



A brief diagnostic screening instrument for mental disturbances in general medical wards

Per Fink^{a,*}, Eva Ørnbøl^a, Frits J. Huyse^b, Peter de Jonge^b, Antonio Lobo^c, Thomas Herzog^d,
Joris Slaets^e, Volker Arolt^f, Graca Cardoso^g, Marco Rigatelli^h, Morten Steen Hansen^a

^aResearch Unit for Functional Disorders, Psychosomatics and CL-Psychiatry, Aarhus University Hospital,
Noerrebrogade 44, DK-8200 Aarhus N, Denmark

^bPsychiatric CL Service, VU Medical Center, Amsterdam, The Netherlands

^cPsychosomatic and CL-Psychiatric Service, Hospital Clinico, Universitario, Zaragoza, Spain

^dDepartment of Psychotherapy and Psychosomatic Medicine, Goepingen, Germany

^eGroningen University Hospital, Groningen, The Netherlands

^fDepartment of Psychiatry and Psychotherapy, University of Muenster, Germany

^gDepartment of Psychiatry, Hospital Fernando Fonseca, Amadora, Portugal

^hDepartment of Psychiatry, University Hospital, Modena, Italy

Received 25 April 2002; accepted 24 April 2003

Abstract

Objective: Mental illness is prevalent among general hospital ward patients but often goes unrecognised. The aim of this study was to validate the SCL-8d as a brief questionnaire for mental disturbances for use in general hospitals. **Methods:** The study included 2040 patients, 18 years or older, consecutively admitted to 11 general internal medicine wards in seven European countries. All patients were screened on admission by means of the SCL-8d questionnaire. The psychometric performance (i.e., the internal validity) of the SCL-8d scale was tested using modern item response theory (IRT) in the form of the Rasch model. **Results:** Differences between sample characteristics were considerable. Even so, the SCL-8d scale showed a remarkable, statistically significant fit in terms of internal homogeneity ($P > .01$) in all individual settings, except in Spain and Germany where the item “Everything is an effort” had to be excluded to

obtain a fit. When pooling data from all centres, an excellent statistical significance of fit ($P > .05$) was obtained by exclusion of the “Effort” item. The scale was homogeneous as to gender ($P > .05$), but not age as it performed better among young patients than among patients older than 60 years ($P < .01$). In these two patient groups both internal and external homogeneity (gender, median age) was achieved. The SCL-8d sum score showed a marked correlation with current and previous treatment for mental illness. **Conclusion:** Apart from the “Effort” item ranking differently on the latent severity dimension as to age, the SCL-8d seems very robust from a psychometric point of view. Besides being short, the SCL-8d scale contains only emotional symptoms. It would therefore seem to be an excellent diagnostic tool for use in medical settings.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Mental disorders; Screening two-phase sampling; Validity; Item response theory; Rasch model; Medical patients; Depression; Anxiety

Introduction

Patients admitted to general hospital wards often suffer from mental illness and emotional distress [1–4]. Mental illness among such patients raises the pressure on health care in general and may, if untreated, prolong physical recovery, extend required sickness benefit periods and

reduce the patient’s general well-being and quality of life [5–8]. Mental illness or distress may also have a direct, negative effect on physical disease, and depression is reported to increase the risk of mortality after myocardial infarction [9–11]. Four recently published reviews on screening questionnaires for depression for use in non-psychiatric setting have reached conflicting results on whether screening may improve detection rates and patient outcome [12–15].

The recognition and possible treatment of mental disturbances in nonpsychiatric settings is therefore of

* Corresponding author. Tel.: +45-89494310; fax: +45-89494340.
E-mail address: flip@akh.aaa.dk (P. Fink).

paramount importance as is the availability of psychiatric diagnostic and screening tools that can be used by nonpsychiatric health care professionals in their everyday clinical practice in medical settings [4,14–18].

Most current interviews and questionnaires used to elicit mental disorders and emotional distress have been developed in psychiatric patient populations without sufficient evaluation and testing in medical settings. This represents a serious problem because tools developed in the former cannot always be used in the latter as physical and mental symptoms may overlap. Loss of weight and appetite, asthenia, decrease of libido and insomnia may, e.g., be attributed both to mental and medical conditions. Moreover, many current screening tools are too time consuming for regular use in medical settings or screening studies. In studies on the validity of diverse scales, the focus has been on the external validity, i.e., what the scales are measuring, and they have been tested against an external criterion as for example a psychiatric interview estimating sensitivities, specificities and positive predictive values, etc. [12–15]. By contrast reviews on the validity of the most commonly used instruments of this type do not or only briefly discuss the internal validity (i.e., that the measurements show some extent of homogeneity). This indicates that the internal validity of the most commonly used instruments of this type have only been cursorily tested [12–15,19] even though it is a precondition for undertaking an external validation that the internal validity of the instrument is warranted.

The time is therefore ripe for the introduction of a psychiatric diagnostic screening tool meeting the needs of clinical medical practice, which also fulfils the demands for a scale from a modern psychometric point of view. The eight-item dichotomised version of the symptom checklist (SCL-8d) (appendix, available from the authors) may be such a tool [20]. It is a reduced version of the SCL-25 consisting mainly of items of the depression and anxiety subscales of the SCL-90-R [21]. The number of items is reduced on the basis of a Nordic multicentre study in primary care [20], in which the external validation was checked against a standardised psychiatric Present State Examination interview (PSE).

The aim of this study was to examine the internal validity of the SCL-8d as a psychiatric diagnostic tool used by nonpsychiatric health care professionals in internal medical departments where it is applied either as a paper and pencil test or as an interview. The scale was validated by use of modern psychometric methods in terms of item response theory (IRT); a very powerful and sophisticated statistical methodology for scale analysis and construction. IRT emerged in the 1980s, replacing the classical test theory as the state of the art methodology [22–25]. In a second paper we have tested the external validity of the scale by testing its results against those of a standardised psychiatric interview [26]. A third paper will explore the usefulness of an extended scale, including a separate depression, anxiety and hostility scale.

Method

The study was undertaken by a European research group collaborating within the framework of the European Union Biomed1 program. The group's aim was to develop a short screening questionnaire or interview for use in general hospitals by nonpsychiatric health care professionals that would facilitate detection of mental illness and psychological and behavioural problems complicating medical and surgical diseases (i.e., the ARSI "Admission Risk Screening Instrument") [17,27]. The common English version of the interview was translated into different European languages by a local expert panel using existing standardised translations of subscales.

This study only reports data on the included SCL-8 questionnaire.

Setting and sample

The study was conducted from March 1996 to December 1997 in 11 general internal medicine wards in seven European countries [Denmark (1 ward), Germany (2), Hungary (2), Italy (3), the Netherlands (1), Portugal (1) and Spain (1)]. During the study period, an average period of 3 months was agreed upon for each participating ward, during which all consecutive newly admitted patients were asked to participate in the study. The following patients were excluded (Fig. 1): patients who did not give informed consent, who were younger than 18, who could not be interviewed (due to a language problem or a cognitive deficit), or who were admitted indirectly, i.e., through another hospital or ward. An attempt was made to perform the interview on the first day of the patient's admission. When an interview could not be made due to severity of

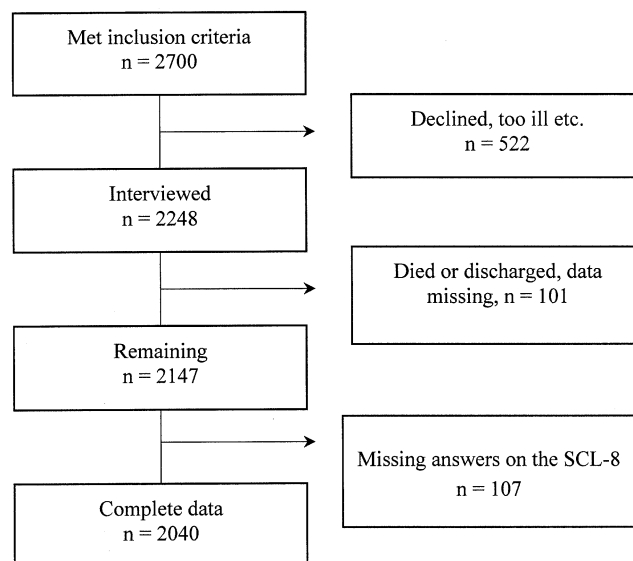


Fig. 1. Inclusion of patients.

Table 1
Background data of the total sample

Country	N	Male (%)	Retired (%)	Job (%)	Unplanned admission (%)	LOS median (Q ₁ –Q ₃)	Age mean (S.D.)
NL	130	53.8	34.6	30.0	59.2	9 (4–15)	58.5 (19.4)
SP	194	50.0	35.6	14.9	60.8	9 (5–17)	62.3 (15.4)
IT	337	53.1	59.9	23.4	74.2	9 (5–16)	62.3 (17.3)
PT	114	48.2	47.4	33.3	98.2	8.5 (6–14)	56.5 (19.2)
DK	257	56.0	45.5	31.5	83.3	4 (2–9)	58.4 (16.6)
HU	501	49.0	50.7	22.2	9.1	10 (7–14)	60.5 (13.6)
GE	507	57.2	49.5	22.3	74.2	10 (6–16)	61.1 (16.2)
Total	2040	53.2	48.6	24.0	58.5	9 (5–4)	60.5 (16.2)

LOS=length of stay; Q₁, Q₃=25%, 75% quartile.

illness, an attempt was made to conduct the interview the following day. After three failed attempts the patients were excluded from the study. Patients who died during admission were afterwards excluded from the study.

In total 2770 patients met the inclusion criteria. A total of 522 patients were excluded as they declined the invitation to participate in the study or were too ill, had language or cognitive deficits making an interview impossible, or they were excluded due to organizational problems, i.e., the patient may have been transferred to another hospital or department or discharged before the interview. In total 2248 patients were interviewed, but 101 died during admission or discharge data were missing. Thus, data for 2147 patients remain, and only patients who answered all eight SCL questions were included in this study, 2040 in total (Fig. 1). The smallest of the 11 subsamples (Portugal) included 114 patients, the largest (Lübeck, Germany) 507 patients (see Table 1).

Procedures and assessment

The patients were interviewed by a health care professional (i.e., a nurse, medical student, doctor, medical officer, etc.) as soon as possible after admission. All interviewers had previously participated in brief tutorials to train the use of the highly structured interview, which included among others the SCL-8 questionnaire. At the interview the questionnaire was filled in by the interviewer or by the patient, according to the latter's preference. Each

item of the SCL-8 had four response categories ranging from “not at all” to “severe”.

In addition, the treating consultant and caring nurse were asked different questions about the patients' condition.

Item response theory

IRT is a theory on how to model responses to a set of items or an instrument [28–31]. It is assumed that responses to particular items/instruments all reflect continuous variables or dimensions (here named latent continua) that cannot be observed or measured directly. The majority of models are concerned with only one latent continuum. In that case the relation of a binary response of each item to the latent continuum is modelled by an S-shaped curve called an item-characteristic curve (ICC). Each item response is characterised by a threshold figure and a slope. The threshold is the point on the latent continuum at which there is a 50% probability that the item will elicit a positive response. Persons whose latent continuum values lie above the threshold are more likely to display the symptom (i.e., elicit a positive response to the corresponding item) than persons whose values lie below the threshold.

The slope shows how well the item separates individuals on the latent continuum. Some symptom items are linked to specific diagnoses, while others may occur in several unrelated illnesses, or even among normal individuals. Items of specific diagnoses are likely to have a steeper slope than items of the two latter. The ICC may be modelled

Table 2
The SCL-8d scale among European internal medical inpatients (N=2040)

Name	Positive response (%) N=2040	Two-parameter model one dimensional ^a		One-parameter Rasch model ^b	
		Factor load	Slope	Item parameter	S.E.
1 Feeling blue	Depressed 34.9	0.82	1.42	–0.98	0.06
2 Feeling everything is an effort	Effort 34.6	0.68	0.93	–0.96	0.05
3 Nervousness or shakiness inside	Nervous 33.9	0.73	1.07	–0.90	0.04
4 Worrying too much about things	Worrying 32.0	0.79	1.28	–0.76	0.06
5 Feeling hopeless about the future	Hopeless 23.9	0.84	1.55	–0.12	0.06
6 Feeling worthless	Worthless 15.4	0.80	1.33	0.72	0.07
7 Feeling fearful	Fearful 15.0	0.75	1.14	0.76	0.07
8 Spells of terror or panic	Panic 6.1	0.68	0.93	2.24	0.10

^a By means of NOHARM.

^b By means of LPC-m.

Table 3
Test of the internal and external homogeneity for the one-parameter Rasch model

	All centres combined					Age group ≤ 60		Age group >60	
	Internal ^a		External			Internal ^a	External Age ^c	Internal ^a	External Age ^d
	All items	omitting Item 2	Gender	Country	Age ^b				
χ^2	30.2	3.7	13.9	224.9	49.6	13.0	13.0	17.4	15.8
df	7	6	7	42	7	7	7	7	7
P value	<.01	.717	.053	<.001	<.001	.072	.049	.015	.027
Cronbach's alpha	0.80	0.78				0.79		0.80	

^a Split at score = 1.
^b Split at median age = 60.
^c Split at median age = 48.
^d Split at median age = 73.

in numerous ways, most often by a logistic or a normal ogive function. We apply two models for binary data, a one-parameter model [32] and a two-parameter model [33,34]. In the two-parameter model both the slope as well as the threshold may be different for each item. In the one-parameter model (the Rasch model) the slopes of the ICC are assumed equal for all items, and they are only characterised by the threshold parameter, named the item parameter. If the Rasch model fits the data, several desirable properties follow, e.g., the sum of positive responses is a sufficient statistics of a person's position on the latent continuum, and there are no item bias or differential item functioning, meaning that the item parameters display homogeneity [23,35]. The Rasch model is also a powerful instrument to solve problems of linking and equating especially educational tests [35].

Statistical methods

In the statistical analysis the responses were dichotomised so that “not at all” and “mild” categories combined were characterised as negative responses, and “moderate” and “severe” combined were characterised as positive responses.

To assess the fit of a Rasch model to the dichotomised responses of SCL-8, we proceeded as follows: Firstly, an explorative factor analysis was performed to investigate the unidimensionality of the SCL-8d. The number of latent continua was inspected by means of factor loadings of items and change in root mean square of residuals in models with one to four factors, i.e., latent continua. Secondly, the slope estimates in a two-parameter model in one latent continuum were computed to evaluate if the use of the simpler one-parameter model was acceptable. Thirdly, a range of conditional likelihood ratio tests were used to test item homogeneity, i.e., to test if the Rasch model item parameters were the same in two subdivisions of the sample vs. the alternate that separate item parameters applied in each subdivision.

The item homogeneity consists of internal and external homogeneity, where the internal uses a subdivision of the sample according to the sum score of the positive responses,

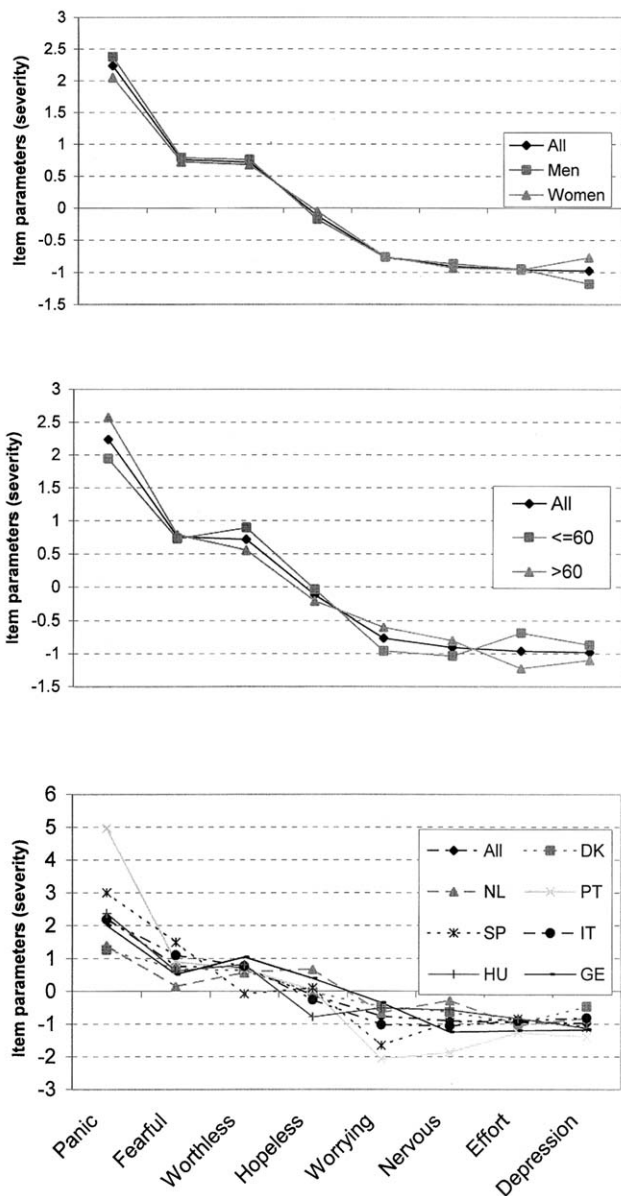


Fig. 2. Rasch model results.

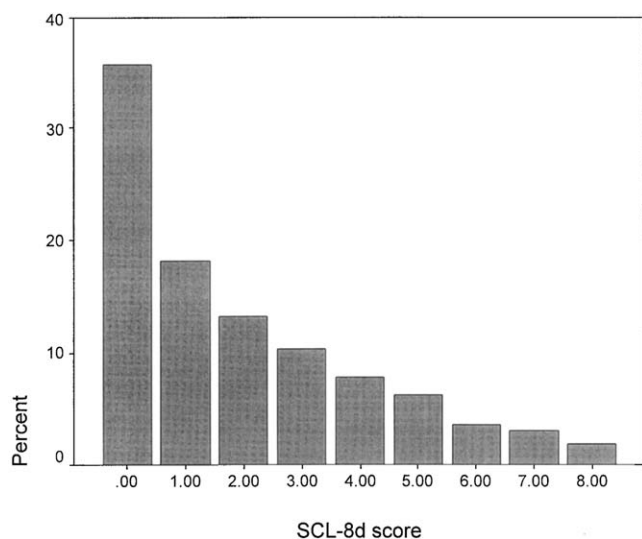


Fig. 3. SCL-8d score distribution in 11 wards ($N=2040$).

often score = 1. External homogeneity uses external variables to divide the sample, e.g., gender, median age and country.

In addition, the Cronbach's alpha and the association between the SCL-8d and treatment of mental illness was computed.

The data were processed by means of the SPSS Windows release 10.0 [36], STATA [37], NOHARM [38], and LPC-m [39].

Results

Patient characteristics differed markedly among included internal medicine wards in the seven countries (Table 1). This may be a reflection of the differences in the practice of medicine and in the health care systems of the included countries.

For example, the Danish sample showed an average stay of 4.0 bed days, which is equivalent to the Danish average for hospital stays, whereas the average was around 9 days in the other countries. In the Portuguese sample 98.2% of the admissions were unplanned compared with only 9% of the

admissions in the Hungarian sample. The explorative factor analysis revealed that one factor was appropriate.

Table 2 shows the individual SCL-8d symptoms (i.e., items), the positive response frequencies, the factor loadings of a one-factor model and slope estimates from the two-parameter model, and the item parameters and their standard errors from a one-parameter Rasch model. In Table 2 it appears that none of the slope estimates differ markedly from the rest. The one-parameter Rasch model is therefore used in the analysis.

The test of internal and external homogeneity for the overall sample is displayed in Table 3. The internal homogeneity ($P = .717$) was reached when Item 2 "Everything is an effort" was excluded. The external homogeneity was obtained for gender ($P = .059$), but not for age and country ($P < .001$). The same pattern was seen when Item 2 "Everything is an effort" was excluded (results not shown). A division of the overall sample by median age (median = 60 years) led to internal homogeneity ($\leq 60 P = .072$; $>60 P = .015$) in the two subsamples and to external homogeneity for age ($\leq 60 P = .049$; $>60 P = .027$), (Table 3). External homogeneity was obtained for gender ($\leq 60 P = .196$; $>60 P = .073$), but not for country ($\leq 60 P < .001$; $>60 P < .001$). The Rasch model results for total population, gender, age groups and for separate countries are illustrated in Fig. 2, where the item parameter for each item is plotted for a variety of samples. The items are displayed according to overall ranking.

The items were ranked in a nice, stepwise order as to the latent severity dimension. The difference between old and young may be ascribed to the different ranking in the latent severity dimension of Items 2 "Everything is an effort" and 6 "Feeling of worthlessness".

Country profiles clustered around the overall profile as illustrated in Fig. 2, although the rank order of symptoms was slightly more skewed for Spain and Portugal. A test for internal homogeneity resulted in the following P values for the participating countries: DK $P = .017$, NL $P = .501$, PT $P = .282$, IT $P = .221$, HU $P = .068$, SP $P = .005$ (without Item 2 $P = .096$), GE $P = .005$ (without Item 2 $P = .042$).

The score distribution on the SCL-8d is displayed in Fig. 3. The same pattern is found when looking at the score distribution per country.

Table 4.
Association between SCL-8d score and treatment for mental illness

		SCL-8d score			Gamma ^a	χ^2 Test
		0%	1–2%	3–8%		
Outpatient treatment ever for mental illness/problems	Yes ($n = 312$)	15.7	26.9	57.4	0.5	$P < .001$
	No ($n = 1721$)	39.3	32.4	28.3		
Admitted to mental health care facilities ever	Yes ($n = 120$)	14.2	25.8	60	0.5	$P < .001$
	No ($n = 1909$)	37.0	31.9	31.1		
Currently being treated by a mental health professional	Yes ($n = 71$)	14.1	19.7	66.2	0.6	$P < .001$
	No ($n = 1960$)	36.3	31.9	31.5		

^a Goodman–Kruskal gamma.

Patients previously tested for mental illness in psychiatric health care facilities scored significantly higher than other patients (Table 4). Furthermore, patients undergoing treatment for mental disturbances during the interview period virtually all had a positive and high SCL-8d score (20% 1–2; 66% 3 or higher).

Discussion

The use of scales builds on the assumption that a person's position on a latent continuum (i.e., depression or anxiety) can be indirectly inferred from that individual's response to a set of well-chosen items and that when confronted with other items from the same domain his/her behaviour can be predicted from this position [40,41]. The major strength of the present study is that it investigated the use of the SCL-8d scale in internal medical settings, applying IRT in the form of the Rasch model [40] testing the above assumption [41]. Modern IRT outperforms traditional methods, such as classical test theory, by not requiring the inclusion of several nearly identical questions (items) to adjust for random error in the measurements. Furthermore, the number of items may hence be reduced by excluding items that, from a mathematical point of view, are almost identical, i.e., they are located in the same position of the (latent) severity dimension, even if they may have completely different wordings. The IRT methodology is thus quite suitable in constructing short scales.

The Rasch model used in this study is the simplest, but also the most restrictive model in the family of modern IRT. Except from a few scales on depression [42] only very few currently used scales concerning mental distress have been tested by means of IRT, and few of them are likely to fit a Rasch model. Strictly speaking a comparison to other instruments is therefore not possible. This may be a somewhat restrictive position, as we cannot deny that other less refined scales may also serve this purpose well. However, if a scale meets the requirements of the Rasch model, a trustworthy scoring system is guaranteed.

The samples studied were remarkably different owing mainly to the fact that they came from 11 different internal medicine wards from seven different European countries. From a psychometrical point of view the heterogeneous sample is a strength of the study, and it is a strong and impressive support for the validity of the SCL-8d scale that it had a statistically significant fit concerning internal homogeneity in all separate settings (i.e., $P > .01\%$), except in Spain and Germany where the "Effort" item had to be excluded to obtain a significant fit. The item "Everything is an effort" also had to be excluded in the test of internal homogeneity in the overall sample to obtain a significant fit. This may be so because this item is a common symptom even among nondistressed elderly individuals as age sets a natural limit to physical stamina. In the younger age group the overall model fits well, even

when the "effort" item is included. We therefore do not recommend that this item is excluded entirely from the scale, but its inclusion probably introduces a bias towards elderly individuals. A weakness of the study is that we did not further explore why the "Effort" item also caused problems in the Spanish and German samples, i.e., whether it was due to problems with the translation, cultural differences or other factors. However, as the "Effort" item fits well in the other centres, we recommend to include it in the scale.

We may conclude that although the patient populations were very heterogeneous and the SCL-8d was administered both as an interview and as a paper and pencil test, it did indeed prove to be highly robust when applied across disparate populations. A high measure of generalisability, which is one of the most important aspects of validity, would hence seem to be guaranteed, despite the problems with the "Effort" item. We did not explore whether it made any difference if the questions were read aloud by the interviewer, or if the patients filled in the questionnaires themselves during the interview.

Depression, anxiety and other mental disorders and disturbances share many symptoms with well-defined physical diseases. It is therefore of paramount importance to minimize the number of false positive results due to symptom overlap [43]. The SCL-8d scale enjoys the great advantage over other scales used in medical setting that it tests solely for emotional symptoms that are not a part of the symptomatology of physical diseases. The symptoms "effort" and "hopelessness" may, however, be explained otherwise than by mental distress. "Hopelessness" may, for instance, be a natural reaction to a diagnosis of a physical disease carrying a poor prognosis. However, both items belong to the distress dimension, though "effort" did not have as high an affinity with the dimension as the other items of the scale in the older age group as well as in some countries.

Only few studies have compared different screening questionnaires for psychiatric distress, of which the SCL and the GHQ are the most common. The two instruments were compared in a general population sample, where they performed equally well although the SCL was better at detecting long-standing disorders and the GHQ detected fewer false positives [44]. The wording of the SCL questions facilitates a psychiatric interpretation as the symptoms asked for are found in the diagnostic criteria for both anxiety and depressive disorders. The GHQ questions are more general and therefore less transparent. A review and comparison of the many different scales suggested as screening tools for mental distress is far beyond the scope of this paper. However, a cursory review of the literature, including recent reviews on scales for depression [12–15], did not point out any particular instruments as outstanding compared to others. Furthermore, it shows that only very few have been tested using modern psychometric methods, and until other instruments have been tested in the same rigorous way, it cannot be concluded whether they are better or worse than the SCL-8d.

A primary care multicentre study [20] produced results similar to those presented here, i.e., internal homogeneity was observed for each participating centres and external homogeneity was observed for gender, but there were age differences and problems with transferability between the centres. The transferability issue demonstrates that from a narrow statistical viewpoint the Rasch analysis does not allow direct comparison of sum scores between centres. In the present study, seven countries were involved and external homogeneity will be lost if only one of these countries deviates from the overall sample estimates; something that could hardly be avoided as some of the countries had rather few participating patients, e.g., 114 in Portugal.

The individual items of the SCL were dichotomised, which implies that not all information is used. On the other hand, this makes the scale much simpler to use. It is likely that the scale could be improved if multiple responses were included, but testing this awaits the development of more robust statistical models.

The SCL-8d score was clearly associated with current treatment for mental illness, previous admission to psychiatric wards or outpatient psychiatric treatment. A second paper based on the Danish subsample of the present study and a sample of new neurological patients analysed the external validity of the SCL-8d using ICD-10 diagnoses made by means of the SCAN interview (Schedules for Clinical Assessment in Neuropsychiatry) [45] as gold standard. We found a good performance measured by sensitivity and specificity of the SCL-8d [46].

Satisfactory sensitivity and specificity of the SCL-8d as to mental disorders according to a standardised psychiatric interview have also been found in a primary care study [20].

Other studies using the SCL-8d revealed that the SCL-8d score is strongly associated with health care utilization prior to hospital admission [46,47]. It also has some predictive power as to health care utilization and to self-rated health after discharge from hospital [48], even if adjusting for severity of physical disease. Likewise, in primary care it could predict frequent attendance [47]. Finally, a high score on the SCL-8d has been shown to be associated with persistence of musculoskeletal problems in primary care patients [49,50].

A weakness of the SCL-8d is that it does not specify the nature of the mental problem. A positive test would thus have to be combined with a clinical assessment or a structured interview, as for example the COMPRI [51], to make psychiatric diagnoses [18,52,53]. A later paper will explore the usefulness of an extended scale that includes separate depression and anxiety subscales. These scales were not tested in the European setting as they were only included in the Danish subsample. However, a separate anxiety and depression scale does not necessarily mean that the SCL-8d is superfluous, as in some cases a very brief instrument is needed and/or the split into the two dimensions is not needed or even undesirable.

Another weakness of the study is that we did not explore the impact of the screening instrument on detection rate, treatment and outcome of patients. The SCL-8d is currently being tested as to these aspects in a randomised controlled study in primary care in Denmark.

Acknowledgments

Dr. Luigi Grassi, MD, Institute of Psychiatry, University of Ferrara, Italy, Dr. Costanzo Gala, MD, Institute of Psychiatry, University of Milan, Italy, Dr. N. Balogh, MD, Institute of Cardiology, Budapest, Hungary have participated planning the study and collecting data. The study has been supported by grants from: The European Union Biomed1 (Grant BMH1-CT93-1180), The Danish Medical Research Council (Grant 9601898), The Danish Health Insurance Fund, The Hede Nielsen Foundation, the fund “Puljen til Styrkelse af Psykiatrisk Forskning”.

References

- [1] Arolt V, Driessen M, Dilling H. The Lübeck General Hospital Study. I: prevalence of psychiatric disorders in medical and surgical inpatients. *Int J Psychiatry Clin Pract* 1997;1:207–16.
- [2] Mayou R, Hawton K. Psychiatric disorder in the general hospital. *Br J Psychiatry* 1986;149:172–90.
- [3] Silverstone PH. Prevalence of psychiatric disorders in medical inpatients. *J Nerv Ment Dis* 1996;184(1):43–51.
- [4] Hansen MS, Fink P, Frydenberg M, Oxhøj M, Sondergaard L, Munk-Jørgensen P. Mental disorders among internal medical inpatients: prevalence, detection, and treatment status. *J Psychosom Res* 2001;50(4):199–204.
- [5] Fink P. Admissions of persons aged 17–49 years to nonpsychiatric departments. *Ugeskr lAeger* 1989;151:307–10.
- [6] Fink P. Mental illness and admission to general hospitals: a register investigation. *Acta Psychiatr Scand* 1990;82:458–62.
- [7] Fink P. Admission patterns of persistent somatization patients. *Gen Hosp Psychiatry* 1993;15:211–8.
- [8] Hansen MS, Fink P, Frydenberg M. Follow-up on mental illness in medical inpatients: health care use, self-rated health, and self-rated physical fitness (submitted for publication).
- [9] Frasure-Smith N, Lesperance F, Talajic M. Depression following myocardial infarction. Impact on 6-month survival [published erratum appears in JAMA 1994 Apr 13;271(14):1082] *JAMA* 1993;20;270(15):1819–25.
- [10] Musselman DL, Evans DL, Nemeroff CB. The relationship of depression to cardiovascular disease: epidemiology, biology, and treatment. *Arch Gen Psychiatry* 1998;55(7):580–92.
- [11] Rozanski A, Blumenthal JA, Kaplan J. Impact of psychological factors on the pathogenesis of cardiovascular disease and implications for therapy. *Circulation* 1999;99(16):2192–217.
- [12] Williams JW, Pignone M, Ramirez G, Stellato CP. Identifying depression in primary care: a literature synthesis of case-finding instruments. *Gen Hosp Psychiatry* 2002;24:225–37.
- [13] Gilbody SM, O’House A, Sheldon TA. Routinely administered questionnaires for depression and anxiety: systematic review. *BMJ* 2001;322:406–9.
- [14] Meakin CJ. Screening for depression in the medically ill—the future of paper and pencil tests. *Br J Psychiatry* 1992;160:212–6.
- [15] Hickie IB, Davenport TA, Ricci CS. Screening for depression in

- general practice and related medical settings. *Med J Aust* 2002; 177:S111–6 (supplement).
- [16] de Jonge P, Huyse FJ, Slaets JP, Herzog T, Lobo A, Lyons JS, Opmeer BC, Stein B, Arolt V, Balogh N, Cardoso G, Fink P, Rigatelli M, van Dijk R, Mellenbergh GJ. Care complexity in the general hospital: results from a European study. *Psychosomatics* 2001;42(3):204–12.
- [17] de Jonge P, Huyse F, Herzog T, Lobo A, Slaets JPJ, Lyons JS, Opmeer BC, Stein B, Arolt V, Balogh N, Cardoso G, Fink P. Risk factors for complex care needs in general medical inpatients: results from a European study. *Psychosomatics* 2001;42(3):213–21.
- [18] Huyse FJ, de Jonge P, Slaets JPJ, Herzog T, Lobo A, Lyons JS, Opmeer BC, Stein B, Arolt V, Balogh N, Cardoso G, Fink P. Compri—an instrument to detect patients with complex care needs: results from a European study. *Psychosomatics* 2001;42(3):222–8.
- [19] Hermann C. International experiences with the hospital anxiety and depression scale—a review of validation data and clinical results. *J Psychosom Res* 2003;42(1):17–41.
- [20] Fink P, Jensen J, Borgquist L, Brevik JI, Dalgard OS, Sandager I, Engberg M, Hansson L, Holm M, Nordström G, Stefansson CG, Sørensen L, Munk-Jørgensen P. Psychiatric morbidity in primary public health care. A Nordic multicenter investigation: Part I. method and prevalence of psychiatric morbidity. *Acta Psychiatr Scand* 1995; 92:409–18.
- [21] Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale—preliminary report. *Psychopharmacol Bull* 1973; 9(1): 13–28.
- [22] Fischer GH, Molenaar IW. Rasch models—foundations, recent developments, and applications. New York: Springer-Verlag, 1995.
- [23] van der Linden WJ, Hambleton RK. Handbook of modern item response theory. New York: Springer-Verlag, 1997.
- [24] Gibbons RD, Clark DC, VonAmmon CS, Davis JM. Application of modern psychometric theory in psychiatric research. *J Psychiatr Res* 1985;19(1):43–55.
- [25] Lohr KN. Health outcomes methodology symposium: summary and recommendations. *Med Care* 2000;38(Suppl. 9):II194–208.
- [26] Fink P, Hansen MS, Ørnboel E, Søndergaard L, De Jonge P. Detecting mental disorders in general hospitals by the brief SCL-8 scale. *J Psychosom Res* 2004;56(3):371–5.
- [27] Huyse FJ, Herzog T, Malt UF, Lobo A. The European Consultation-Liaison Workgroup (ECLW) collaborative study. I. General outline. *Gen Hosp Psychiatry* 2001;18:44–55.
- [28] Andersen EB. The statistical analysis of categorical data. Berlin: Springer-Verlag, 1990.
- [29] Duncan-Jones P, Grayson DA, Moran PAP. The utility of latent trait models in psychiatric epidemiology. *Psychol Med* 1986;16:391–405.
- [30] Goldberg DP, Bridges K, Duncan-Jones P, Grayson DA. Dimensions of neurosis seen in primary-care settings. *Psychological Med* 1987; 17:461–70.
- [31] McDonald RP. Linear versus nonlinear models in item response theory. *Appl Psychol Meas* 1982;4:379–96.
- [32] Rasch G. Probabilistic models for some intelligence and attainment tests. Studies in mathematical psychology I. Copenhagen: Pedagogic Institute in Denmark, 1960.
- [33] Birnbaum A. Some latent trait models and their use in inferring an examinee's ability. In: Lord FM, Novick MR, editors. Statistical theories of mental test scores. Reading (MA): Addison-Wesley, 1968. pp. 397–479.
- [34] Baker FB. Item response theory: parameter estimation techniques. New York: Marcel Dekker, 1992.
- [35] Gustafsson J-E. Testing and obtaining fit of data to the Rasch model. *Br J Math Stat Psychol* 1980;33:205–33.
- [36] SPSS 10.0 for Windows. SPSS Inc., 2000.
- [37] StataCorp J-E. Stata Statistical Software: Release 5.0. College Station: Stata, 1997.
- [38] Dunner DL. The issue of comorbidity in the treatment of panic. *Int Clin Psychopharmacol* 1998;13(Suppl. 4):19–24.
- [39] Fischer GH, Ponocny-Seliger E. Structural Rasch modeling: handbook of the usage of LPCM-WIN 1.0. Vienna: Dept. of Psychology, University of Vienna, 2001.
- [40] Fischer GH, Molenaar IW. Rasch models—foundations, recent developments, and applications. New York: Springer-Verlag, 1995.
- [41] van der Linden WJ, Hambleton RK. Handbook of modern item response theory. New York: Springer-Verlag, 1997.
- [42] Olsen LR, Jensen DV, Noerholm V, Martiny K, Bech P. The internal and external validity of the Major Depression Inventory in measuring severity of depressive states. *Psychol Med* 2003;33(2):351–6.
- [43] Kaplan CP, Miner ME, Mervis L, Newton H, McGregor JM, Goodman JH. Interpretive risks: the use of the Hopkins Symptom Checklist 90-revised (SCL 90-R) with brain tumour patients. *Brain Inj* 1998; 12(3):199–205.
- [44] Goldberg DP, Rickels K, Downing R, Hesbacher P. A comparison of two psychiatric screening tests. *Br J Psychiatry* 1976;129:61–7.
- [45] WHO P. SCAN. Schedules for Clinical Assessment in Neuropsychiatry, version 2.1. Geneva: World Health Organization, Division of Mental Health, 1998.
- [46] Hansen MS, Fink P, Frydenberg M, Oxhøj ML. Use of health services, mental illness, and self-rated disability and health in medical inpatients. *Psychosom Med* 2002;64(4):668–75.
- [47] Vedsted P, Fink P, Olesen F, Munk-Jørgensen P. Psychological distress as a predictor of frequent attendance in family practice: a cohort study. *Psychosomatics* 2001;42(5):416–22.
- [48] Hansen MS, Fink P, Frydenberg M, de Jonge P, Huyse FJ. Complexity of care and mental illness in medical patients. *Gen Hosp Psychiatry* 2001;23:319–25.
- [49] Jørgensen KC, Fink P, Olesen F. Psychological distress among patients with musculoskeletal illness in general practice. *Psychosomatics* 2000;41(4):321–9.
- [50] Jørgensen KC, Fink P, Olesen F. Psychological distress and somatisation as prognostic factors in patients with musculoskeletal illness in general practice. *Br J Gen Pract* 2000;50:537–41.
- [51] Huyse FJ, Lyons JS, Stiefel FC, Slaets JP, de Jonge P, Fink P, Gans RO, Guex P, Herzog T, Lobo A, Smith GC, van Schijndel RS. “INTERMED”: a method to assess health service needs: I. Development and reliability. *Gen Hosp Psychiatry* 1999;21(1):39–48.
- [52] Stiefel FC, de Jonge P, Huyse FJ, Guex P, Slaets JP, Lyons JS, et al. “INTERMED”: a method to assess health service needs: II. Results on its validity and clinical use. *Gen Hosp Psychiatry* 1999;21(1):49–56.
- [53] Huyse FJ, Lyons JS, Slaets J, de Jonge P, Latour C. Operationalizing the biopsychosocial model: the intermed. *Psychosomatics* 2001; 42:5–13.