THE VALIDITY OF GENERAL HEALTH QUESTION-NAIRES, GHQ-12 AND GHQ-28, IN MENTAL HEALTH STUDIES OF WORKING PEOPLE

ZOFIA MAKOWSKA, DOROTA MERECZ, AGNIESZKA MOŚCICKA and WOJCIECH KOLASA

Department of Work Psychology
Nofer Institute of Occupational Medicine
Łódź, Poland

Abstract. The purpose of the study was to determine such cut-off points in the scores of General Health Questionnaires (GHQ-12 and GHQ-28) that allow for optimal identification of people with mental health disorders in the Polish working population attending primary health care settings. The groups under the study covered 419 and 392 patients for GHQ-12, and GHQ-28, respectively. In the GHQ-12 group, 90 and in the GHQ-28 group, 80 subjects filled in the questionnaires and agreed to participate in the second stage of the study – a psychiatric interview. The criterion validity of the GHQs was a mental health diagnosis, based on the Munich version of Composite International Diagnostic Interview. The complete computerized version of interview, covering all diagnostic sections, has been adopted. In the mental health diagnosis only disorders, which currently troubled patients were taken into consideration and disorders which created problems in the distant past were excluded. In the group covered by GHQ-12 examination, 55.6% of persons had at least one type of mental disorder diagnosed, based on the criteria of both Diagnostic and Statistical Manual for Mental Disorders (DSM-IV) and the International Classification of Diseases (ICD-10). In the GHQ-28 group, the percentage of persons with mental disorders was 47.5%. After excluding patients with nicotine dependence disorder only, the frequency of mental health problems decreased to 45.5% and 33.8%, respectively.
The proposed cut-off points, 2/3 points for GHQ-12 and 5/6 points for GHQ-28, were established at the level of the highest possible sensitivity and specificity not lower than 75%. These principals have been accepted for a practical reason, as the acceptance of the lower level of specificity forces medical practitioners to devote too much time to practically healthy people.
At the above mentioned cut-off points for GHQ-12 sensitivity is 64% and specificity – 79%, while for GHQ-28 the values are 59% and 75%, respectively. These validity coefficients were calculated from distributions of groups, from which persons with nicotine dependence as the only disorder were excluded. Incorporation of these people in the whole sample reduced the questionnaires' validity. Modification of responses scoring from the standard one – GHQ to CGHQ has not improved the validity of questionnaires.
Lower validity coefficients of GHQ-28, in comparison to GHQ-12 validity are the effect of greater influence of somatic disease on the results acquired in this scale version of the questionnaire.

Key words:
D. Goldberg's General Health Questionnaires, Criterion validity, Working people, Mental health diagnosis, Composite International Diagnostic Interview

INTRODUCTION

Over the last few years, a systematic increase in the prevalence of mental disorders in various groups of population has been observed [1]. The reports published by WHO suggest that individual and social costs resulted from significant deterioration of mental health of the whole population are comparable (or in some mental disorders even higher) with costs of somatic diseases. Therefore, the significance of early detection of mental health disorders

must be emphasized in view of increasing opportunity to benefit from professional intervention, which can diminish the risk of the development of serious disorders leading to long-term work disability [2].

Early detection of mental health disorders is understood as mental health monitoring in a given population at the primary medical care level.

Standardized screening questionnaires can be used for this purpose. They will allow to detect from among entirety of patients, in a short period of time, those persons whose psychological state presages mental health disorders.

Such studies have not yet been carried out in Poland in spite of the fact that a significant part of the population has mental problems. For example, the research carried out in 1996 by the Central Statistical Office (GUS) proved that 24.7% of women and 18.2% of men, aged above 15 years complain of psychological discomfort. In addition, neurosis was diagnosed in 14.9% of women and 7.1% of men [3]. The research did not cover information concerning other mental disorders. Health state indicators of the population in 1992–1995 provide evidence that the mental status of the Polish population deteriorated in that period [4]. The results of analyses showed both an increase in the number of outpatients treated in psychiatric consulting settings and the number of patients treated in stationary psychiatric wards.

An example of such questionnaires, which comply with conditions laid down for the screening tools of implementing the above mentioned aims is a family of David Goldberg's General Health Questionnaires – GHQ. A high validity of the GHQs has been confirmed by the results of many studies carried out in the 1980s [5] and the1990s [6]. The qualities of these questionnaires were a decisive argument in preparing the Polish adaptation of two of them: GHQ-12 and GHQ-28.

The purpose of the presented studies is to evaluate validity coefficients for both questionnaires and to find the best threshold score to achieve optimum sensitivity and specificity, which is necessary to identify potential "cases" in Polish working population. The routine monitoring of mental health in this population seems to be very important since mental health disorders are thought to be one of the ten most common diseases and complains related to work [7].

MATERIALS AND METHODS

Research methods

A two-stage design of the validity study has been applied. Stage I was an examination aided with GHQ-12 or GHQ-28. Stage II was a standardized psychiatric interview, used as a criterion for GHQ score.

General Health Questionnaires, GHQ-12 and GHQ-28

The fundamental assumption of the GHQ construction is conviction that there are many traits and characteristics which discriminate psychiatric patients as a class from individuals considered themselves to be healthy. The questionnaire concentrates on two fundamental groups of problems: inability to carry out one's normal "healthy" functions and the appearance of new phenomena of a distressing nature. It focuses on break in normal functioning and not on permanent traits. The main version of the questionnaire contains 60 questions; shorter versions are derived from the main one, e.g. 12-item GHQ-12 and 28-item – GHQ-28 (the latter is also called Scaled GHQ). GHQ-28 consists of four 7-item scales: somatic symptoms, anxiety and insomnia, social dysfunction and severe depression. It allows for mental health assessment on four dimentions corresponding with these four scales. The structure of all questions is always the same, regardless of a type of version. The patient is asked to assess changes in his/her mood, feelings and behaviors in the period of recent four weeks. The patient evaluates their occurrence on a 4-point response scale. The scale points are described as follow: "less than usual", "no more than usual", "rather more than usual", "much more than usual". The standard scoring method recommended by Goldberg for the need of case identification is called "GHQ method". Scores for the first two types of answers are "0" (positive) and for the two others – "1" (negative). Sometimes a modified scoring method, called Goodchild and Duncan-Jones's method (CGHQ), is suggested, since it is possible to omit long-standing disorders in the classical version of GHQ scoring [5].

In CGHQ scoring of the answer "as always" depends on the consideration: does the positive answer to a question indicates illness (so called negative items of the questionnaire) or health (positive items). Among negative items,

e.g. "feeling unhappy and depressed", a respondent gets 1 point for an answer "no more than usual". Therefore, the scoring of answers for such questions is: 0, 1, 1, 1. In positive items, such as "been able to concentrate on whatever you're doing" the reply "same as usual" has a 0 value. The scoring of positive items is then the same as the standard scoring in the GHQ (0, 0, 1, 1). Both first and second method of scoring give the same theoretical range of global score (for GHQ-12 from 0 to 12 points, and for GHQ-28 from 0 to 28 points).

In the analyses carried out by us, both methods of response scoring were taken into consideration.

Psychiatric interview

The validity criterion for GHQ-12 and GHQ-28 was a clinical diagnosis acquired on the basis of a standardized psychiatric interview – the Munich version of the Composite International Diagnostic Interview (CIDI). Polish translation of M-CIDI has been made in the Department and Clinic of Psychiatry of the Medical Academy in Wrocław.

The CIDI was developed in the context of a joint project between the World Health Organisation and the US Alcohol, Drug Abuse and Mental Health Administration (ADAMHA), registered in 1979. While preparing the questionnaire several stages of work were accomplished in the 1980s [8,9]. The CIDI was prepared to be applied in different cultures under the framework of many research projects concerning various issues. It has been worked out mainly from the point of view of epidemiological studies of psychopathology in a general population. It is a comprehensive, highly structured and totally standardized interview, allowing to diagnose mental disorders in accordance with definitions and criteria of ICD-10 [10] and DSM-IV [11].

The interview questions cover 12 diagnostic sections listed below, dedicated to individual groups of disorders (sequence of the sections – as in the interview):

- Nicotine dependence
- Somatization and dissociative disorders
- Anxiety disorders
- Depression and dysthymic disorder
- Mania and affective, bipolar disorder

- Schizophrenia and other psychotic disorders
- Eating disorders
- Alcohol related disorders
- Obsessive-compulsive disorder
- Disorders related to psychoactive substances abuse
- Mental deterioration (dementia, memory disorder and other cognitive disorders)
- Posttraumatic stress disorders.

The CIDI prepared also in a computerized version was applied in our research. The total standardization of the instrument enables us to administer it and diagnose disorders by nonclinicians who underwent an appropriate training in an authorized institution.

The studies of GHQs validity made in the 1980s were based on two diagnostic criteria: Present State Examination (PSE) and Clinical Interview Schedule (CIS) [5], whereas the studies from the 1990s applied the CIDI as a basis for mental disorder diagnoses [6,12,13].

Research procedure

The research was carried out among patients attended by general practitioners and in co-operation with these professionals, in the primary care settings of a large provincial city and in two occupational medicine centres of local factories. The examination covered only persons employed at that time, reporting to a general practitioner with no acute complains and/or fever. Whenever a person met the above criteria, a physician asked her/him to participate in the study and, when the request was accepted, the patient got either one of the questionnaires, GHQ-12 or GHQ-28 (it depended on the address of the research venue and the setting). The patients also received a written information, containing the description of the study purpose, the method of results exploitation and participation principles (voluntary character, anonymity). Patients filled in the GHQ and immediately afterwards reported to the person carrying out the psychiatric interview (M-CIDI). The subject was interviewed by a psychologist employed in the Nofer Institute of Occupational Medicine. A request to take part in an interview was directed to each person who turned up with a questionnaire whenever a psychologist was not engaged with another patient at that time. If a respondent appeared

during an interview with another person, only the questionnaire was collected. A principle was accepted that a patient did not wait for an interview, which was, in fact, very time-consuming lasting more or less one hour.

The applied research procedure can be defined as a pseudo-random one; an interview, a second part of the study, was held whenever a psychologist was not busy with another patient. It was imposed by the necessity to carry out both stages of the study immediately one after the other. While characterizing positive and negative aspects of various procedures of the research on the GHQ validity, Goldberg and Wiliams [5] stated that the advantage of a procedure based on random trial (or pseudo-random trial) is that these procedures allow for most precise estimation of sensitivity and specificity, and in consequence for establishing the best threshold score.

Study population

The validity study of GHQ-12 was carried out in a group of 90 persons. It was a sample of 419 patients who were willing to answer the questions included in the questionnaire and, at the same time take part in the second stage of the study. Table 1 summarizes the characteristics of this group. Over 50% of the group were women. The prevailing age categories were: 36–45 and 46–55 years. The percentage of people over 55 years of age was relatively small, making only few percent of all batch. The majority had at least secondary education.

The validity study of GHQ-28 was carried out in another group of subjects; 392 patients filled in the questionnaire

and 80 of them were interviewed using the M-CIDI. The demographic characteristics of persons taking part in both studies are presented in Table 2. Like in the former group, men were slightly outnumbered by women. Age categories were similar as well. The 46–55 age category prevailed and was followed by the 36–45 age category. In the whole sample there were only 5% of persons aged over 55 years. A large majority of the subjects had at least secondary education.

The structure of the groups participating in a two-stage study (GHQ and M-CIDI) was similar to that covered only by GHQ-12 and GHQ-28 studies.

Validity indicarors

In our study, five validity indicators, described by Goldberg and Williams [5], were analyzed: sensitivity, specificity, overall misclassification rate, positive and negative predictive values.

Sensitivity is a probability that "true cases" will be correctly identified on the basis of the GHQ score. Specificity is a probability that healthy persons will be correctly identified on the ground of the GHQ score. The overall misclassification rate is the proportion of subjects whose mental health assessments (GHQ score and results of psychiatric interview) disagree. Positive predictive value is a probability that a person, whose GHQ score is above a cut-off point (assumed threshold) will appear to be a true case. Negative predictive value is a probability that a person whose score is below the cut-off point in the questionnaire will be a "true normal" at the psychiatric inter-

Table 1. Demographic characteristics of the subjects administered both GHQ-12 and psychiatric interview

		Sex				
		Women – 58.9 %		Men – 41.1 %		
Age	Years	≤25	26–35	36–45	46–55	≥56
	Persons (%)	13.3	15.6	30.0	35.5	5.6
	$\bar{x}$ = 40.9		SD = 10.8		Me = 43	
Education		Primary school	Vocational education	Secondary school		High school
Persons (%)		2.2	25.6	43.3		28.9

SD – standard deviation.
Me – median.

Table 2. Demographic characteristics of the subjects administered both GHQ-28 and psychiatric interview

		Sex				
		Women – 57.5 %		Men – 42.5 %		
	Years	≤25	26–35	36–45	46–55	≥56
Age	Persons (%)	11.2	18.8	26.2	38.8	5.0
	$\bar{x}$ = 41.7		SD = 11.1		Me = 44	
Education	Primary school		Vocational education	Secondary school		High school
Persons (%)	5.0	16.2	60.0	18.8		

SD – standard deviation.
Me – median.

view. The method of calculating these coefficients was described in Goldberg and Williams' handbook [5].

The validity coefficients described above were estimated for subsequent cut-off points of the GHQ score, from the lowest score in the acquired range and increasing gradually by 1 point. Two different methods of scoring were acquired on the basis of the GHQs: standard method called GHQ scoring and modified called the Goodchild and Duncan-Jones scoring (CGHQ) (types of scoring were described in research methods).

RESULTS

The samples of patients covered by both GHQ examination and interview made about 20% of the entirety of persons who answered the GHQ questions. It was partly because of an assumed principle that a patient who submitted a filled in GHQ should not wait for an interview until a psychologist ends interviewing another person, hence about 20% of respondents dropped out from the 2nd stage of the research. Very significant part of patients, about 60%, refused to take part in an interview because of its long duration. Only a few persons refused to be interviewed without giving any reasons.

In all analyses performed, only current mental disorders diagnosed employing the M-CIDI were taken into consideration. Owing to this procedure the results of GHQ and M-CIDI were comparable.

In the sample of 90 subjects who participated in the GHQ-12 validation study, 50 persons (55.6%) had at least one type of mental disorders that met the DSM-IV and ICD-10 diagnostic criteria; 31.1% of the patients had two or more disorders diagnosed simultaneously. In the sample of 80-subjects who participated in the GHQ-28 validation study, the proportion of persons with mental disorders was slightly lower – 38 persons (47.5%); "comorbidity" of at least two disorders was found in every second person in the group. Table 3 presents categories of disorders, and the percentage of subjects in both groups; GHQ-12 and GHQ-28.

In the group GHQ-12, 36.7% of subjects suffered from neurotic, stress-related and somatoform disorders. In this category, anxiety disorders as phobias (23.3%) and persistent somatoform pains disorders (18.9%) prevailed. In addition, panic disorder (2 persons), generalized anxiety disorder (4 persons) and one case of dissociative amnesia were diagnosed. In the category of mental and behavioral disorders due to psychoactive substance use, nicotine dependence (22.2%) predominated and in 9 persons it was the only type of disorder. Mood disorders were diagnosed in 18 persons (20.0%) with dysthymia and recurrent depressive disorders as the most frequent pathology (10.0% and 8.9%, respectively).

Mental and behavioral disorders due to psychoactive substance use were prevailing in the GHQ-28 group; they affected almost 29% of the subjects. The most frequent was again nicotine dependence (28.8%) and in 11 persons it was the only type of disorder. This was followed by neurotic disorders, among which anxiety disorders like phobias (16.2%) and somatoform pain disorders (12.5%)

Table 3. Prevalence of mental health disorders according to the ICD-10 categories

Categories of disorders	GHQ-12 group (N = 90)		GHQ-28 group (N = 80)	
	N	%	N	%
Mental and behavioral disorders due to psychoactive substance use	22	24.4	23	28.8
Affective disorders	18	20.0	11	13.8
Neurotic, stress-related and somatoform disorders	33	36.7	22	27.5
Organic, including symptomatic mental disorders	1	1.1	–	–
Behavioral syndromes associated with physiological disturbances and physical factors	–	–	1	1.2

were most frequent, and mood disorders (13.8%), including bipolar affective disorder, dysthymia, severe depressive episode and recurrent depressive disorder, were diagnosed.

To calculate the GHQ validity coefficients, patients who both filled in GHQ and went through an interview were divided into sub-groups. First, according to their GHQ results (i.e. a result below threshold value and above threshold value, at every GHQ cut-off point raised by 1 point). Second, according to criterion interview and affiliation to "normals" or "cases" sub-groups, and the diagnosis based on the M-CIDI. At the beginning, GHQ-12 and GHQ-28 validity coefficients were calculated from a group distribution considering the dichotomy: "normals" and "cases" on the ground of all current disorders diagnosed by the mean of the M-CIDI.

The results are presented in Table 4 for GHQ-12 and in Table 5 for GHQ-28 columns "a".

Single values of validity coefficients provide information only about the performance of a screening tool at one cut-off point. The necessity to define the best threshold for practical needs compels a compromise solution, between values of sensitivity and specificity. According to Goldberg and Williams' handbook [5], the optimal threshold value should correspond with the level in GHQ score, in which the questionnaire has simultaneously optimal sensitivity and specificity.

Sensitivity and specificity of GHQ-12 have the value most close to each other, when the cut-off points equals 1/2; sensitivity is 64% and specificity 65%, but specificity level seems especially difficult to be accepted in practice. In our opinion, it should not be lower than 75%. From this point of view, assuming a cut-off point level at 2/3 score seems to be more correct, and then GHQ-12 specificity is much better reaching 80% with sensitivity of 58%. At this GHQ threshold, the overall misclassification rate is lower and

Table 4. Validity coefficients of GHQ-12 for consecutive cutting scores for GHQ using M-CIDI as a criterion: (a) in the group of cases with all the disorders diagnosed, (b) in the group of cases with all disorders excluding nicotine dependence

Cutting score for GHQ	Sensitivity		Specificity		Overall misclassification rate		Positive predictive value		Negative predictive value	
	a	b	a	b	a	b	a	b	a	b
0/1	74	79	53	52	36	36	66	59	62	74
1/2	64	69	65	65	36	33	70	63	59	70
2/3	58	64	80	79	32	28	78	73	60	72
3/4	48	52	88	85	34	30	83	76	57	67
4/5	40	45	90	90	38	31	83	79	55	65
5/6	32	36	93	92	41	34	84	79	52	62
6/7	26	29	93	92	44	38	81	75	50	59

Table 5. Validity coefficients of GHQ-28 for consecutive cutting scores for GHQ using M-CIDI as a criterion: (a) in the group of cases with all the disorders diagnosed, (b) in the group of cases with all the disorders excluding nicotine dependence

Cutting score for GHQ	Sensitivity		Specificity		Overall misclassification rate		Positive predictive value		Negative predictive value	
	a	b	a	b	a	b	a	b	a	b
0/1	95	96	36	30	36	48	57	41	88	94
1/2	87	89	43	38	36	45	58	42	78	87
2/3	79	81	55	49	34	40	61	45	74	84
3/4	71	70	62	55	34	40	63	44	70	78
4/5	58	63	67	64	38	36	61	47	64	77
5/6	50	59	76	75	36	30	66	55	63	78
6/7	42	48	81	79	38	31	67	54	61	75
7/8	37	48	81	83	40	29	64	59	59	76
8/9	34	44	81	83	41	30	62	57	58	75
9/10	32	44	86	89	40	26	67	67	58	76
10/11	29	41	90	92	39	25	73	73	58	75
11/12	29	41	93	94	38	24	79	79	59	76
12/13	26	37	93	94	39	25	77	77	58	75

the coefficients of predictive value, important from a practical point of view, are higher, which is positive.

Acquisition of low GHQ-12 scores by a relatively high percentage of patients diagnosed as persons with disorders could evidence that the questionnaire is not sensitive enough in respect to some disorders in the group examined. Following an analysis of specific diagnoses and related number of persons, it was decided to exclude patients with nicotine dependence as the only disorder from the "cases" sub-group. Recalculated GHQ-12 validity coefficients, showed in column "b" of Table 4, indicated a few percent improvement in sensitivity. For the cut-off point equal to 2/3 of GHQ score, the validity coefficient was higher in comparison to the earlier one (see column "a") by 6% and was equal to 64%. The specificity changed only by 1% (it is now 79%) and the overall misclassification rate lowered from 32% to 28%. Both the positive and negative predictive values of GHQ-12 was 70% for the cut-off point of 2/3.

Table 5, column "a" presents the results of analyses performed in the GHQ-28 group, when subjects were divided into two sub-groups: "cases" and "normals" with respect to all the M-CIDI diagnoses. These results prove that sensitivity of the GHQ is satisfactory at low levels of the cut-off point of GHQ score, but at the same time specificity is low. Specificity of the questionnaire at the acceptable

level ≥ 75% coexists with sensitivity of maximum 50% – at GHQ cut-off level equal to 5/6 points.

We can get a few percent improvement in the GHQ-28 sensitivity, if patients with nicotine dependence as the only disorder are excluded from the subgroup "cases" (see: Table 5, column "b"). At the cut-off point of GHQ result at the level of 5/6, specificity of 75% is accompanied by the sensitivity coefficient equal to 59%, thus higher by 9% (comparing to column "a"). In this situation, the overall misclassification rate is lower by 30% in comparison with 36% in the "a" situation. A negative predictive value is significantly higher (by 15%) but, unfortunately, a positive predictive value is lower (55% only). Taking account of these validity coefficients, important from a practical point of view, suggests that GHQ cut-off point should also be considered on higher GHQ scores between 11 and 12, as an alternative solution regarding 5/6 cut-off point. The GHQ-28 predictive value will be then over 75% with also high specificity (94%) and relatively the lowest coefficient of the overall misclassification rate of patients. This error is, however, made mainly in the classification of people with disorders.

Table 6 shows validity coefficients of both General Health Questionnaires, with respect to a modified CGHQ scoring method. Applying the same criteria for acceptance of validity coefficients, no improvement of

Table 6. Validity coefficients of GHQ-12 and GHQ-28 for consecutive cutting scores for GHQ using the Goodchild and Duncan-Jones scoring method

Cutting score for GHQ	GHQ-12			GHQ-28		
	Sensitivity	Specificity	Overall misclassific ation rate	Sensitivity	Specificity	Overall misclassific ation rate
0/1	96	0	47	100	0	53
1/2	88	20	72	100	7	49
2/3	80	35	40	100	10	48
3/4	78	50	34	100	12	46
4/5	58	63	40	100	17	44
5/6	54	75	37	97	24	41
6/7	44	83	39	92	29	41
7/8	34	90	41	89	40	36
8/9	28	93	43	79	50	36
9/10				71	57	36
10/11				66	62	36
11/12				61	74	33
12/13				47	74	39
13/14				42	81	38
14/15				39	83	38

Table 7. Descriptive statistics of GHQ-12 and GHQ-28 scores using both standard scoring method (GHQ) and modified scoring method (CGHQ)

GHQ version	Scoring method	Mean	Median	Skewness	Scoring zero (%)	Standard deviation
GHQ-12	GHQ	2.96	2.0	1.13	37.8	3.53
	CGHQ	5.19	4.0	0.51	2.2	3.40
GHQ-28	GHQ	5.81	4.0	1.29	21.3	6.22
	CGHQ	11.28	11.0	0.43	0	5.65

GHQ-12 sensitivity was found after employing CGHQ scoring method. At the optimal threshold level 5/6, specificity equals 75% and sensitivity – 54%. Both coefficients are then lower than in the classical method of scoring GHQ data and optimal threshold value was calculated at 2/3 (compare with Table 4, column "a"). The effects of changing the scoring method on GHQ-28 validity coefficients were slightly different. By employing the CGHQ scoring method we acquired an increase in sensitivity by 11% at a slight difference in specificity (-2%). The optimal cut-off point was 11/12 with GHQ-28 sensitivity of 61% and specificity of 74%. One must point out, however, that the above mentioned sensitivity and specificity levels only insignificantly differed from the ones acquired when applying a standard scoring method with the exclusion of nicotine dependence disorder (compare with Table 5, column "b"; threshold value 5/6).

If responses are scored according to CGHQ, the optimum cut-off points in both GHQ-12 and GHQ-28 will be found at higher levels of scores. Theoretically, there is a possibility to acquire a larger number of points for cases with long-standing disorders. The basic descriptive statistics for the distribution of results calculated by the standard and modified methods are shown in Table 7. The data prove that distributions of results scored with the CGHQ method are more normal. Much lower percentage of patients is scored zero, the mean value is higher and skewness coefficient of the results distribution, scored according to CGHQ, is lower in comparison to the GHQ scoring system.

DISCUSSION

The investigation of the criterion validity of GHQ-12 and GHQ-28 questionnaires described here was a two-stage

study with a standardized psychiatric interview, and a diagnosis of mental disorder, based on M-CIDI interview, became a validity criterion for GHQ questionnaires. Both the method of validity study and the tool being the criterion basis satisfy the recommended and accepted procedures in this kind of research.

The proposed optimal threshold (for both GHQ-12 and GHQ-28) to identify persons with disorders resulted from the analysis of validity coefficients calculated for many cut-off points (from the lowest threshold value, and growing gradually by 1 point). They were fixed at the level of the highest possible sensitivity of the questionnaires and, at the same time, their specificity was not lower than 75%. These principles have been adopted for a practical reason, as the acceptance of the lower level of specificity causes that GHQ results of too high percentage of "non-disordered" persons are above the threshold, which forces medical practitioners to devote too much time to practically healthy people. This principal is in disagreement with Goldberg and Williams' opinion presented in the User's Guide to the General Health Questionnaire [5] and their later recommendation, which gave a priority to the sensitivity level versus specificity and adoption of low threshold values [12].

The suggested cut-off point for GHQ-12 at the level of GHQ scores equal to 2/3 is a recommended threshold [14] also obtained in many other studies [5]. The threshold proposed for GHQ-28, equal to 5/6 GHQ points is also the one quite often indicated as a result of various studies [5] in spite of the fact that the range of advocated cut-off points is far wider for this questionnaire than for GHQ-12. The optimum cut-off level for the results of the considered questionnaire, arising from our studies and applied criteria, was only by 1 point higher than the one suggested by Goldberg and Williams' on the ground of the research from the 1980s, in which the Present State Examination and the Clinical Interview Schedule were the criterion for GHQ [14]. It was also equal to the one recommended as a result of studies carried out in the 1990s under the patronage of WHO, in which disorders diagnosed according to the CIDI were the criterion for GHQ [6]. The indicated cut-off points concern the results

acquired by the standard scoring method, since it was not proved that the modified way of CGHQ scoring, which was an attempt to take into account the occurrence of long-lasting disorder symptoms, significantly improved validity coefficients of the questionnaires.

The validity coefficients: 64% sensitivity and 79% specificity for GHQ-12, and 59% sensitivity and 75% specificity for GHQ-28 are not particularly high. Their values are below the median value, which exceeds 80% in other studies. These validity data are not totally unique. The similar coefficients for GHQ-12 (sensitivity 68% and specificity 70%) were proved by Gureye and Obikoya [13] in a study carried out in Nigeria. An especially wide range of validity coefficients values were observed in the studies of GHQ-28 – sensitivity between 44% and 100%, and specificity between 74% and 93% [5]. In comparison with these values our results were not particularly low.

Relatively not too high sensitivity obtained for the GHQ in our study would be difficult to explain by the economic and/or cultural features of our general population. The studies of the GHQ-12 and GHQ-28 validity, carried out in the 1990s under the patronage of the World Health Organisation, covering almost 5500 patients of general practice settings in various countries, showed that General Health Questionnaires work well in developed and developing countries, with little differences in validity resulted from minor variations in the criteria used for defining "a case" [6]. The values of validity coefficients calculated in our study can be related rather to the specific character of the community examined than to the specific character of the population as a whole. This community, consisting of only working people, could be more homogenous than an entirety of adult patients of general practice settings, i.e. the community in which other studies have been usually carried out. Considering the homogeneity of the working community, one may say about its mental and somatic state of health that the state of health is not so bad and the disorder is not so acute to influence fundamentally the work ability of these people. This fact in connection with diseases accompanying somatic illnesses, which were the reasons for visiting a general practitioner, could lower the GHQ validity as a detector of men-

tal disorders in this particular group. Comparing GHQ-12 and GHQ-28 as far as validity coefficients are concerned, one can see that the GHQ-28 validity is worse. For sure, such a result was not influenced by differences in reliability of both questionnaires. The most popular reliability coefficient - Cronbach's α was 0.911 for GHQ-28 and 0.890 for GHQ-12. The reason behind the differences in criterion validity might be, however, a greater influence of somatic disease on GHQ-28 than on GHQ-12 scores. In the studies employing General Health Questionnaires, carried out during preventive, periodical examinations of working people, there was lack of significant differences between general GHQ-12 scores in a group of people with a diagnosed somatic disease and the results in a group of persons considered to be healthy, while there were significant differences, in terms of statistics, in the GHQ-28 scores influenced by the results of scale A, i.e. positive answers to questions concerning the occurrence of somatic symptoms in people physically ill [15].

CONCLUSIONS

1. The optimal cut-off points that can serve as a tool for identifying persons with mental disorder were established at the level of 2/3 points for GHQ-12 and 5/6 for GHQ-28. The results obtained from the study of Polish samples are similar to those reported in other studies.
2. Comparison of reliability coefficients (sensitivity and specificity) of two questionnaires under study proved better applicability of GHQ-12 in the studies aimed at identifying psychiatric cases in the working population attending primary care settings.

ACKNOWLEDGEMENTS

REFERENCES

1. Harnois G, Gabriel P. *Mental health and work: impact, issues and good practices.* Geneva: World Health Organization; 2000.

2. Sauter SL, Murphy LR, Hurrell JJ. *Prevention of work-related psychological disorders. A national strategy proposed by the National Institute for Occupational Safety and Health (NIOSH).* Am Psychol 1990; 45: 1146–58.

3. Kiejna A, Wojtyniak B, Stokwiszewski J, Łupiński P: *Frame of mind among Polish population in research carried out by the Central Statistical Office – an initial analysis.* Psychiatr Pol 2000; 34: 903–18 [in Polish].

4. Karski JB, Tomaszewska E, Koronkiewicz A. *Health status indicators of Polish population in 1992–1995 years.* Zdrow Publicz 1997} C VII: 221–36 [in Polish].

5. Goldberg DP, Williams P. *A User's Guide to the General Health Questionnaire.* Berkshire: NFER-Nelson Publishing Company Ltd.; 1991.

6. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureye O, et al. *The validity of two versions of the GHQ in the WHO study of mental illness in general health care.* Psychol Med 1997; 27: 191–7.

7. Quick JC, Murphy LR, Hurrell JJ. *Stress and Well-being at Work. Assessments and Interventions for Occupational Mental Health.* Washington: American Psychological Association; 1993.

8. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, et al. *The Composite International Diagnostic Interview*. Arch Gen Psychiatry 1988; 45: 1069–77.

9. Wittchen HU. *Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review.* J Psychiat. Res 1994; 28: 57–84.

10. World Health Organization. *The ICD-10 Classification of Mental Behavioural Disorders: Clinical Description and Diagnostic Quidelines.* Geneva: WHO; 1992.

11. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* 4th ed. Washington DC: American Psychiatric Association; 1994.

12. Goldberg DP, Oldehinkel T, Ormel J. *Why GHQ threshold varies from one place to another.* Psychol Med 1998; 28: 915–21.

13. Gureye O, Obikoya B. *The GHQ-12 as a screening tool in a primary care setting.* Soc Psychiatry Psychiatr Epidemiol 1990; 25: 276–80.

14. Mc Dowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires.* New York, Oxford: Oxford University Press; 1987.

15. Makowska Z, Merecz D. *Usefulness of General Health Questionnaires for diagnosis of employees' mental health*. Med Pr 2000; 51: 589–601 [in Polish].