The survey form of SCAN: the feasibility of using experienced lay survey interviewers to administer a semi-structured systematic clinical assessment of psychotic and non-psychotic disorders

T. S. BRUGHA, F. NIENHUIS, D. BAGCHI, J. SMITH and H. MELTZER

From the Section of Social and Epidemiological Psychiatry, University of Leicester; and Social Survey Division, Office for National Statistics, London

ABSTRACT

Background. The success of large scale surveys depends on well designed questionnaires and the skills of lay interviewers. Discrepancies in prevalence rates between epidemiological surveys and poor agreement between survey interviewer and clinician diagnostic interviews are giving rise to increasing concern among researchers, public health planners and policy developers. New approaches to information collection are called for. The feasibility of training experienced survey interviewers in semi-structured, clinical, diagnostic interviewing has never been investigated systematically across the range of neurotic and psychotic disorders.

Methods. Eight experienced survey interviewers from the Office for National Statistics (ONS) were selected and underwent extended training in a Survey Form of SCAN (SCAN-SF). Sixty-four adults, including a majority of psychiatric in-patients were assessed by ONS interviewers and re-interviewed within a week by SCAN-trained clinicians. Feedback was sought from interviewers and trainers.

Results. Trainers found lay interviewers coped at least as well with psychotic as with neurotic symptoms. Concordance for any disorder was 0.74 (95% CI: 0.57 to 0.91); for any specific psychotic disorder 0.63 (0.40 to 0.86); for any specific neurotic disorder 0.63 (0.43 to 0.83). Sensitivity ranged from 0.6 to 0.9 and specificity from 0.8 to 0.9. There was no evidence of rater bias.

Conclusions. These preliminary findings are very promising. However, before the SCAN-SF, administered by carefully trained lay interviewers, can be recommended in large scale surveys, further evaluations of its feasibility and reliability in the general population are needed.

INTRODUCTION

In recent years, the desire for valid information on psychiatric morbidity has been expressed by psychiatric researchers, public health planners and policy developers commissioning large scale epidemiological surveys. Large scale psychiatric morbidity surveys, until now, have relied on structured diagnostic interviews by lay survey interviewers (Wittchen, 1994) because of the cost and lack of availability of clinically experienced interviewers. However, comparisons of lay and clinician administered survey interviews in the general population consistently furnish discrepant findings suggesting that they yield different information (Brugha et al. 1999a; Wittchen et al. 1996). Furthermore, prevalence estimates derived from general population surveys conducted by lay interviewers, which use similar structured diagnostic interviews continue to furnish different disorder rates. This has considerable implications for determining treatment need in the context of finite healthcare resources (Regier et al. 1998). Among
researchers and those responsible to government for mental health policy and planning, there are growing concerns about how the most relevant information can be sought and can be obtained reliably from the general population.

**Structured diagnostic interview assessments**

A recent study comparing two structured diagnostic interviews, the CIS-R and the CIDI-Auto with the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) among adults living in private households (Brugha *et al.* 1997, 1999b) confirm the findings of researchers using earlier such interviews of poor agreement between lay and clinician systematic interviews of householders for depressive and other specific non-psychotic disorders (Anthony *et al.* 1985; Helzer *et al.* 1985; McLeod *et al.* 1990). In these evaluations a common set of classification rules is used for both approaches so that any differences found must be due mainly to differences in the nature of the interview assessment method and the information it gathers. These findings beg the question whether more clinically-valid information on neurotic disorders could be provided by increased use of semi-structured methods (such as the SCAN) in future general population surveys.

In the general population, non-clinically trained interviewers have carried out structured interviews to identify psychotic disorders (Robins & Regier, 1991; Kendler *et al.* 1996) and psychotic symptoms (Jenkins *et al.* 1997). If anything, even greater difficulties have been encountered with the identification of psychotic rather than neurotic disorders. Most recently, the diagnostic validity of a structured assessment of non-affective psychosis by non-clinically trained interviewers was assessed by clinician re-interviews and it was found that the diagnoses so obtained agreed very poorly with clinical diagnoses (Kendler *et al.* 1996).

The feasibility of using lay interviewers to assess psychotic symptoms by means of a systematic, semi-structured interview, has never, to our knowledge, been systematically evaluated.

**Using lay survey interviewers to administer a semi-structured clinical psychiatric interview for neurotic symptoms**

Until now the authors of the Present State Examination (PSE) have accepted the feasibility of employing lay interviewers to assess non-psychotic symptoms but have advised against the use of clinically inexperienced interviewers to assess psychotic symptoms (Wing, 1980a, b). Our accumulated experience of teaching the PSE and the SCAN over two decades provides ample, albeit anecdotal, evidence, that the reliable assessment of depression may be at least as difficult as for psychotic symptoms.

We are aware that semi-structured interviews have been used by non-clinicians in quite large scale community surveys of children and adults (Costello *et al.* 1996). Two earlier reports suggest that lay interviewers may be able to master the neurosis components of the Present State Examination in adult community surveys (Rodgers & Mann, 1986; Sturt *et al.* 1981). Both investigators found evidence that lay interviewers may rate symptoms at a lower threshold. Sturt and her colleagues found that two out of eight agency (lay) interviewers ‘had thresholds that were lower than those of their own clinical interviewers for common symptoms including the key symptom of depressed mood’. Similarly, Rodgers & Mann, when they re-examined 526 tape recorded PSE interviews by nurses employed to study a community longitudinal cohort, found 6.7% were allocated as cases by nurses and 4.9% by the second author (an experienced clinician). In one other community survey in which the PSE was administered by trained lay interviewers (Dean *et al.* 1983), it was found that the rating threshold adopted was such as to increase the prevalence estimate of specific disorders by approximately a third (Surtees, personal communication).

A feature of all three of the studies just mentioned was the relative inexperience of the interviewers in conducting health and, in particular, mental health surveys. With the recent achievement of full scale national surveys of psychiatric morbidity (Kessler *et al.* 1994; Jenkins *et al.* 1997b) it has been generally recognized that full time professional survey interviews offer a considerable body of experience of health and related social surveys. Their ability to achieve high response rates and to carry out structured psychiatric diagnostic interviews is by now well accepted; but their potential as semi-structured interviewers for mental disorders has never been assessed. Success in this respect would potentially create a
breakthrough in the quality and usefulness of future psychiatric epidemiological data.

The Office for National Statistics (ONS) survey of psychiatric morbidity among adults in the general population used a design in which lay interviewers visited addresses and administered a questionnaire which, among other things, acted as a sift to find adults with possible psychosis (Meltzer et al. 1995; Jenkins et al. 1997a, b). Clinical psychiatrists then visited those adults who had sifted positive and attempted to carry out a SCAN Version 1 interview (Wing et al. 1990) to assess the presence of severe and psychotic psychiatric disorders. At the time the survey was carried out it was felt that the SCAN interview, because of its semi-structured nature and reliance on clinical judgment, was not appropriate for use by ONS interviewers.

A major drawback of the two phase design, a clinical interview following a lay interview, is that there will always be a level of drop out between the initial screening interview and the follow up SCAN. In the general population, about 60% of adults sifted eligible for the SCAN interview by a psychiatrist (working in his or her spare time) were interviewed within a pre-set time limit (Jenkins et al. 1997a). It is, therefore, clearly an advantage if lay interviewers can be trained to carry out a full semi-structured interview or, alternatively, that the considerable cost of recruiting and funding psychiatrists to work full time on such surveys can be obtained.

Aim and objectives

Given the differences between systematic lay and clinical interview measures and the drawbacks of two phase designs in private household surveys, it would be highly desirable to be able to conduct systematic clinical assessments in large scale surveys. Clinically more relevant information would be obtained, yielding more reliable estimates of cases ‘in need of treatment’.

Therefore, the aim of the present study was to assess the feasibility of training ONS interviewers in carrying out clinical assessments of neurotic and psychotic disorders using the semi-structured SCAN interview, to yield ICD-10 diagnoses (World Health Organization, 1992). The ONS pool of interviewers are all trained to a high standard in quantitative questionnaire techniques, but contains people from different backgrounds and with different levels of experience, some of whom may be better suited to training for clinical assessment. Therefore, the investigators also wished to try to identify, in very general terms, the characteristics of lay interviewers best suited to SCAN training, and to evaluate the effectiveness of their training.

Although the study was not designed to assess the interviewers’ suitability in any detail, a specific objective of the study was to investigate in a preliminary way the reliability of SCAN interviews carried out by ONS interviewers following an extended period of training. The outcome of this investigation would determine whether the next and most crucial stage of such work could be justified, namely the assessment of the reliability of the method in the general population. The work was carried out as a joint project between ONS and the University of Leicester, bringing together the expertise of WHO SCAN training and development centres in Great Britain and the Netherlands.

**METHOD**

**Design of the study**

To assess accurately the relative abilities of different types of interviewers for SCAN training would require an experimental design involving a large number of interviewers and subjects, and was not within the scope of the current study. Instead, a small number of interviewers were used with the aim of assessing whether or not using interviewers in this way was feasible following intensive training.

Eight lay interviewers were asked to carry out ten interviews each within an acute clinical population setting, followed by an interview of the same subjects by a clinician trained in the SCAN. As the SCAN was designed for clinicians, their interview was regarded as the standard by which the lay interviewers were evaluated. Test–retest reliability statistics were used to measure the degree of agreement between the lay interviewer and the clinician at a diagnostic level where both used the same instrument: SCAN-SF, which is a survey form of SCAN version 1 (Brugha & Nienhuis, 1998a). Ethical approval was sought and obtained for the study from the Leicestershire REC.
Training the interviewers for the study

Details of the training are described in a supplementary training document to the SCAN-SF manual, which is available on request (Brugha & Nienhuis, 1998b). Training in Leicester was carried out, in the main, by WHO accredited trainers at the University of Leicester and the University of Groningen, the Netherlands. Training in SCAN can only be obtained from officially accredited WHO training centres.

Twelve potential trainees from ONS were invited to a 1-day introduction; they were chosen from the survey field force in the English Midlands, both to be able to commute to our centre and because this method of selection would be manageable in a full scale national survey. Each was assessed as to his or her suitability and eight (mainly more experienced) interviewers were selected for SCAN training and to carry out the feasibility study. Two of these interviewers had been with ONS for less than 2 years and the remainder typically for 5 years or more. Three very experienced interviewers had previous experience in qualitative and cognitive interviewing and as trainers. The interviewers were aged from 45 to 55 years.

Interviewers were asked to familiarize themselves with the basic concepts and sections of SCAN-SF by spending 8 hours reading the SCAN Manual and Glossary. A week-long training course in Leicester then provided training in SCAN. To attempt to address their lack of clinical experience and to convey what SCAN was about, interviewers were also issued a first draft of the SCAN-SF introductory manual covering assessment methods in psychiatry and basic concepts of psychopathology (Brugha & Nienhuis, 1998a). This document also specified the sections of SCAN that they should cover and those which should be omitted (see below) from SCAN-SF. The document includes a very basic introduction to the distinction between clinical and structured interviewing (ONS staff are completely familiar with structured interviewing). As far as possible lay terms are used and technical terms (jargon) are defined. It provides introductory definitions to the concept of a symptom, neurotic symptoms (extreme emotions), perceptual abnormalities (abnormal experiences) and delusions (unusual beliefs and inferences) and to the distinction between these and diagnosis (and the classification of disorders). As training proceeded requests for further clarification of terms used in the SCAN interview schedule and glossary were given.

Having completed the standard SCAN training course, ONS interviewers carried out practice interviews with approximately ten acute psychiatric patients and sat in on ten interviews of a colleague. Interviewers had access to a SCAN trainer to answer queries, and these clinicians supervised approximately half of the practice interviews. Care was taken that each interviewer had a mix of patients with neurotic and psychotic disorders. This period of extended training lasted a further 3 weeks.

The field trials

The training and field trials were conducted with the slightly modified Survey Format (Brugha & Nienhuis, 1998a) of SCAN version 1 (World Health Organization Division of Mental Health, 1992). SCAN includes the tenth version of the Present State examination (Wing et al., 1990). It differs from its fore runner, the ninth version of the PSE (Wing et al. 1974) in several ways. One such difference may be crucial to overcoming the possible problem of lay interviewers employing lower symptom rating thresholds. The standard rating scale for neurotic symptoms in the tenth version includes a ‘sub-threshold’ rating. This is used when the respondent complains of a symptom but where the interviewer judges that the criteria in the PSE symptom glossary are not fully satisfied.

The Survey Format covers most neurotic and psychotic sections of SCAN but excludes the somatoform, alcohol, drug and tobacco, and cognitive dysfunction sections and all ‘organic attributions’ (Brugha & Nienhuis, 1998a). These sections require examiners to have an extensive clinical background. We believe that it should be possible to apply the format also to other versions of the PSE and the SCAN. A supplementary training manual for SCAN training centres was also developed during the study (Brugha & Nienhuis, 1998b).

Each interviewer was asked to carry out a SCAN-SF interview with ten subjects, some of whom had a psychotic disorder, some a neurotic disorder and the remainder had no significant
psychiatric problem. The interviewers were blind as to the status of each subject. Each interview lasted between 60 and 90 min, although interviews with non-cases tended to be shorter. Fieldwork took place over a period of 4 weeks.

**Procedures**

Patients who consented to be subjects for the study, were chosen by one of two researchers at the Section of Social and Epidemiological Psychiatry, University of Leicester. They were assisted in this by the clinical staff responsible for patients in the Leicestershire Mental Health Services NHS Trust. They also recruited ‘normal’ volunteers from the local community: no attempt was made to screen for symptoms and some volunteers may well have had one or several of the commoner neurotic symptoms found in the general population. Most patient and volunteer interviews took place in private rooms reserved for the purpose within two acute general psychiatric hospital units in Leicester: Bradgate and Brandon. Follow-up interviews were carried out within 1 week by three psychiatrists and two clinically experienced psychologists, all trained in SCAN. Usually the second interview took place on the day following the first interview.

**Evaluation of feasibility exercise and analyses**

The clinician’s judgement was regarded as the standard and comparisons were made, examining levels of agreement between the ONS interviewers and clinicians. In evaluating the interviewers’ work, a hierarchy of diagnostic boundaries was applied. These ranged from: a two-fold classification (absence or presence of psychiatric disorder); a three-fold classification (absence of psychiatric disorder, neurotic disorder, psychotic disorder); to, specific ICD-10 F codes (F1, F2, F3 etc.), this part of the analysis being non-hierarchical.

At each level 2 × 2 tables were obtained. We examined the degree of concordance between clinical and lay interviews by applying a coefficient of agreement, Cohen’s kappa (Cohen, 1968). It is recognized that complete agreement between lay interviewers and clinicians would be impossible. Before the start of the fieldwork, it was decided that agreement would be judged as follows: a kappa statistic below 0·4 would be deemed poor, while over 0·6 would be considered as good. However, we also took account of the levels of agreement achieved by clinicians using the same test–retest design in the earlier field trials of SCAN (Wing et al. 1998) as a standard to be aimed for, if not necessarily fully achievable. Errors at the two- or three-fold classification were regarded as crucial. Errors at the more detailed level were regarded as descriptive rather than critical. We also estimated the sensitivity and specificity of the ONS findings against the standard assessments by the clinicians. For all statistics calculated 95% confidence intervals were obtained.

We also examined the data for evidence of interviewers making use of different rating thresholds, or ‘rating bias’, such as had occurred in earlier studies in which lay interviewers had rated more symptoms present than clinician interviewers. In addition to the statistical analysis, ONS interviewers were also asked to comment on their training and field work experiences at a post-data collection debriefing.

**RESULTS**

**Response**

Of the 80 planned initial SCAN interviews by ONS interviewers 78 were completed successfully. The two failures occurred at the end of fieldwork when it was not possible to schedule replacements following the last minute refusal by two patients who originally agreed to take part. There were 64 successful SCAN re-interviews by clinicians. Seven patients refused a second interview (9%); one patient terminated their first interview before it was completed and one patient did likewise at their second interview; one patient went on leave (on subsequent occasions when this happened it was possible to arrange a home interview within a week of the first interview); one patient ‘went off sick’ and in two cases a clinician was not available to conduct the interview within the 1 week time period.

**Comparisons of ONS and clinician SCAN interviews**

The data were analysed by two tables comparing: true positives (D); false positives (B); true negatives (A); and false negatives (C). Kappa, sensitivity and specificity statistics were calculated with 95% confidence intervals.
Table 1. Comparison of ONS and clinician administered SCAN data according to ICD-10 Classification (N = 61 available data pairs)

<table>
<thead>
<tr>
<th>Clinician SCAN:</th>
<th>–</th>
<th>–</th>
<th>+</th>
<th>+</th>
<th>Sensitivity</th>
<th>95% CI</th>
<th>Column A</th>
<th>Column D</th>
<th>95% CI</th>
<th>Specificity</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONS SCAN:</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>Kappa</td>
<td>(C+D)</td>
<td>(A+B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Column</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any neurosis</td>
<td>25</td>
<td>2</td>
<td>6</td>
<td>28</td>
<td>0.74</td>
<td>0.57-0.91</td>
<td>0.82</td>
<td>0.66-0.93</td>
<td>0.93</td>
<td>0.76-0.99</td>
<td></td>
</tr>
<tr>
<td>Any psychosis</td>
<td>43</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.63</td>
<td>0.40-0.86</td>
<td>0.67</td>
<td>0.38-0.88</td>
<td>0.93</td>
<td>0.76-0.99</td>
<td></td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>48</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>0.71</td>
<td>0.48-0.95</td>
<td>0.73</td>
<td>0.39-0.93</td>
<td>0.96</td>
<td>0.86-0.96</td>
<td></td>
</tr>
<tr>
<td>delusion disorder</td>
<td>55</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0.64</td>
<td>0.27-1.02</td>
<td>0.6</td>
<td>0.15-0.95</td>
<td>0.98</td>
<td>0.91-1.00</td>
<td></td>
</tr>
<tr>
<td>Affective psychosis</td>
<td>30</td>
<td>6</td>
<td>5</td>
<td>20</td>
<td>0.63</td>
<td>0.43-0.83</td>
<td>0.80</td>
<td>0.59-0.93</td>
<td>0.83</td>
<td>0.67-0.93</td>
<td></td>
</tr>
<tr>
<td>Depressive disorder</td>
<td>38</td>
<td>7</td>
<td>6</td>
<td>10</td>
<td>0.46</td>
<td>0.21-0.71</td>
<td>0.63</td>
<td>0.36-0.85</td>
<td>0.85</td>
<td>0.71-0.94</td>
<td></td>
</tr>
<tr>
<td>Panic disorder</td>
<td>47</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>0.52</td>
<td>0.23-0.81</td>
<td>0.60</td>
<td>0.26-0.88</td>
<td>0.92</td>
<td>0.81-0.98</td>
<td></td>
</tr>
<tr>
<td>Phobia</td>
<td>43</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>0.51</td>
<td>0.25-0.78</td>
<td>0.62</td>
<td>0.32-0.86</td>
<td>0.90</td>
<td>0.77-0.97</td>
<td></td>
</tr>
</tbody>
</table>

The following results are based on 61 complete pairs of SCAN interviews, the first carried out by an ONS interviewer and the second by a clinically experienced SCAN interviewer within the maximum time limit of 1 week between interviews.

ICD-10 F codes (F1, F2, F3 etc.)
The main findings are set out in Table 1. Levels of agreement were generally good, with kappa coefficients of 0.74 for the twofold classification of the presence/absence of any disorder. At the most detailed level there were too few specific subtypes of psychosis cases to make comparisons meaningful (schizophrenia, mania, bipolar psychosis). When all functional psychosis was considered there seemed to be remarkably good agreement, high sensitivity and specificity and no trend suggestive of a bias between raters. Similarly, when all F2 disorders were combined (schizophrenia or delusional disorder, but excluding schizotypal disorder, which is not systematically assessed by SCAN) and when all affective psychoses (parts of F3) were combined, the findings were acceptable.

Similarly, for ‘any’ neurotic disorder the findings were very good but for specific disorders the findings were not as good. For generalized anxiety disorder there were four ONS cases and one clinician case but no concordant cases. Kappa could not be calculated for obsessive–compulsive disorder (one ONS case only and thus not concordant). We could find no evidence of interviewers making use of different rating thresholds, noting a clear lack of ‘rating bias’ throughout the study.

Comments from the ONS interviewers and SCAN trainers
ONS interviewers were asked for their reactions to the whole exercise at a post-data collection debriefing. At first, ONS interviewers found SCAN quite daunting. However, once they had begun to gain experience in carrying out interviews, their confidence grew and with it enthusiasm for the procedure.

Many of the interviewers had worked in one or more of the ONS national psychiatric morbidity surveys: in private households, among homeless people, in institutions or in the national prison survey. They felt happy and confident using the SCAN interview.

The SCAN trainers were generally happy with the progress of the ONS interviewers during training. There were some concerns about one or two less experienced interviewers, but these concerns were eventually overcome through closer attention from trainers. ONS trainees did appear to need the full training period arranged by the SCAN trainers.

DISCUSSION
This small feasibility study appears to show that lay, highly experienced survey interviewers can be taught the SCAN, a clinical, semi-structured interview within a period of 3 to 4 weeks.
Main findings

The results of the study appear to counter previously expressed reservations about the ability of lay interviewers to become sufficiently familiar with psychotic symptoms so that they can assess them. In general there was good agreement between lay and clinician assessments in the test–retest design, which is a stringent method for evaluating measurement consistency. It should be noted that in the WHO PSE-10 SCAN field trials (Wing et al. 1998) disorders with test–retest reliability levels in the intermediate range (kappa = 0.4 to 0.6) typically had very satisfactory inter-rater reliability agreement (0.7 and above), based on a single interview rated simultaneously by two clinicians. We did not use this method as we felt it is open to bias due to the co-rater ‘best guessing’ the likely rating of the principle assessor in such test settings.

The levels of agreement achieved for the broad groupings of psychotic disorders and for neurotic disorders are as good as can be hoped for in that they achieve the levels reported in the WHO PSE-10 SCAN field trials in clinical settings (Wing et al. 1998). The short delay between most pairs of interviews may have minimized the inevitable fluctuations in symptoms expected in acute patients. Interviewers may have found it easier than in a community survey to ‘guess’ which respondents were not patients and this might explain the excellent findings for ‘any psychosis or neurosis’. However, the good findings for other disorders would also have to depend on successful discrimination between symptoms underlying different types of disorder. Findings in the present study for very specific diagnoses are generally less good but most of these can be attributed to the rarity of the disorders and inevitable instability at the margins. The findings for the commoner specific neurotic disorder, including depressive disorder (kappa = 0.46), appear disappointing, and in the example of depressive disorder somewhat surprising in comparison with other semi-structured and structured interviews used in clinical populations. But these findings are difficult to explore further because the numbers of cases are two small. The findings for anxiety disorders are, however, as good or better than those achieved under identical test–retest conditions by experienced clinicians in the earlier field trials (Wing et al. 1998). Sensitivity was just about acceptable; but specificity and lack of bias were excellent.

Even in a clinically acute sample little fall off in positively rated symptoms would be expected in a study in which most interviews have been achieved within a day of each other. Therefore, the lack of bias comparing ONS and clinician SCAN interview data suggests that we may have made progress in overcoming the persistent problem of lay rater bias identified in previous studies (Rodgers & Mann, 1986). However, only an evaluation in a household sample can verify this.

Training

The reliability findings for the commoner specific neurotic disorders were somewhat disappointing. Further analyses, comparing individual item ratings, may also help to identify whether a few items can be identified that require extra attention and perhaps extended and less ambiguous glossary definitions for lay users. The interviewers commented that it was easier to be sure about the ‘abnormal’ significance of psychotic symptoms, than about neurotic symptoms, such as depressed mood, a symptom that exists on a continuum. One had to work quite hard sometimes to establish whether severity and pervasiveness of the complaint was out of proportion to the day-to-day circumstances of the patient, which hardly ever posed a difficulty with psychotic items. The training course placed a heavy emphasis on psychosis and less attention was given to the SCAN depression sections both in the patients chosen for training interviews and in the training videos used on the course. The possibility that the assessment of neurotic and particularly depressive symptoms will require more particular attention in training materials and in training interviews in future courses with lay interviewers needs to be evaluated further. The advent of videos with digitally identified segments will allow us to prepare examples of definite (present, absent) and borderline examples of the commoner depressive and other neurotic symptoms. As digital encoding and
retrieval of video segments is becoming less costly we envisage trainees being able to call up examples of items they find more difficult to rate (e.g. judging when is low mood following a loss event normal or pathological and the use of cross questioning to assess impairment in such examples), which could be in the form of self-directed learning material. A final important point is that we were gaining experience and learning lessons about training as this first study proceeded: building on this experience, future courses should be more successful in producing good quality interviewers.

Feasibility
Because we were restricted in choosing locally available interviewers we have no doubt that sufficient numbers of the same calibre would be available throughout Great Britain from the field force of a nationally based survey provider such as ONS and that the same would apply for any other country with an active economic and social survey programme. The ample availability and relatively low cost of the interviewers does need to be balanced against the high cost of training, which requires about three times more time than that needed by clinically experienced interviewers. During fieldwork attention would need to be given to quality of data and to identifying and correcting any tendency for rating thresholds to drift, but arguably, such attention should be given also when clinicians are employed.

Conclusion and recommendation
The financial cost of such surveys would be greater than for current methods. Central and local funding allocation estimates for health and social care (Brugha et al. 1997) should be based on information that is relevant to the nature of the problem and the very considerable level of public expenditure involved (at least a fifth of the NHS budget). The choice of information collection method used should be informed by debate about the underlying nature of differences in the survey interviews themselves (Brugha et al. 1999a) as well as by a consideration of the reliability of the data that both kinds of methods can provide through careful attention to quality control.

Fresh solutions must be found to the problems that have emerged with the use of lay diagnostic interviews. A variety of solutions could be evaluated including, for instance, the collection by lay interviewers of examples of symptoms in the form of vignettes for later rating by an expert, a method which is being considered in the forthcoming national children survey. However, one approach would be to make it possible for lay interviewers to make full use of clinically based measures following an extended training period. The initial success of the work described here leads us to believe that further work on SCAN with lay interviewers is warranted. The next stage of development and feasibility evaluation should be carried out in a larger sample of householders over a wide geographical area. Until the feasibility and reliability of procedures such as SCAN-SF administered by carefully trained and experienced lay survey interviewers has been demonstrated in household populations, its potential for meeting the needs of researchers and policy developers must remain uncertain.

This paper is based on a Report to the Department of Health, Statistics Division, London, who provided funding for the training and fieldwork. We are indebted to the clinical staff and patients at the Leicestershire Mental Health Services Trust, to the following psychiatrists who provided SCAN interviewing time: Dr Alex Witcomb and Dr Mangesh Marudkar; and to the social survey interviewers from the Office for National Statistics for their dedication and enthusiasm.

REFERENCES
The survey form of SCAN


