Soc Psychiatry Psychiatr Epidemiol (1995) 30: 20–25

ORIGINAL PAPER

G. Lewis · R. I. Araya

Is the General Health Questionnaire (12 item) a culturally biased measure of psychiatric disorder?

Accepted: 23 June 1994

Abstract There have been suggestions that some self-administered questionnaires designed to assess psychiatric disorder tend to overestimate prevalence in samples from Latin America. This phenomenon may be obscured when the General Health Questionnaire (GHQ) is used, as it is recommended that researchers determine the threshold in each setting by comparing the GHQ with a standardised interview. Reports in the literature suggest that Latin American samples have a higher threshold for case definition using the GHQ than that found in British samples. The present study confirmed this finding when comparing the 12-item GHQ in a Chilean primary care sample with a sample of primary care attenders from the United Kingdom. The increase in GHQ scores in the Chilean sample persisted after adjustment for age, sex, marital status and the score on the Revised Clinical Interview Schedule (CIS-R). The increase in scores seen in the Chilean sample was only found in that half of the GHQ that asks about negative aspects of mental health.

Introduction

The General Health Questionnaire (GHQ; Goldberg 1972) is one of the most widely used self-administered questionnaires designed to assess psychiatric disorder. It has been studied and employed in a variety of developed and developing countries (Goldberg and Williams 1988).

G. Lewis (✉)
Section of Epidemiology and General Practice, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK

G. Lewis
Epidemiology Unit, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, UK

R. I. Araya
Public Health Unit, School of Medicine, University of Chile, Condell 300, Santiago, Chile

Many epidemiological studies compare rates between groups, for example, by comparing rates of disorder between different social classes, between different regions of a country or between countries. The conclusions from this sort of study will be invalid if the assessment of disease is biased as a result of differences between the groups. This study was designed to investigate the possibility that the GHQ shows an ascertainment bias (Last 1988) when used to compare rates of psychiatric disorder between the United Kingdom and Latin America. An ascertainment bias of this kind could be described as a cultural bias, in the sense that any difference between respondents in the United Kingdom and Latin America presumably depends upon some cultural influence. Culture has been defined as the "'social heritage' of the community" including the "the mental and spiritual 'artifacts' (systems of symbols), ideas, beliefs, aesthetic perceptions, values etc.)" and "transmitted from generation to generation" (Fletcher 1977).

The possibility that self-administered questionnaires have led to the biased assessment of psychiatric disorder was discussed by Dohrenwend in 1966. He drew attention to two forms of potential bias: social desirability bias and 'yeasaying' or acquiescence bias (Couch and Keniston 1960). Dohrenwend noted that the Puerto Rican ethnic groups in Washington Heights had higher scores on the Langner (1962) 22-item questionnaire than other disadvantaged groups, and this effect persisted after adjusting for income and occupation. He also cited work that found, using the same scale, that the general population in Mexico City scored almost as much (mean 5.4) as psychiatric patients in New York City (mean 6.1) and almost twice as much as community respondents in New York City (mean 2.8). Dohrenwend argued that a social desirability bias must contribute to the higher scores observed amongst Puerto Rican immigrants and Mexicans.

Interest in this work has recently been renewed by Guarnaccia et al. (1990) who have commented on the finding that Puerto Ricans and Puerto Rican immi-

Table 1 Validity studies of the 12 item General Health Questionnaire (GHQ-12) in the UK and Latin America

Authors	Setting	Threshold
Latin American Studies		
Mari and Williams (1986)	Brazil, primary care	3/4
Araya et al. (1992)	Chile, primary care	4/5
British Studies		
Goldberg (1972)	Primary care	1/2
Wessely and Lewis (1989)	Dermatology clinic	1/2
Banks (1983)	Unemployed youths	2/3

grants to the United States have higher scores on the Centre for Epidemiologic Studies Depression scale than other community respondents (CES-D; Radloff 1977). In contrast, a recent survey in Puerto Rico (Canino et al. 1987) using a standardised interview (the Diagnostic Interview Schedule; DIS; Robins et al. 1981) has found similar rates of disorder in Puerto Rico and in Mainland USA. Guarnaccia et al. (1990) argue that the presumed bias underlying the CES-D results from the Latin American concept of 'ataques de nervios', which results in items on the CES-D being seen as more socially desirable than in other North American ethnic groups. One weakness of these studies is that they did not use both a self-administered questionnaire and a standardised interview in the same study.

When the GHQ is used as a means of case definition, it is recommended that a threshold is determined by comparing the GHQ with a standardised interview (Goldberg and Williams 1988). The threshold score is then determined in order to maximise the sensitivity and specificity of the GHQ. This procedure may have tended to inhibit the comparison of scores across cultures and the discussion of the possibility that the GHQ may be subject to a cultural bias. It is, therefore, of interest to compare the thresholds determined in Latin America with those determined in the United Kingdom and North America.

In general there seems to be an increase in the threshold for the GHQ-12 when used in Latin America compared with that used in the United Kingdom (Table 1). Mari and Williams (1985) found that a threshold of 3/4 was needed in a sample of over 250 general practice attenders in a poor area of Sao Paulo, Brazil and Araya et al. (1992) found that a threshold of 4/5 was optimal for the GHQ-12 in a primary care sample in Chile. In contrast, Goldberg's (1972) original validity study in the United Kingdom found that a threshold of 1/2 was optimal for the GHQ-12 (disembedded). Banks (1983) found the best threshold was 2/3 in a sample of 17-year-old school leavers though the Present State Examination (PSE; Wing et al. 1974) was used as the criterion and, on occasions, this interview appears to have a higher threshold than the Clinical Interview Schedule (CIS; Goldberg et al. 1970) used by Goldberg (1972). The data published by Wessely and Lewis (1989) illustrated that the best threshold was 1/2 for the GHQ-12 in a British dermatology clinic. Medina-Mora et al. (1983) conducted a study with the GHQ-30 in Mexico and found that 4/5 was the best threshold. This, in contrast with the previous results cited for the GHQ-12, is the same as the threshold determined by Goldberg (1972) for the GHQ-30 in the United Kingdom. Two studies conducted in Chile (Trucco et al. 1979; Torres and Alvarez 1987) did not use a standardised assessment to define cases and are therefore difficult to include in this comparison.

One problem in attempting to assess bias in psychiatric research is the absence of any criterion measure of psychiatric disorder. It is widely assumed that standardised interviews such as the CIS or PSE are more accurate and more valid measures of psychiatric disorder than self-administered questionnaires such as the GHQ. Furthermore, in many of the currently used standardised interviews, including the PSE and CIS, the psychiatrist is encouraged to use clinical judgements in deciding upon the presence or absence of psychopathology (Lewis and Williams 1989). It is possible, therefore, that differences between countries in the results from standardised interviews arise from the idiosyncratic views of the interviewers. Standardised interviews now exist that are more highly standardised than those previously available and these include the Revised Clinical Interview Schedule (CIS-R; Lewis et al. 1992) and the DIS (Robins et al. 1981).

Despite these problems, there are some arguments that support the assumption that standardised interviews are more valid as measures of psychiatric disorder than a self-administered questionnaire such as the GHQ. First, there is a marked difference between the sort of questions asked in the GHQ and those asked in a standardised interview. The GHQ tends to include questions such as "Have you recently felt on the whole you were doing things well?" and "Have you recently been having disturbed restless nights?". In contrast, standardised interviews ask about specific symptoms, for example, "Have you had spells of feeling low in mood, sad or miserable recently?". Supplementary questions in standardised interviews usually acquire more detailed information about the severity and frequency of symptoms than are available from self-administered questionnaires. This method of enquiry in standardised interviews is, therefore, much closer to the methods used by clinical psychiatrists. Furthermore, the tendency to use idiomatic forms of language in self-administered questionnaires may be expected to encourage differences in interpretation dependent upon language and culture.

The GHQ was designed to include questions asking about both positive and negative aspects of mental health. Duncan-Jones et al. (1986), using latent trait analysis, and Lewis (1992), with principal components analysis, have both argued for the usefulness of treating the two halves of the GHQ as two sub-scales. The positive mental health sub-scale is tapped by items that ask about normal functioning and contribute to the score when answered "less than usual"; for example, "Have you recently been satisfied with the way you've carried

Table 2 Means (95 % CI) for the GHQ and Revised Clinical Interview Schedule (CIS-R) for the two samples and the differences between the samples before and after adjustment for CIS-R score, age, sex and marital status

	GHQ total score	GHQ negative score	GHQ positive score	CIS-R score	Sample size
Chile	16.2 (14.9–17.5)	8.5 (7.8–9.3)	7.6 (7.0–8.3)	14.6 (12.7–16.6)	163
UK	12.3 (11.2–13.5)	4.9 (4.2–5.6)	7.5 (6.9–8.0)	13.1 (11.3–14.9)	107
Chile UK difference	3.8 (2.0–5.7)	3.6 (2.5–4.7)	0.1 (–0.8–1.1)		
Chile UK difference after adjustment	3.1 (1.6–4.5)	3.1 (2.3–4.0)	–0.2 (–1.0–0.6)		

out your task(s)?". The other sub-scale asks about symptoms, therefore assessing negative mental health, and includes items such as "Have you recently lost much sleep over worry?". A "more than usual" response leads to a score on such items. Whether one views these sub-scales as a methodological artefact or as an important and useful description of different facets of mental health, they may be relevant to the possibility of bias in the GHQ. It is interesting to note that Krause et al. (1990) suggest that Punjabi primary care attenders seem to show a slightly different pattern of scores on the 28-item GHQ compared with primary care attenders of white British origin.

The present study was designed to investigate the possibility of a cultural bias between the United Kingdom and Latin America affecting the GHQ. The aims were first, to compare total scores and scores on the positive and negative sub-scales of the GHQ between samples in Chile and the United Kingdom. We argued that standardised interviews such as the CIS-R are more valid in assessing psychiatric disorder than self-administered questionnaires. Our second aim was, therefore, to adjust the comparison of GHQ scores for differences in CIS-R scores and also for some comparable demographic variables between the samples. We also wished to examine the possibility that the mode of administration (interviewer versus self-administered) would influence any findings of bias.

Method

The Chilean sample was collected in a primary care clinic, Lo Prado, serving a large working class urban area of Santiago, Chile. Consecutive attenders at the daily medical clinics answered the GHQ-12 and a sociodemographic enquiry. Illiterate patients had their questionnaires read out. Subsequently, the CIS-R (Lewis et al. 1992) was administered by lay interviewers or a psychiatrist (RIA). The CIS-R was translated into Spanish by RIA and translated back into English by two Chilean nationals (one a British-trained psychiatrist, the other a non-psychiatrist) who speak Spanish as a first language. RIA is a Chilean, Spanish-speaking psychiatrist who has trained in psychiatry in the United Kingdom.

The British sample was drawn from attenders at a health centre in Thamesmead, South-East London. No information was collected on non-responders. Subjects were given two interviews, administered either by a psychiatrist or lay interviewer, using the CIS-R. Results from the first interview are used here. The self-administered GHQ-12 was completed by all subjects and demographic information was collected by interview. Further details are given by Lewis et al. (1992).

Three scores were calculated from the GHQ-12. The GHQ total score was the sum of all items scored using the Likert method (Goldberg 1972). Items 2, 5, 6, 9, 10 and 11 were classified as assessing negative mental health as indicated by a "more than usual" response. These items were summed as the negative GHQ scale. The remaining items were summed as the positive GHQ scale. When the GHQ was used to define cases using a threshold score, it was scored in the traditional manner (Goldberg 1972). The difference between the GHQ scores in the Chilean and British samples were adjusted for age, sex, marital status and CIS-R scores using multiple regression performed with the GLM procedure of the program SAS (SAS Institute Inc. 1985).

Two other samples available to the authors were also used in the study. The 12-item GHQ was disembedded from the 30-item GHQ used in the Health and Lifestyle Survey. Details of method are published by Cox et al. (1987). This was a large-scale community survey conducted in the United Kingdom in 1984 and 1985. A sample of 12,254 people was selected using the electoral register. The second sample was a survey of 173 new attenders at a dermatology clinic in which the GHQ-12 was used. Details of method are given by Wessely and Lewis (1989).

Results

In the Chilean sample, 170 consecutive attenders were included in the study of whom 7 refused to take part for various practical reasons. The mean age was 38 years; 74 % of the attenders were women; 53 % were married. Only 11 % of the sample were illiterate; 85 % of the sample (or their spouses) were in unskilled labouring jobs.

The British sample consisted of 107 subjects with a mean age of 39 years of whom 72 were female; 27 % were single and 57 % married; 13.6 % of participants were born outside the United Kingdom. No information was collected on non-responders and this sample probably represented about 10 % of the attenders at the clinic. Other details on both samples are given by Lewis et al. (1992).

The means of the GHQ scores and the CIS-R scores for the two samples are shown in Table 2. The mean CIS-R score was somewhat higher for the Chilean sample than the British sample but this was not statistically significant. However, the GHQ total score was markedly higher in the Chilean sample (mean difference 3.8) and the difference was explained almost entirely by the difference in the negative scale of the GHQ (3.6). The difference between the Chilean and British samples on the GHQ positive scale was not statistically significant. The British sample had a higher mean for the positive

Table 3 Percentage of cases (95 % CI) in the two samples using different case definitions

Case definition	Chile	UK
CIS-R	52.8 % (45.1–60.5)	50.5 % (41.0–60.0)
GHQ-12 (threshold 1/2)	76.1 % (69.6–82.6)	52.8 % (43.3–62.3)
GHQ-12 (threshold 4/5)	52.8 % (45.1–60.5)	27.9 % (19.4–36.4)

Table 4 Means (95 % CI) for the sub-scales of the GHQ in the Chilean sample, by mode of administration

	Self-administered	Read by interviewer because of illiteracy	Difference between means	*P* value
GHQ negative	8.3 (7.5, 9.1)	10.6 (7.7, 13.5)	2.3 (–0.2, 4.8)	$t_{161} = 1.79$ $P = 0.07$
GHQ positive	7.6 (6.9, 8.3)	7.9 (5.1, 10.7)	0.3 (–2.3, 2.9)	$t_{161} = 0.23$ $P = 0.8$
Sample size	145	18		

Table 5 Means (SD) for the negative and positive scales of the GHQ-12 from two British samples: the Health and Lifestyle Survey and a dermatology clinic

	Health and Lifestyle Survey	Dermatology clinic
GHQ negative	4.3 (3.69)	5.00 (4.00)
GHQ positive	6.5 (1.93)	7.0 (2.68)
Sample size	6487	129

scale than for the negative scale. In contrast, subjects in the Chilean sample had higher scores on the negative scale than the positive scale of the GHQ.

The difference in GHQ scores between the samples was adjusted for the scores on the CIS-R and the age, sex and marital status of the individuals. Table 2 gives also the results after adjustment. There was only a modest change in the size of the differences between the samples after adjustment.

It is common for researchers to classify subjects as 'cases' or 'non-cases' using a threshold score on the GHQ. This has been done in Table 3 to illustrate the size of the bias observed in these samples. The prevalence using the CIS-R was approximately 50 % in both samples, so adjustment for CIS-R scores was not necessary. Over 70 % of the Chilean sample scored more than the threshold determined in the British sample (1/2). In contrast, 30 % of the British sample scored more than the 4/5 threshold most suitable for the Chilean sample.

The illiterate subjects in the Chilean sample responded to the GHQ after having the questions and responses read out by an interviewer. The illiterate subjects had higher total GHQ scores than the remainder of the sample, and as can be seen in Table 4, the illiterate group still had higher scores on the negative than the positive sub-scale of the GHQ. The mean GHQ negative score in both the self-administered (8.3) and illiterate group (10.6) was still considerably larger than the mean GHQ negative score in the British sample (4.9; Table 2).

Results from the two other British samples are given in Table 5. In both the Health and Lifestyle Survey and the dermatology clinic sample, the subjects had higher scores on the positive scale of the GHQ than on the negative scale of the GHQ.

Discussion

The data presented here suggested that subjects in Chile had higher scores on the GHQ relative to a standardised assessment of psychiatric disorder, the CIS-R. This evidence of an ascertainment bias seemed only to occur on the half of the GHQ assessing negative aspects of mental health – those questions that score when the subject responds "more than usual" to the question. This result persisted after adjustment for the age, sex and marital status of the subjects. No attempt was made to adjust for socioeconomic status variables as there were concerns about the cultural equivalence of these. Both samples were drawn from relatively deprived areas of both countries.

There is a problem concerning the generalisability of these results. The British sample was not representative of attenders at the primary care clinic. Furthermore, any primary care sample is selected in the sense that the subjects choose to consult with a doctor. However, evidence from two other samples, a large community survey from the United Kingdom and a smaller survey of dermatology clinic attenders, supports the results from the British primary care sample. In all these three British samples the subjects scored about 2 points more on the positive scale of the GHQ than the negative scale. In contrast, subjects from the Chilean sample scored more on the negative than the positive scale of the GHQ. This latter result has also been found in a sample of Brazilian primary health care attenders (Mari and Iacoponi, personal communication).

The conclusion of an ascertainment bias rests, in part, upon the assumption that the use of the CIS-R in the two cultures was not affected by any cultural bias. This position is difficult to justify in the absence of any gold standards in psychiatry. However, the observation that the relationship between the two GHQ scales differed between the two cultures supports the notion that the GHQ is prone to a cultural bias in this context. It is also important to point out that the CIS-R was translated by a Chilean psychiatrist to ensure that the concepts used were appropriate for use in Latin American culture.

The results for the Chilean sample stratified by literacy indicated that the mode of administration of the GHQ did not affect the bias described here, though the number of illiterate subjects was small. The illiterate group, if anything, seemed to show an increase in the difference between the negative and positive scales of the GHQ, though this could well have something to do

with other characteristics of the illiterate group. Therefore, the differing performances of the GHQ and CISR when comparing Latin America and the United Kingdom cannot be explained by the fact that the GHQ is self-administered. Any explanation for this bias must therefore rely upon other differences between the assessments.

A cultural bias in responses to the GHQ is a major problem in using the GHQ in cross-cultural research. To a certain extent this problem can be overcome by classifying individuals 'cases' or 'non-cases' and determining the threshold within the population to be studied, as advocated by Goldberg and Williams (1988). However, minor psychiatric disorder is probably best viewed as a continuum, and Rose (1985) has persuasively argued that dividing subjects into those with and those without disease is not the best approach towards comparing populations and in investigating public health.

These findings also raise the perhaps more serious problem of a cultural bias between groups within a single population. Stansfeld and Marmot (1992) have demonstrated a bias between socioeconomic groups in an occupational cohort in the United Kingdom. Those in the lower occupational grades in the civil service are more likely to be false-negatives on the GHQ when the CIS is used as a criterion. This presumed bias has the effect of eliminating a social class gradient in the prevalence of minor psychiatric disorder. Furthermore, evidence from studies conducted in Brazil (Mari and Williams 1986) and Chile (Araya et al. 1992) have reported that those with less education are more likely to be false-positives on the GHQ.

This work may have relevance for other self-administered questionnaires designed to assess minor psychiatric disorder. The literature suggesting a similar cultural bias in the CES-D when used in Latin America was reviewed in the Introduction. The Symptom Reporting Questionnaire (SRQ; Harding et al. 1980) is widely used in many developing countries, though rarely in developed countries. All the questions in the SRQ assess negative aspects of mental health (and some ask about somatic symptoms), and this questionnaire may also be susceptible to a similar bias to that described here for the GHQ.

It is of interest to speculate about the possible reasons for the GHQ's behaviour in Latin America. Without further research including qualitative methods it is impossible to say very much. However, our results were similar to those described by Guarnaccia et al. (1990) and supported their notion that symptoms of psychiatric disorder are more socially acceptable in Latin American culture than in other groups within North America and the United Kingdom.

Conclusion

Responses to the GHQ are probably influenced by cultural factors and Latin American populations have higher scores on the part of the GHQ assessing negative aspects of mental health. This limits the usefulness of the GHQ in cross-cultural comparisons and raises the possibility of biased assessment when comparing populations of different cultures within a single country. This phenomenon does not appear to result merely because the GHQ is self-administered while standardised interviews are administered by an interviewer. A cultural bias may also occur when other self-administered questionnaires, apart from the GHQ, are used to assess psychiatric disorder.

Acknowledgements We would like to thank the Bethlem Royal and Maudsley Research Fund and Prof. Michael Shepherd for providing funding for RIA to conduct the Chilean study. The data from the United Kingdom was collected during a study financed by the Health Promotion Research Trust, and we thank Dr. Anthony J. Pelosi and Caroline Croft who helped to collect the data and to Dr. Simon Wessely who helped to collect the dermatology clinic data. We are very grateful to Jair Mari and Eduardo Iacoponi for providing us with previously unpublished analyses. Our thanks to Prof. David Goldberg and Dr. Keith Lloyd who provided many useful comments on an earlier draft of the paper.

References

Araya R, Wynn R, Lewis G (1992) A comparison of two self-administered psychiatric questionnaires (GHQ-12 and SRQ) in primary care in Chile. Soc Psychiatry Psychiatr Epidemiol 27: 168–173

Banks MH (1983) Validation of the General Health Questionnaire in a young community sample. Psychol Med 13: 349–354

Canino GJ, Bird HR, Shrout PE, Rubio-Stipec M, Bravo M, Martinez R, Sesman M, Guevara LM (1987) The prevalence of specific psychiatric disorders in Puerto Rico. Arch Gen Psychiatry 44: 727–735

Couch A, Keniston K (1960) Yeasayers and naysayers: agreeing response set as a personality variable. J Abnormal Soc Psychol 60: 151–174

Cox BD, Blaxter M, Buckle ALJ, Fenner NP, Golding JF, Gore M, Huppert FA, Nickson J, Roth M, Stark J, Wadsworth MEJ, Whichelow M (1987) The Health and Lifestyle Survey. Health Promotion Research Trust, Cambridge

Dohrenwend BP (1966) Social status and psychological disorder: an issue of substance and an issue of method. Am Sociol Rev 31: 14–34

Duncan-Jones P, Grayson DA, Moran PAP (1986) The utility of latent trait models in psychiatric epidemiology Psychol Med 16: 391–405

Fletcher R (1977) Culture. In: Bullock A, Stallybrass O (eds) The Fontana dictionary of modern thought. Fontana, London

Goldberg DP (1972) The detection of psychiatric illness by questionnaire. Oxford University Press, London

Goldberg D, Williams P (1988) User's guide to the General Health Questionnaire. NFER-NELSON, Windsor

Goldberg DP, Cooper B, Eastwood MR, Kedward HB, Shepherd M (1970) A standardised psychiatric interview for use in community surveys. Br J Prev Soc Med 24: 18–23

Guarnaccia PJ, Good BJ, Kleinman A (1990) A critical review of epidemiological studies of Puerto Rican Mental health. Am J Psychiatry 147: 1449–1456

Harding TW, De Arango MV, Baltazar J et al (1980) Mental disorders in primary health care: a study of their frequency and diagnosis in four developing countries. Psychol Med 10: 231–241

Krause I-B, Rosser RM, Khiani ML, Lotay NS (1990) Psychiatric morbidity among Punjabi medical patients in England measured by General Health Questionaire. Psychol Med 20: 711-719

Langner TS (1962) A twenty-two item screening score of psychiatric symptoms indicating impairment. J Health Hum Behav 3: 269–276

Last JM (ed) (1988) A dictionary of epidemiology, 2nd edn. Oxford University Press, Oxford

Lewis G (1992) Dimensions of neurosis Psychol Med 22: 1011–1018

Lewis G, Williams P (1989) Clinical judgement and the standardised interview in psychiatry. Psychol Med 19: 971–980

Lewis G, Pelosi AJ, Araya RI, Dunn G (1992) Measuring psychiatric disorder in the community: a standardised assessment for use by lay interviewers. Psychol Med 22: 465–486

Mari JJ, Williams P (1985) A comparison of the validity of two psychiatric screening questionnaires (GHQ-12 and SRQ-20) in Brazil, using Relative Operating Characteristic (ROC) analysis. Psychol Med 15: 651–659

Mari JJ, Williams P (1986) Misclassification by psychiatric screening questionnaires. J Chronic Dis 39: 371–378

Medina-Mora M, Padilla GP, Campillo-Serrano C, Mas CC, Ezban M, Caraveo J, Corona J (1983) The factor structure of the GHQ: a scaled version for a hospital's general practice service in Mexico. Psychol Med 13: 355–361

Radloff LS (1977) The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1: 385–401

Robins LN, Helzer JE, Croughan J, Ratcliff KS (1981) National Institute of Mental Health Diagnostic Interview Schedule; its history, characteristics and validity. Arch Gen Psychiatry 38: 381–389

Rose G (1985) Sick individuals and sick populations. In J Epidemiol 14: 32–38

SAS Institute Inc. (1985) SAS user's guide: basics. Version 5. SAS Institute Inc., Cary, N. C.

Stansfeld SA, Marmot MG (1992) Social class and minor psychiatric disorder in civil servants: a validated screening survey using the General Health Questionnaire. Psychol Med 22: 739–749

Torres P, Alvarez V (1987) Validez de un cuestionario para detectar desordenes emocionales (GHQ-12) en estudiantes universitarios: comunicación preliminar. Rev Psiquiatr 4: 255–264

Trucco M, Campusano ME, Larrain S (1979) Un cuestionario para detectar desordenes emocionales. Estudio de validación preliminar. Rev Chil Neuropsiquiatr 17: 20–25

Wessely SC, Lewis GH (1989) The classification of psychiatric morbidity in attenders at a dermatology clinic. Br J Psychiatry 155: 686–691

Wing JK, Cooper JE, Sartorius N (1974) The measurement and classification of psychiatric symptoms. Cambridge University Press, Cambridge