

Development of a scale to measure symptoms of anxiety and depression in the general UK population: the psychiatric symptom frequency scale

Malin Lindelow, Rebecca Hardy, Bryan Rodgers

Abstract

Objectives—The psychiatric symptom frequency (PSF) scale was developed to assess symptoms of anxiety and depression (ie affective symptoms) experienced over the past year in the general population. This study aimed to examine the distribution of PSF scores, internal consistency, and factor structure and to investigate relationships between total scores for this scale and other indicators of poor mental health.

Participants—The Medical Research Council national survey of health and development, a class stratified cohort study of men and women followed up from birth in 1946, with the most recent interview at age 43 when the PSF scale was administered.

Main results—The PSF scale showed high internal consistency between the 18 items (Cronbach's $\alpha=0.88$). Ratings on items of the scale reflected one predominant factor, incorporating both depression and anxiety, and two additional factors of less statistical importance, one reflecting sleep problems and the other panic and situational anxiety. Total scores were calculated by adding 18 items of the scale, and high total scores were found to be strongly associated with reports of contact with a doctor or other health professional and use of prescribed medication for "nervous or emotional trouble or depression," and with suicidal ideas.

Conclusions—The PSF is a useful and valid scale for evaluating affective symptoms in the general population. It is appropriate for administration by lay interviewers with minimal training, is relatively brief, and generates few missing data. The total score is a flexible measure which can be used in continuous or binary form to suit the purposes of individual investigations, and provides discrimination at lower as well as upper levels of symptom severity.

(*J Epidemiol Community Health* 1997;51:549-557)

Depressive and anxiety disorders together constitute a considerable proportion of all psychiatric disorders in adults. The United States national comorbidity survey estimated the 12

month prevalence of depressive disorders at 11.3% and that of anxiety disorders at 17.2%.¹ There is substantial comorbidity of these two groups of disorders in clinical and general population samples²⁻⁴ and, particularly for less severe disorders, it appears arbitrary to attempt to draw a distinction. The terms "common mental disorders"⁵ and "neurotic disorders"⁶ have been applied to this broader category, as has "affective disorders",⁷ although usage of this last term in the United States typically excludes anxiety disorders.⁸ The above prevalence estimates do not include the many individuals who experience symptoms that fall short of criteria for formal diagnoses. In the OPCS surveys of psychiatric morbidity in Great Britain,⁶ in which the clinical interview schedule was used, a category of mixed anxiety and depressive disorder was defined. This covered symptoms of anxiety and depression which did not meet criteria for psychosis, depressive disorder, generalised anxiety disorder, phobias, obsessive-compulsive disorder, or panic disorder. The one week prevalence for this mixed category was 77 per thousand out of a total of 160 per thousand for all neurotic disorders. Overall, patients with affective symptoms account for considerable use of services, especially primary care,⁹ and a substantial proportion of time off work,¹⁰ not to mention the contribution these symptoms make to social role handicaps and personal distress.

This subject is therefore of major importance in health services and epidemiological research. However, investigating affective disorders, as for psychiatric disorders in general, is hampered by difficulties in measurement. Clinical diagnostic evaluation is time consuming and costly, and impractical for large scale community studies. This has led to the development of alternative approaches, including self completion inventories, such as the Center for Epidemiologic Studies depression scale (CES-D)¹¹ and the general health questionnaire (GHQ),^{12,13} and more extensive structured interviews such as the present state examination (PSE)^{14,15} and the composite international diagnostic interview (CIDI).¹⁶ There are now many instruments, but selecting one for a particular investigation is not necessarily a simple process. They vary in many respects, including the time and resources needed for training interviewers, the time and cost of administration, the necessity for main-

MRC National Survey of Health and Development, University College London Medical School, Department of Epidemiology and Public Health, 1-19 Torrington Place, London WC1E 6BT.
M Lindelow
R Hardy

NH&MRC Social Psychiatry Research Unit, The Australian National University, B Rodgers

Correspondence to:
Dr R Hardy.

Accepted for publication
January 1997

taining standards during fieldwork, the demands placed on respondents, the range of symptoms included, the time frame of assessment, the dimensions of severity encompassed, the provision or not of subscales or diagnostic subgrouping, the distribution of total scores (for continuous measures), and the threshold for identifying disorders (for discrete measures). The advantages and disadvantages of individual methods of evaluation must be considered in the context of any given study's resources, scope, and objectives

In planning the 43 years of age (1989) follow up of the Medical Research Council's national survey of health and development (NSHD), for which assessment of mental health comprised only a part of the full interview, such issues had to be addressed in choosing a method for the evaluation of affective disorders. The NSHD is a prospective longitudinal study of 5362 men and women born in England, Scotland, and Wales during the week 3-9 March 1946.¹⁷ All legitimate, single births from non-manual and agricultural backgrounds, and one quarter of children from manual class backgrounds were selected for inclusion in the study. Data have been collected at regular intervals, at least every two years during childhood and less frequently during adulthood, with the previous collection having been in 1982 when the survey members were 36 years of age.

The short form of the PSE had been used in the 36 year follow up, but concerns regarding its suitability had arisen. From a practical point of view, the considerable time required, both for training interviewers and administration during fieldwork, made the PSE expensive, while also limiting the time available for assessing other aspects of psychological functioning. There was a tendency for interviewers' ratings to 'drift' over the course of the data collection¹⁸ and although audiotaping a proportion of interviews helped to maintain standards, this was both time consuming and more intrusive than wished. Furthermore, the very skewed distribution of total PSE scores, with many subjects scoring zero, was a particular disadvantage in a longitudinal study such as the NSHD where differences in symptom levels in the subclinical range may be important for predicting future symptoms of greater severity and where the study of well being (in contrast to 'ill being') may be of interest. Finally, the diagnostic classification given by the PSE did not prove useful in that few variables (past or contemporary) were found to discriminate between diagnostic groups.^{19, 20}

These concerns led to the specification of particular requirements for the instrument to be used at the 1989 data collection as follows.

- In addition to the obvious necessity for internal reliability, face validity, and predictive validity with respect to external criteria, it was necessary that items should first cover the symptoms indicative of both depressive and anxiety disorders.
- The time frame for assessment should be sufficiently long to avoid high severity ratings

KEY POINTS

- The psychiatric symptom frequency (PSF) scale is reliable and valid for evaluating symptoms in the general population.
- It is suitable for use by lay interviewers takes little time, and generates few missing values.
- It provides discrimination at the lower, as well as the higher, end of the scale.
- ROC analysis is a valuable means of quantifying the validity of symptom scales.

arising from brief recent episodes, but not so long as to introduce problems of recollection.

- The format of the schedule should be suitable either for administration by lay interviewers or for self completion.
- The need for training of interviewers or for instruction of subjects should be kept to a minimum.
- The time for administration/completion should be short, preferably less than 10 minutes.
- Resources for monitoring the quality of assessment during fieldwork should be minimal, and any methods of quality control should not interfere with assessment.
- The schedule should provide a scale which gives discrimination at the lower, as well as the upper, end of its range. Its distribution should be suitable for the application of different cut off points identifying high scoring individuals, and for studying psychological well being as a continuum.

When the data collection was planned in 1988, there was no single instrument that met all these criteria, so a new scale was developed (by BR) to incorporate all seven desirable characteristics listed above. Of the existing instruments, the psychiatric epidemiology research interview (PERI)²¹ perhaps came the closest to the specifications. However, the items included in the PERI did not have a close correspondence to symptoms contained in standard classifications of the relevant psychiatric disorders. There were additional difficulties arising from the scales being developed for the United States population which meant that many items would have required rewording. Other possible candidates included the CES-D scale,¹¹ but this typically assesses symptoms over just one week. The various forms of the general health questionnaire²² were considered inappropriate because of the unspecified period of assessment and because the response frame calls for comparison with how individuals usually are or usually feel. A number of symptom check lists, such as the Malaise scale,²³ were problematic because items scored only as zero or one (ie absent or present), when they have high thresholds and are substantially intercorrelated, will generate a large proportion of zero values for total scores and yield poor discrimination at lower severity levels.

This paper describes the development of the new instrument in the context of the NSHD, highlighting the characteristics required. The aim of the analysis is to examine the distribution, internal consistency and factor structure of the PSF scale, as well as its relationships with suicidal ideation, and with use of prescribed medication and contact with a doctor for 'nervous or emotional trouble or depression'.

Methods

INSTRUMENT DEVELOPMENT

Possible items were selected from the short form of the PSE, used previously in the study,¹⁸ and from other instruments covering symptoms of unipolar depressive disorders, agoraphobia, social phobia, panic disorder, and generalised anxiety disorder. Items from the NSHD 36 year study that had contributed least to the discrimination of those seeking treatment included worrying, tension pains, free floating anxiety, specific phobias, anxiety avoidance, social withdrawal, self depreciation, lack of self confidence, ideas of reference, guilty ideas of reference, pathological guilt, expansive mood, ideomotor pressure, grandiose ideas/actions, obsessional checking/repeating, obsessional cleanliness/rituals, obsessional ideas/rumination, derealisation, and depersonalisation. A number of these symptoms are comparatively rare in the general population.¹⁸ The least discriminating items were omitted. Somatic items were excluded because of fears that they could contribute to age differences if used in future follow ups as the inclusion of somatic items has been the subject of concern for other instruments.^{22 24 25} Crying (or tearfulness) was also excluded because of the particularly strong gender difference for such an item in adult samples.

Wording of items followed the principles for the study's interview as a whole, being as simple as possible. In some instances, PSE items were suitable or required only minor modifications, eg "Have you felt on edge or keyed up or mentally tense?" When PSE wording was inappropriate, items from other instruments were adapted. In order to maximise the information obtained from relatively few questions, it was decided to utilise multiple response categories or a rating scale for each item. A number of dimensions of severity were considered, including: frequency of onset, duration, subjective intensity and degree of associated disability. Frequency and duration appeared the most appropriate, and permitted the use of a common response frame for all items.

A period of assessment of at least a few months was thought necessary in order to minimise the impact of brief recent episodes. A long time period would also reduce the proportion of subjects who had never experienced a particular symptom, and therefore minimise the number scoring zero on the full scale. Although accurate recollection over one year may be problematic, the choice of 12 months had the advantage of giving a simple reference point for the start of the period, while also limiting possible seasonal variations in interviews conducted at different points in the

year. This one year time frame was also consistent with other sections of the questionnaire requiring retrospective information.

PILOT STUDY

A pilot study for the 1989 interview was carried out using a sample close in age to the NSHD cohort, identified from GP registers in two geographical areas of England. Interviews were conducted by NSHD scientific staff in the respondents' homes. A trial version of the PSF scale was included in the study, with subjects asked 'how often' they had felt or experienced each symptom over the last year and given an answer card to indicate responses on a six point scale. These responses and their respective ratings were never (0), occasionally (1), sometimes (2), quite often (3), very often (4), and always (5). Problems with the questionnaire were identified not only from routine administration, but also by asking subjects at the end of each interview about their reaction to and interpretation of specific questions, and by double checking responses given during the interview and requesting elaboration on these. One difficulty with the symptom ratings was identified very early in the pilot phases. Specifically, there were large individual differences in the meaning attached to the terms on the answer card used to indicate frequency. The same objective frequency of symptoms could span as many as three points on the subjective scale and, consequently, the instrument was modified so that open-ended responses were elicited without use of an answer card. These responses were collated and used to construct a rating scale for interviewers as follows:

- 0 = never in the last year.
- 1 = up to 10 days in total, less than once a month.
- 2 = a spell up to one month, once or twice a month, 'a months worth'.
- 3 = a spell up to four months, once or twice a week, three to ten times a month.
- 4 = a spell of over four months, three or more times a week, 11 or more times a month.
- 5 = every day in the last year.

This scale reflected the tendency for spontaneous replies to be given either in the form of an episode or a rate (typically weekly or monthly), but did not distinguish between these two modes of responding.

The pilot interview also led to some modifications to the wording of items. One item covering difficulties in making decisions was deleted because it was often construed by subjects as a personality characteristic independent of psychological distress or disorder. This left 18 items plus one set of questions relating to suicidal ideation and acts that were reformulated into a composite Guttman type scale, ranging from never feeling that life was hardly worth living in the last year through to attempting to take one's own life, still yielding a 0 to 5 rating, but forsaking responding in terms of frequency of experience. The resultant questions, shown in the Appendix, were used in the interviews conducted with the NSHD

sample, with interviewers rating open-ended responses on symptom frequency.

MAIN STUDY: SAMPLE AND DATA COLLECTION

The interviews at which the PSF was administered were conducted by trained nurses during 1989-1990 when survey members were 43 years of age. At that time it was known that 365 (6.8%) of the original cohort of survey members had died, and 618 (11.5%) had left the country. Of those thought to be still living in Britain 3262 (74.5%) were successfully interviewed. Of those not interviewed, 646 (14.8%) had refused to participate further in the study, 276 (6.3%) were untraced, and 195 (4.5%) were not willing to be interviewed on this particular occasion. However, the population interviewed at the age of 43 years were, in most respects, representative of the native born population of that age.²⁶ Certain small groups of individuals are, however, under-represented, specifically those with poor literacy skills, learning disabilities, and serious psychiatric disorders.

The total score on the PSF scale was calculated by adding the scores of the first 18 items of the instrument, excluding the questions on suicidal ideation. Most subjects interviewed (3196/3262=98%) provided a response for every item, thus suggesting that the PSF has a good completion rate. The group of individuals with at least one missing item were no different to those with no missing items in terms of gender, socioeconomic status, and educational achievement. For subjects missing only one or two of the contributing items, total PSF scores were estimated using the technique of equivalent centile points. As a consequence, only 20 subjects, with more than two missing items, had to be excluded from the analysis due to missing data on the PSF.

The item referring to suicidal ideation and acts was dichotomised to contrast individuals who reported neither (at most stating that at some point over the past year they had felt that life was hardly worth living) with those who admitted having thought that they would be better off dead, thought of taking their own life, made plans to take their own life, or attempted to take their own life. Only a small proportion of the respondents (4.1%) fell into this latter group.

In a prior part of the interview, survey members were asked about their service contact and treatment histories regarding a wide range of medical complaints experienced over the past year. Within this context they were asked if they had seen a doctor or other health professional for 'nervous or emotional trouble or depression'. Survey members who responded affirmatively to this were contrasted with those who denied such contact, independently of the number of consultations that had occurred. Similarly, information regarding the use of prescribed medication for 'nervous or emotional trouble or depression' was requested. Again, those reporting such use were contrasted with those giving negative answers.

In the NSHD population, 5.8% reported contact with a doctor or other health profes-

sional, and 4.3% reported that they were taking prescribed medication. As expected, almost all of those taking such medication also reported having seen a doctor or other health professional. The rate for reported contacts with a health professional is, as expected, lower than reported prevalence rates for affective disorders. However, the rates for service use found in the NSHD correspond well with those found in the OPCS *Morbidity Statistics from General Practice*,⁹ where using their findings for annual prevalence of 'neurotic disorders' diagnosed in visits to general practices, a rate of 4.3% in a population the age of the NSHD would be expected.

Exclusion from all analyses of those survey members with missing values for either these additional variables or the PSF scale resulted in a final sample size of 3058 (94% of those interviewed).

STATISTICAL METHODS

Statistical analysis was undertaken in two steps. First the internal consistency and factor structure of the scale were examined using Cronbach's α and factor analysis. Second the criterion-related validity of the scale was examined by reference to contact with a doctor or other health professional, use of prescribed medication, and suicidal ideation. Here, both binary and continuous versions of the PSF were investigated, using χ^2 tests and receiver operating characteristic (ROC) curves²⁷ respectively.

An ROC curve may be used to examine whether any proposed diagnostic score, measured on a continuous numerical scale, can discriminate between a "diseased" group and a "normal" group. Essentially, it examines the degree of overlap of the distributions of the diagnostic score for groups defined by any specified criterion variable. In the present study, no formal clinical diagnoses of "affective disorders" were made, so that a strict definition of "diseased" and "normal" was not available. Such clinical diagnoses may, in any case, be considered too stringent for many epidemiological purposes and are themselves of questionable validity. Hence, other criteria indicative of poor mental health were used. These indicators were whether an individual had been taking prescribed medication for "nervous or emotional trouble or depression", had been in contact with a doctor or other health professional for the same reason, or had reported suicidal ideation.

For each point, or cut off, on the PSF scale, subjects may be classified as scoring high or low, ie scoring above or below that cut off. Individuals classified as high and who are also positive on a given criterion variable, for example those taking medication, are termed true positives (correctly ascertained by the PSF), while those classified as high and who are negative on the criterion variable, that is those not taking medication, are termed false positives (incorrectly ascertained by the PSF). The ROC curve itself is a plot of the true positive rate against the false positive rate for every point of the PSF scale, that is sensitivity plotted

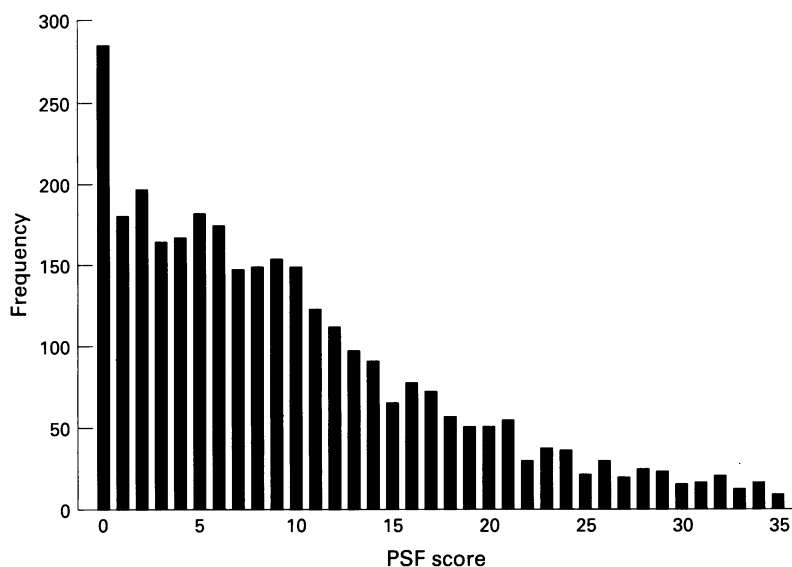


Figure 1 Frequency distribution of the lowest 95% of the psychiatric symptom frequency (PSF) score.

against one minus specificity. The degree of overlap, in terms of their PSF scores, between the two groups defined by the criterion variable, is measured by calculating the area under the curve (AUC). The AUC and its associated standard error were calculated using the non-parametric method of Hanley and McNeil.²⁸

ROC curves can also be used to identify cut off points for dichotomising continuous scales. Clearly there can be no unequivocal solution, as any dichotomy is arbitrary, and consideration must be given to costs and benefits of misclassification in the specific clinical or research situations in which the dichotomous variable is used. A cut off derived from the ROC curve was compared with a cut off between the scores of 22 and 23, identifying 12.9% of the sample as scoring above the threshold, that has been used in previous analyses.^{29, 30}

Results

DISTRIBUTION OF TOTAL PSF SCORE

The distribution of the total PSF score, excluding the highest 5%, is shown in figure 1. Fewer than 10% of survey members scored 0, with the remaining scores ranging up to 90. 95% of scores were equal to or less than 35, thus giving the distribution a long right-hand tail.

INTERNAL CONSISTENCY

The internal consistency of the scale, as indicated by Cronbach's α was relatively high (0.88). Furthermore, values for α with each item deleted in turn showed that eliminating any one of the 18 items from the scale resulted in little change in the value of α . This suggests that none of the items detracted from the reliability of the overall scale.

FACTOR ANALYSIS

Exploratory factor analysis of the first 18 items was undertaken, using principal component analysis. Three factors with eigenvalues larger than 1.0 were extracted, and a varimax rotation

performed. This gave one predominant factor (eigenvalue 6.12), accounting for 34% of the variance, and two subsidiary factors which accounted for less than 7% of the variance each. All except four items were strongly associated with the first rotated factor, making it an overall factor incorporating symptoms of both depression and anxiety. The other four items were linked to the second and third factors. One of these factors was concerned with two items reflecting sleep disturbance and the other with two items covering panic and situational anxiety, and both arose from the relatively high correlations between each pair of items (0.56 and 0.54 respectively). When analysis was restricted to the first (unrotated) factor, all 18 items showed factor loadings in the range 0.43 to 0.71.

ROC ANALYSES PREDICTING SERVICE USE AND SUICIDAL IDEATION

The ROC curve for the PSF score with use of medication as the criterion variable is shown in figure 2. The diagonal line represents the hypothetical case of no separation between the two distributions, and can be referred to as the line of no information. Hence, the further the ROC curve lies from this line of no information, the better the performance of the scale in discriminating between the criterion groups. Hence, in the case considered here, there is certainly some degree of discrimination by the PSF scale (fig 2).

The AUC for a random ROC curve is 0.5, whereas a perfect scale, completely separating the two groups, gives an AUC of 1.0. Hence, the relatively large AUCs for all three criterion variables considered here (table 1) indicate that the PSF score distinguishes significantly between the two groups of interest. Subjects with high scores on the PSF were more likely to have sought professional help in relation to emotional problems, to be prescribed medication and to have had suicidal feelings. ROC curves were then produced and AUCs calculated for men and women separately for each of the three criterion variables. The null hypothesis that the AUC for men was equal to the AUC

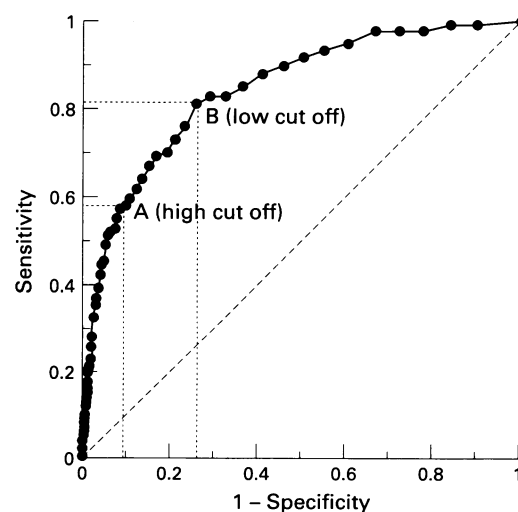


Figure 2 Receiver operating characteristic (ROC) curve for the PSF score where the diagnostic criterion is use of prescribed medication.

Table 1 Results from receiver operating characteristic (ROC) analysis

Criterion variable	Area under curve (SE)	Cut off maximising specificity and sensitivity
Contact with a doctor	0.84 (0.016)	13/14
Use of prescribed medication	0.86 (0.014)	13/14
Suicidal ideation	0.92 (0.012)	19/20

Table 2 Percentage of those with low and high psychiatric symptom frequency (PSF) scores reporting contact with a doctor, use of prescribed medication, and suicidal ideation (PSF cut off between 22 and 23)

Criterion variable	Low PSF score (n=2694)	High PSF score (n=364)	p value (χ^2 test)
Contact with a doctor	2.7% (n=74)	28.0% (n=102)	0.000
Use of prescribed medication	2.1% (n=56)	21.2% (n=77)	0.000
Suicidal ideation	1.0% (n=27)	25.5% (n=93)	0.000

for women was tested in each case. Results indicate that for both service use criterion variables, the PSF scale performs equally well for each gender. For suicidal ideation, however, the result of the test is significant at the 5% level with the AUC for men (AUC=0.95, SE=0.013) being greater than that for women (AUC=0.90, SE=0.018).

The cut off used in previous analyses, that is between 22 and 23, was applied to the 18-item scale, shown as point A in figure 2. Although in the past this cut off had been applied to a 19 item scale with suicidal ideation and acts included, the low prevalence of positive responses to this item meant that its omission caused minimal alteration to the distribution of total scores. In respect of identifying subjects who have taken prescribed medication, this had high specificity (91%), but relatively low sensitivity (58%). Alternatively, the point on the ROC curve nearest to the top left-hand corner (point B) is the one giving the best balance of specificity and sensitivity. In table 1, this is referred to as the cut off maximising specificity and sensitivity and is between scores of 13 and 14. The same cut off (13/14) was found for contact with a doctor or other health professional as for use of prescribed medication, and this classified 28.5% of the total sample as high scorers. In subsequent binary analyses, the two cut offs of 13/14 and 22/23 were compared.

BINARY ANALYSES USING SERVICE CONTACT AND SUICIDAL IDEATION AS CRITERIA

The associations between binary versions of the PSF and respondents' reports of service contact for "nervous or emotional trouble or depression" and suicidal ideation were investigated, applying χ^2 tests of significance. This confirmed that reports of health service contact, use of prescribed medication, and suicidal ideation were all strongly related to high PSF scores. For the higher cut off, typically a

Table 3 Percentage of those with low and high psychiatric symptom frequency (PSF) scores reporting contact with a doctor, use of prescribed medicine, and suicidal ideation (PSF is cut off between 13 and 14)

Criterion variable	Low PSF score (n=2187)	High PSF score (n=871)	p value (χ^2 test)
Contact with a doctor	1.4% (n=31)	16.6% (n=145)	0.000
Use of prescribed medication	1.1% (n=25)	12.4% (n=108)	0.000
Suicidal ideation	0.6% (n=12)	12.4% (n=108)	0.000

quarter of those with high PSF scores reported each of having had contact with a doctor or other health professional, having used prescribed medication, or having experienced suicidal ideation over the past year, compared to under 3% of those with low scores (table 2). For the lower cut off, only about 1% of those with low scores reported affirmatively on criterion variables, compared with 12-17% of those with high scores (table 3). In all cases these comparisons were highly significant.

Discussion

The advantages and disadvantages of the PSF scale for epidemiological research are discussed here in relation to three aspects of the instrument:

- The content and format of the scale,
- Its psychometric properties, including inter-correlations of items and associations with external criterion variables and
- The utility of resulting measures, both in discrete and continuous forms.

This will act as a guide to the suitability of the scale for particular research purposes and to the feasibility of incorporating components of the general approach into the development of new instruments.

CONTENT AND FORMAT

The items of the scale cover many of the symptoms of the target anxiety and depressive disorders.⁸⁻³¹ The most prominent exclusion, of somatic symptoms which could arise from physical illnesses, should be advantageous for samples that include the physically ill, the elderly or those experiencing physical stresses.³² The inclusion of items covering sleep problems could introduce difficulties with physically ill or elderly subjects. Sleep problems do appear, in fact, to hold up as indicators of depression in elderly subjects,³³ but their prevalence is also known to rise with age.⁹ In some circumstances it may be advisable to omit these items from the scoring of the instrument or to contrast findings with them included or not. The omission of items on self worth may be a weakness in certain contexts, although one that could be addressed if necessary. Since the development of the PSF, a report from the epidemiologic catchment area (ECA) programme has identified "worthlessness or guilt" and "trouble concentrating or thinking" as the two symptom groups most predictive of the onset of major depression.³⁴ However, these analyses were conducted for subjects who had no lifetime history of major affective disorder at the initial interview when they reported these symptoms, and their implications for the prediction of contemporary disorder are not clear. Elsewhere, data from the same study have indicated that "feelings of worthlessness" are relatively uncommon amongst subjects who currently meet criteria for major depressive disorder,³⁵ indicating poor sensitivity. It is notable, too, that the symptoms of anxiety and depression (SAD) scale, derived by the selection of items discriminating patient and normal groups,³⁶ contains no items relating to self worth. There are some additional minor peculiarities reflect-

ing the roots of the PSF scale. For example, enclosed space is mentioned in the item covering situational anxiety even though fear of enclosed spaces is usually classified with simple phobias.

The instrument satisfied the practical and administrative requirements laid down in the introduction. The format of the scale, requiring rating of open-ended responses is straightforward to carry out, although inappropriate for self-completion, and requires minimal training and monitoring of interviewers. Interviewers need to be able to cope with queries or conditional responses, for example not to include sleep problems resulting from noisy neighbours, and also to elicit and aggregate responses when subjects have experienced both acute episodes and periodic symptomatology. In general, the approach of rating the frequency of symptoms, even over a 12 month period, appears to be satisfactory. The use of the 12 month assessment period is well suited to studies of chronic adversity, risk associated with more distant past events and circumstances (eg childhood adversity), and changes in symptomatology associated with long-term transitions, as identified by demographic or socioeconomic factors. It would be inappropriate for studies wishing to investigate short term changes in symptomatology, such as those following life events. For the latter type of study, it would be possible to adapt the PSF to inquire of symptoms over a shorter period and to amend the rating scale for responses. The full scale, including questions on suicidal ideation and acts, is relatively brief, giving an average administration time of about seven minutes. This is particularly useful for studies with broad interests, where limited time is available for assessing mental health.

PSYCHOMETRIC PROPERTIES

The PSF scale has high internal consistency ($\alpha = 0.88$), thus indicating good reliability. This compares well with other similar scales, such as the GHQ-30 which was found to have an α value of 0.9,³⁷ the CES-D with a value of 0.85,¹¹ and the Malaise with a value of 0.80.³⁸ Such a level of internal consistency, as well as indicating good reliability, can also be interpreted as showing considerable redundancy of items. This is not, however, attributable to the inclusion of synonymous items, as care was taken to avoid such duplication. The implication of this redundancy is that even briefer scales could be developed with little loss of reliability and, hopefully, validity. It is inevitable, however, that reducing the number of items will increase the proportion of individuals scoring zero on the scale, and this would have some disadvantages. Two prominent specific examples of redundancy involved sleep problems and the association between situational anxiety and autonomic symptoms of panic. It is likely that the correlations between the relevant pairs of items, resulting in the second and third factors of the principle component analysis, arise from reports referring to the same occasions, that is subjective anxiety and panic occurred together for many individuals, as did trouble getting off

to sleep and waking up early. Although difficulty getting to sleep and early waking have been considered to have some diagnostic specificity, it may be acceptable in the context of measuring general affective symptomatology to incorporate both symptoms into a single item. The failure to identify two separate factors representing depression and anxiety is consistent with other studies.^{2,3} Even when separate constructs have been indicated by more sophisticated item-analysis techniques, the resultant subscales are highly correlated and do not map perfectly onto the symptom constellations contained in classification systems.³⁹

The ROC and contingency table analyses using external criterion variables gave the unexceptional findings that PSF scores were strongly associated with seeking help from health professionals, being prescribed medication and suicidal ideation and acts. Indeed, a scale professing to measure "affective disorders" would be expected to produce such associations and hence the results are encouraging in this respect. The scale was most successful in discriminating between subjects with and without suicidal ideation, as indicated by the greater area under the curve. Whereas this may be a genuine reflection of reluctance to seek help by many individuals suffering from significant psychopathology,^{40,41} it is also possible that it is an effect of the methods whereby different criterion variables were collected. That is, information about suicidal ideation was obtained in conjunction with the PSF scale, whereas information about health service contact and use of prescribed medication for nervous or emotional trouble or depression was collected at an earlier part of the interview, during which subjects were also asked about their physical health. These circumstances may well have influenced reporting.

In the absence of gold standards for diagnoses in psychiatry, ROC analysis provides a valuable means of quantifying the validity of symptom scales. The incorporation of similar criteria in other studies, or in investigations using more than one method of assessment, would permit direct comparisons of different instruments. The traditional difficulty of establishing predictive validity in psychiatric epidemiology has led to an over-reliance on reliability and construct validity, and ROC analysis offers an alternative approach. It is advisable to obtain several criterion variables in such studies, as any single outcome (such as help-seeking) may reflect unwanted biases, eg propensity to consult a doctor independent of severity of symptoms. In addition to the criteria used in the present study, sickness absence from work and interference with social roles would be suitable candidates.

UTILITY OF THE TOTAL PSF SCORE

A very large proportion of subjects had no more than two missing items on the scale (99.4%) and could therefore be ascribed a total PSF score. The long upper tail of the distribution of this total score is a desirable attribute, permitting the application of alternative or

multiple cut points to suit the aims of specific investigations. The results for binary forms of the PSF, using two different cut off points, illustrate how both may be useful in epidemiological research, but for different purposes. When using the higher cut off, a relatively large proportion of subjects with high scores report positively on the criterion variables, although a certain number of those with health service contact, use of prescribed medication, and suicidal ideation still fall into the group with low scores. This may therefore be appropriate for research where interest is in defining a population at high risk of clinical disorder. The lower cut off was successful in putting a greater number of those with affirmative answers on the criterion variable in the high scoring group, and is therefore more appropriate as a screening measure.

The discrimination at the lower end of the PSF scale, less than 10% of subjects scored zero compared to almost half on the corresponding total PSE score obtained in 1982,¹⁸ is a prerequisite for the study of well being and is of particular use in longitudinal studies where symptoms at a subclinical level may be predictive of later disorder. The overall distribution of total scores remains very skewed, however, but the relatively fine gradation of the scale allows transformation to a more normal distribution if appropriate for necessary analyses. This would still leave truncation at the lower extremity, but the fact that a significant number of individuals do not report experiencing any of 19 symptoms at any time during a 12 month period suggests that further attempts to achieve greater discrimination could be futile. Either there is a group of very resilient (or protected) individuals in the general population or there is a reluctance to admit to affective symptoms.

The PSF performed well in this nationally representative sample of 43 year olds. The results of the ROC analyses were, overall, similar for men and women. The difference between the AUCs for suicidal ideation, although being statistically significant, was small in terms of any practical implications as for both sexes the discrimination of the PSF scale was high. However, since the NSHD is a cohort of individuals who are all of the same age, no conclusions can be drawn about how the PSF operates across different age ranges. Furthermore, since there are very few individuals in this sample with serious psychiatric disorders, it is not known whether the scale is suitable for such a group and hence, studies looking at the properties of the scale in clinic populations would be informative. In addition, there are other aspects of the performance of the scale, such as the test-retest and inter-rater reliability, which have not been assessed in the present study, but which could be considered in future work.

In conclusion, the PSF scale is a useful addition to the set of instruments available for evaluating affective symptoms in adults. It is particularly suited to studies of general population samples where data collection is by lay interviewers. The PSF requires little time for administration and yields a very small propor-

tion of missing values. The total score represents one predominant dimension, consistent with the construct of "common mental disorders" or "neurotic disorders", incorporating symptoms of anxiety and depression. This score is extremely flexible and appropriate for use as a continuous measure of affective distress or as a categorical outcome measure, using a variety of cut points, where high scoring individuals are likely to have disorders of clinical significance. Its discrimination in the lower range is particularly valuable for longitudinal studies of onset of disorder. The quantification of discriminability by reference to several external criterion variables provides the opportunity for comparison with other possible instruments, and this general approach is recommended as a means of assessing validity in preference to the traditional reliance on indices of reliability.

The authors wish to thank Professor Mike Wadsworth for helpful comments on previous drafts of this paper.

Funding: the research was funded by the Medical Research Council.

Conflicts of interest: none.

- 1 Kessler RC, McGonagle KA, Zhao S, *et al.* Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. *Arch Gen Psychiatry* 1994;51:8-19.
- 2 Eaton WW, Ritter C. Distinguishing anxiety and depression with field survey data. *Psychol Med* 1988;18:155-66.
- 3 Feldman LA. Distinguishing depression and anxiety in self-report: Evidence from confirmatory factor analysis on non-clinical and clinical samples. *J Consult Clin Psychol* 1993;61:631-38.
- 4 Hunt C, Andrews G. Comorbidity in the anxiety disorders: the use of a life-chart approach. *J Psychiatric Res* 1995;29:467-80.
- 5 Goldberg DP, Huxley P. *Common mental disorders: a bio-social model*. London: Routledge, 1992.
- 6 Meltzer H, Gill B, Petticrew M, Hinds K. *The prevalence of psychiatric morbidity among adults living in private households*. OPCS surveys of psychiatric morbidity in Great Britain Report 1. London: OPCS Social Survey Division, 1995.
- 7 Paykel ES, ed. *Handbook of affective disorders*. 2nd ed. New York: Guilford Press, 1992.
- 8 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th ed, rev. Washington, DC: American Psychiatric Association, 1994.
- 9 McCormick A, Fleming D, Charlton J. *Morbidity statistics from general practice: fourth national study 1991-92*. Series MB5 No 3. London: HMSO, 1995.
- 10 Department of Health and Social Security. *Digest of statistics analyzing certificates of incapacity June 1969-May 1970*. London: HMSO, 1973.
- 11 Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977; 1: 385-401.
- 12 Goldberg DP. *The detection of psychiatric illness by questionnaire*. London: Oxford University Press, 1972.
- 13 Goldberg DP, Hillier VF. A scaled version of the general health questionnaire. *Psychol Med* 1979; 9: 139-45.
- 14 Wing JK. Use and misuse of the PSE. *Br J Psychiatry* 1983; 143: 111-17.
- 15 Wing JK, Cooper JE, Sartorius N. *Present State examination*. London: Cambridge University Press, 1974.
- 16 World Health Organization. *Composite International diagnostic interview*. Geneva: WHO, 1995.
- 17 Wadsworth MEJ. *The imprint of time: childhood, history and adult life*. Oxford: Clarendon Press, 1991.
- 18 Rodgers B, Mann SA. The reliability and validity of PSE assessment by lay interviewers: a national population survey. *Psychol Med* 1986; 16: 689-700.
- 19 Rodgers B. Behaviour and personality in childhood as predictors of adult psychiatric disorder. *J Child Psychol Psychiatry* 1990; 31: 393-414.
- 20 Rodgers B. Socio-economic status, employment and neurosis. *Soc Psychiatry Psychiatr Epidemiol* 1991;26:104-14.
- 21 Dohrenwend BP, Shrout PE, Egri G, Mendelsohn FS. Non-specific psychological distress and other dimensions of psychopathology. *Arch Gen Psychiatry* 1980; 37: 1229-36.
- 22 Goldberg DP, Williams P. *A user's guide to the general health questionnaire*. Windsor: NFER-Nelson, 1988.
- 23 Rutter M, Tizard J, Whitmore K. *Education, health and behaviour*. London: Longman, 1970.
- 24 Hirst MA. Evaluating the Malaise inventory: an item analysis. *Soc Psychiatry* 1983; 18: 181-84.
- 25 McGee R, Williams S, Silva PA. An evaluation of the Malaise inventory. *J Psychosom Res* 1986; 10: 147-52.

- 26 Wadsworth MEJ, Mann SL, Rodgers B, Kuh DL, Hilder WS, Yusuf EJ. Loss and representativeness in a 43 year follow-up of a national cohort. *J Epidemiol Community Health* 1992; **46**: 300-4.
- 27 Hsiao JK, Bartko JJ, Potter WZ. Diagnosing diagnoses: receiver operating characteristic methods and psychiatry. *Arch Gen Psychiatry* 1989; **46**: 664-67.
- 28 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**: 29-36.
- 29 Rodgers B. Pathways between parental divorce and adult depression. *J Child Psychol Psychiatry* 1994; **35**: 1289-308.
- 30 Rodgers B. Reported parental behaviour and adult affective symptoms. 1. Associations and moderating factors. *Psychol Med* 1996; **26**: 51-61.
- 31 World Health Organization. *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. Geneva: World Health Organization, 1993.
- 32 Grant G, Nolan M, Ellis N. A reappraisal of the Malaise inventory. *Soc Psychiatry Psychiat Epidemiol* 1990; **25**: 170-78.
- 33 Mackinnon A, Christensen H, Jorm AF, Henderson AS, Scott R, Korten AE. A latent trait analysis of an inventory designed to detect symptoms of anxiety and depression using an elderly community sample. *Psychol Med* 1994; **24**: 977-86.
- 34 Dryman A, Eaton WW. Affective symptoms associated with the onset of major depression in the community: findings from the US National Institute of Mental Health epidemiologic catchment area program. *Acta Psychiatrica Scandinavica* 1991; **84**: 1-5.
- 35 Eaton WW, Dryman A, Sorenson A, McCutcheon A. DSM-III major depressive disorder in the community: A latent class analysis of data from the NIMH Epidemiologic catchment area programme. *Br J Psychiatry* 1989; **155**: 48-54.
- 36 Bedford A, Foulds GA, Sheffield BF. A new personal disturbance scale (DSSI/SAD). *Br J Soc Clin Psychol* 1976; **15**: 387-94.
- 37 Goodchild ME, Duncan-Jones P. Chronicity and the general health questionnaire. *Brit J Psychiat* 1985; **146**: 55-61.
- 38 Rodgers B, Power C, Hope S. Parental divorce and adult psychological distress: evidence from a national birth cohort. In press.
- 39 Goldberg DP, Bridges K, Duncan-Jones P, Grayson D. Dimensions of neuroses seen in primary-care settings. *Psychol Med* 1987; **17**: 461-70.
- 40 Henderson JG, Pollard CA, Jacobi KA, Merkel WT. Help-seeking patterns of community residents with depressive symptoms. *J Affect Disorders* 1992; **26**: 157-62.
- 41 Maier W, Lichterman D, Oehrlein A, Fickinger M. Depression in the community: A comparison of treated and non-treated cases in two non-referred samples. *Psychopharmacol* 1992; **106** (Suppl): 79-81.
- (3) have you felt particularly low or depressed first thing in the morning?
- (4) have you had the feeling that something terrible might happen?
- (5) have you had days when your thoughts were muddled or slow?
- (6) have you had no appetite, not counting periods of physical illness?
- (7) have you been in situations, such as in a crowd or an enclosed space or meeting people, when you have become unduly anxious?
- (8) have you been in situations when you felt shaky or sweaty or your heart pounded or you could not get your breath?
- (9) have you had trouble getting off to sleep?
- (10) have you had trouble with waking up and not being able to get back to sleep?
- (11) have you been frightened or worried about becoming ill or dying?
- (12) have you felt fidgety or restless?
- (13) have you found it hard to concentrate on things or found your thoughts drifting off to other things?
- (14) have there been days when you tired out very easily?
- (15) have there been days when you found it difficult to get things done or had trouble getting started on things?
- (16) have you had the feeling that the future does not hold much for you?
- (17) have you been so caught up in your own thoughts that you neglected things?
- (18) have you seemed to lose interest in things?

Additional item concerned with suicidal ideation and acts

In the last year have you ever:-

- (i) felt that life is hardly worth living? No...0
 - (ii) thought that you really would be better off dead? No...1
 - (iii) thought about taking your own life? No...2
 - (iv) made plans to take your own life? No...3
 - (v) attempted to take your own life? No...4 /Yes...5
- (1) have you felt on edge or keyed up or mentally tense?
 - (2) have you been in low spirits or felt miserable?

Appendix

THE PSYCHIATRIC SYMPTOM FREQUENCY (PSF) SCALE

I would like to get some idea about how you have been feeling about things over the last year. How often:-