psychiatry*, **130** (12), 1327–31.

LANG, P. J., MELAMED, B. G. & HAST, J. A. (1970) A psychophysiological analysis of fear modification using an automated desensitization procedure. *Journal of Abnormal and Social Psychology*, **76**, 220–34.

LEWIS, A. (1936) The problem of obsessional illness. *Proceedings of the Royal Society of Medicine*, **29**, 325–6.

LUCAS, R. W. (1977) A study of patients' attitudes to computer interrogation. *International Journal of Man-Machine Studies*, **9**, 69–86.

—— MULLIN, P. J., LUNA, C. B. X. & MCINROY, D. C. (1977) Psychiatrists and a computer as interrogators of patients with alcohol-related illnesses: a comparison. *British Journal of Psychiatry*, **131**, 160–7.

MARKS, I. M. (1978) *Living with Fear*. New York: McGraw-Hill.

—— HALLAM, R. S., CONNOLLY, J. & PHILPOTT, R. (1977) Nursing in Behavioural Psychotherapy. London: *Royal College of Nursing Research Report*.

—— & MATHEWS, A. M. (1979) Brief standard self-rating for phobic patients. *Behaviour Research and Therapy*, **17**, 263–7.

Anthony C. Carr, M.Phil., M.R.C.P., M.R.C.Psych., *Senior Registrar and Lecturer*

*Anindya Ghosh, M.D., M.A.M.S., M.R.C.Psych., *Senior Registrar and Hon. Lecturer*
Institute of Psychiatry, 99 Denmark Hill, London SE5 8AF

* Reprints.

(*Received 14 August; revised 17 November 1981*)