

Ethno-psychometric evaluation of the General Health Questionnaire in rural China

DOMINIC T. S. LEE^{1,2*}, WINNIE C. M. YIP³, YENFONG CHEN⁴, QINGYUE MENG⁵
AND ARTHUR KLEINMAN^{1,6}

¹ *Department of Social Medicine, Harvard Medical School, Boston, MA, USA*; ² *Department of Psychiatry, The Chinese University of Hong Kong, Shatin, Hong Kong*; ³ *Department of Population and International Health, Harvard School of Public Health, Boston, MA, USA*; ⁴ *Hui Long Guan Hospital, Beijing, China*; ⁵ *Centre for Health Management and Policy, Shandong University, Shandong, China*; ⁶ *Department of Anthropology, Harvard University, Cambridge, MA, USA*

ABSTRACT

Background. Most mental health research tools are developed in Western, urban contexts. Few studies have evaluated the applicability of these research tools in rural populations of non-Western countries. We examined the cultural acceptance and psychometric performance of the 12-item General Health Questionnaire (GHQ) in China's rural villages.

Method. Ethnographic investigations were conducted to assess the cultural applicability of self-report rating scales among villagers. This was followed by a survey of 1401 rural residents, randomly selected from 48 villages of Shandong province using stratified multistage cluster sampling. The respondents were administered the GHQ and the Composite International Diagnostic Interview (CIDI).

Results. The GHQ, when administered by trained interviewers, was culturally acceptable to rural residents. The scale had good psychometric properties in the study population. The area under the curve was 0.86. At a cut-off of 1/2, the sensitivity and specificity were 80.6% and 79.3% respectively.

Conclusions. The ethno-psychometric evaluation showed that the GHQ was both culturally valid and psychometrically sound in the Chinese rural context.

INTRODUCTION

Whilst half of the world's population live in rural areas, comparatively little mental health research has been conducted among rural populations, especially the 91% of the rural population living in lesser developed countries (Murphy, 1998; United Nations Population Division, 2002).

It is not surprising that most mental health research tools are developed in Western, urban contexts, and only rarely do these research tools receive evaluation in rural villages outside

the Western world. Furthermore, even when research tools are developed to cater for lesser developed societies and for cross-cultural comparison, the unique circumstances of rurality are rarely reflected in the design and content of the tools. This situation is unsatisfactory as properly validated and culturally sensitive research tools are fundamental to quality research (Furukawa *et al.* 2002).

We report a multi-method evaluation of the General Health Questionnaire (GHQ) among people living in rural China. The GHQ has been commonly used to examine mental health in lesser developed countries. While validation studies of the GHQ in lesser developed countries are mostly positive, those studies have been

* Address for correspondence: Professor Dominic T. S. Lee, Department of Psychiatry, The Chinese University of Hong Kong, Shatin, NT, Hong Kong SAR, China.
(Email: dominiclee@cuhk.edu.hk)

conducted in urban or suburban populations (McDowell & Newell, 1996). Hence, their findings may not necessarily be applicable to rural populations, whose local language and indigenous culture can differentially shape the experience of mental illnesses. Moreover, few studies have explored how to administer self-report rating scales like the GHQ in order to study a population that is largely illiterate. It is also uncertain if there is a willingness among rural respondents to reveal their inner emotional thoughts and psychosocial problems to survey researchers, whom they regard as outsiders.

ETHNOGRAPHIC VALIDATION

Method

The present study is part of a larger research project on the mental health impact of economic development in rural China. This study began with an ethnographic enquiry of the acceptance and applicability of psychological self-report rating scales among rural residents. Ethnography is the scientific description of the customs of individual peoples and cultures. The qualitative methodology involves the close study of a local world, i.e. a village, an urban neighbourhood, an institution, a network, or a patient population. Ethnography evaluates local language, conceptual categories, indigenous practices, and above all experiences and meanings (Kleinman, 2000).

In a review of the literature we found that there were few reports on the evaluation of self-report rating scales in non-Western rural populations. To discover unpublished information and knowledge, we consulted five Chinese psychiatric epidemiologists and journal editors for guidance. We were advised to administer the self-reporting rating scales with trained interviewers, while the respondents read a copy. We were also warned that rural respondents might have difficulties with sophisticated mental health concepts.

In August 2003, our survey team and ethnographic team jointly conducted field visits in three villages of Shandong province. A battery of scales [Center of Epidemiologic Study Depression Scale (CES-D), State Trait Anxiety Inventory (STAI), Perceived Stress Scale (PSS), Hospital Anxiety Depression Scale (HADS),

Satisfaction with Life Scale (SWLS), and General Health Questionnaire (GHQ)] were evaluated among 40 respondents purposely selected for education level and exposure to an urban lifestyle (sociodemographic characteristics of respondents are available from the authors). First, the interviewers asked the respondents to self-administer all the scales. If the respondents had difficulty in completing the scales by themselves, the interviewers would then administer the scales.

The ethnographic team also examined how the villagers perceived the survey. The team was initially concerned that the villagers would be unwilling to reveal their private socio-emotive experiences to researchers whom they regarded as outsiders. Moreover, there is no culture of surveys or opinion polls in China's villages. Revealing family conflicts to outsiders is also taboo in many cultures, and this too may deter respondents from reporting their emotional distress. It should also be considered if past political trauma (e.g. the Cultural Revolution) may inhibit respondents from reporting their emotional distress, which may be connected with their socio-political lives (e.g. social injustice and discontent with government).

Findings

We originally planned to 'debrief' the respondents after they had completed the scales to identify misunderstandings and misinterpretations of questionnaire items. However, we quickly realized that nearly all respondents, including the educated group, strongly preferred the interviewers to administer the scale. This was mainly due to the respondents being unaccustomed to reporting emotional feelings in a written format. The respondents were willing to talk about their personal feelings to the interviewers, but were unwilling to reveal the same information in a written questionnaire.

Moreover, nearly all respondents had difficulties in understanding the items of rating scales designed for specific affective constructs (CES-D, STAI, PSS, HADS). For instance, the respondents could not differentiate affective terms that were closely related (e.g. tense, strained and jittery). Some affective terms were also alien to the respondents, and even if the respondents had previously heard of the terms, they were not entirely sure of the meanings.

In contrast, the respondents found the rating scales of general psychological well-being (GHQ, SWLS) to be acceptable and understandable. Explanations and interpretations, however, were still needed for unfamiliar terms. The respondents, however, generally preferred the interviewers to administer the scales.

We interviewed five key village dignitaries (village chief, village doctors, teacher, and party official) and 20 villagers (purposely selected on age, gender, income and education) regarding the cultural acceptability of community surveys and self-report questionnaires on psychosocial pathologies. Most of the respondents believed that the research team must convince the villagers that they were not affiliated with local government. Only when the villagers were assured of privacy and confidentiality, would they feel secure enough to reveal their emotional and social state.

PSYCHOMETRIC VALIDATION

Participants

The validation of the GHQ involved a cross-sectional sample of residents from 48 villages of Shandong province. We first chose three counties with different levels of economic growth. Within each county, we stratified towns into quartiles based on economic development and randomly chose one town within each quartile. We further stratified villages within each town into quartiles and randomly chose one village from each. Within each village, we randomly chose around 10% of the households. All household residents aged between 16 and 80 years were invited to participate in the study. Informed consents were obtained from the respondents before the interview began. Eligible participants were only excluded if informed consent was not provided. The study protocol was approved by the Human Subject Protection Committees of both the Harvard School of Public Health and Shandong University.

Protocol

Participants were given a thorough description of the study before completing written informed consent. The GHQ and SWLS were then administered by the trained interviewers, with the participants reading a copy (if participants were literate). To ensure that the GHQ and SWLS

were administered consistently, the interviewers had received prior training on the administration of the scales. To ensure the quality and completeness of data collected during the survey, daily meetings were held among the interviewers. After the GHQ and SWLS, the Composite International Diagnostic Interview (CIDI) was administered to the participants by trained interviewers to establish 'gold-standard' psychiatric diagnosis for criterion comparison.

Assessment tools

The GHQ was originally designed for the identification of current non-psychotic disorders among patients in a general medical setting and respondents in the community (Goldberg & Williams, 1988). The GHQ has been used worldwide as a general measure of psychological well-being, and at least seven different language versions have been validated (McDowell & Newell, 1996). We used the 12-item Chinese version validated by the originator (Pan & Goldberg, 1990). Prior study has demonstrated satisfactory internal consistency, test-retest reliability, concurrent validity, and discriminative validity of the Chinese version (Pan & Goldberg, 1990; Boey, 1999; Cheung, 2002).

The SWLS is commonly used in large-scale surveys to measure happiness and subjective well-being (Diener *et al.* 1985). The 5-item scale has been shown to have satisfactory reliability and validity. The Chinese version has acceptable psychometric properties among the Chinese population (Leung & Leung, 1992).

The CIDI is a comprehensive, fully structured interview designed to make reliable diagnoses (Robins *et al.* 1988). The interview, available in 16 languages, should be administered by trained lay interviewers in large-scale epidemiological studies (Wittchen *et al.* 1991). Satisfactory reliability and validity had been reported cross culturally (Wittchen *et al.* 1991; Rubio-Stipec *et al.* 2004). This report focused on the 1-month diagnoses established with the CIDI interviews. Because of resource limitations, diagnoses that were known to be rare in rural populations were not examined in this study. The included diagnoses were somatoform and dissociative disorders, phobic and other anxiety disorders, depressive disorders and dysthymic disorders, manic and bipolar affective disorders, schizophrenia and other psychotic disorders,

and alcohol use disorders. The omitted diagnoses were: nicotine use disorder, eating disorders, obsessive-compulsive disorder and post-traumatic stress disorder, substance-related disorders, dementia, amnesic and other cognitive disorders, and psychosexual disorders.

The interviewers were trained by an experienced trainer (Y.C.) who was also familiar with the local Shandong culture. The inter-rater reliability was satisfactory ($\kappa = 0.74-0.95$). We had previously evaluated the validity of the CIDI in a rural Chinese population. The results, showing good sensitivity and specificity, will be reported in a forthcoming paper.

Data analyses

We used GHQ scoring (0, 0, 1, 1) to calculate the total score. Internal consistency was measured with item-total correlations and Cronbach's alpha statistics. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and misclassification rate (MR) were calculated for a range of cut-off scores against the 1-month CIDI diagnosis. Receiver operating characteristic (ROC) curve was also constructed, and the areas under the curve (AUC) were measured for the whole study population as well as subpopulations of male, female, older adults, and those with no education. The correlation between GHQ and SWLS was also calculated.

Findings

A total of 2430 individuals were sampled, among whom 342 were found to be working or studying away from home in urban cities. Of the 2088 eligible residents, 25 were unable to participate in the survey due to poor health. Another 634 residents were not at home (e.g. working in the field or studying at school) when the survey team visited the villages, and a further 28 residents declined to participate in the study. Of the 1401 residents (68%) who responded to the survey, the mean age was 45.3 (s.d. = 13.3) years, and 600 were male (43%). Approximately 91% of the respondents were married and 55% had more than 6 years of education. The sociodemographic characteristics of the respondents and non-respondents are summarized in Table 1. Compared with study respondents, the non-respondents were

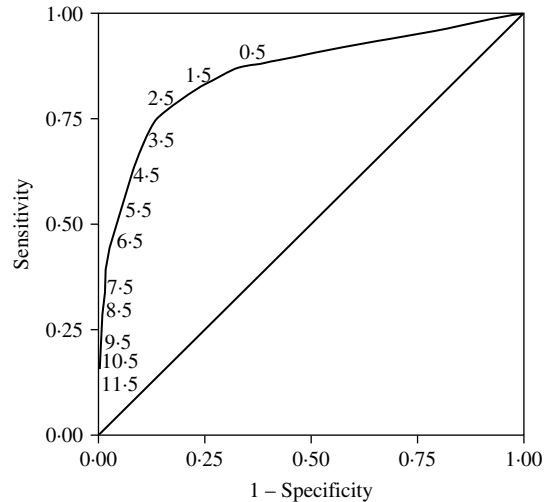


Fig. 1. Receiver operating characteristic (ROC) curve of the GHQ.

more likely to be men, single, younger, better educated, and non-farmer ($p < 0.001$).

The mean GHQ score of the respondents was 1.19 (s.d. = 2.25); and the item-total correlations ranged from 0.48 to 0.64 (Table 2). Out of the 12 items, five had item-total correlations of > 0.55 . The split-half reliability was 0.80 and the alpha coefficient was 0.87.

Of the 1401 respondents, 93 were found to suffer from one or more DSM-IV diagnoses in the month prior to the CIDI interviews, yielding a 1-month prevalence of 6.6% (95% CI 5.4-8.1). The GHQ scores were significantly correlated with the SWLS scores (correlation coefficient -0.26 , 95% CI -0.2 to -0.14).

The psychometric performance of the GHQ at various cut-off points is summarized in an ROC curve (Fig. 1). At a cut-off point of 1/2, the sensitivity was 80.6%, specificity 79.3%, PPV 21.7%, NPV 98.3%, and MR was 20.6%. At a cut-off of 2/3, the corresponding statistics were 75.3%, 86.6%, 28.6%, 98.0%, and 14.1% respectively. The AUC for the whole sample population was 0.86 (s.e. = 0.02), and the AUC for the subsamples of male, female, no education, and > 50 years of age were 0.81, 0.89, 0.85, and 0.88 respectively.

DISCUSSION

There is a lack of simple, inexpensive, easy-to-use, and above all properly validated

Table 1. *Social and demographic characteristics of respondents and non-respondents*

Characteristic	Respondent (N = 1401)		Non-respondent (N = 659)		Analysis		
	n	(%)	n	(%)	χ^2	df	p
Sex					45.4	1	<0.001
Male	600	(42.8)	387	(58.7)			
Female	801	(57.2)	272	(41.3)			
Marital status					872.2	3	<0.001
Single	89	(6.4)	444	(67.4)			
Married	1274	(90.9)	210	(31.9)			
Divorced	5	(0.4)	0	(0)			
Widowed	33	(2.4)	5	(0.8)			
Education (years)					62.3	5	<0.001
Illiterate	392	(28.0)	84	(12.7)			
Primary	368	(26.3)	216	(32.8)			
Junior High	506	(36.1)	277	(42.0)			
Senior High	108	(7.7)	66	(10.0)			
Institute	23	(1.6)	10	(1.5)			
University or above	4	(0.3)	6	(0.9)			
Occupation					812	9	<0.001
Farmer	1108	(79.1)	157	(23.8)			
Worker	61	(4.4)	57	(8.6)			
Party official	19	(1.4)	0	(0)			
Doctor	11	(0.8)	1	(0.2)			
Businessman	36	(2.6)	6	(0.9)			
Small business	30	(2.1)	2	(0.3)			
Working outside	24	(1.7)	28	(4.2)			
Technician	11	(0.8)	4	(0.6)			
Teacher	0	(0)	1	(0.2)			
Other	101	(7.2)	403	(61.2)			
Age (years)					740	2	<0.001
≤30	178	(12.7)	478	(72.5)			
31–50	737	(52.6)	118	(17.9)			
>50	486	(34.7)	63	(9.6)			

Table 2. *Item-total statistics of GHQ items*

Items	Scale mean if item deleted	Scale variance if item deleted	Corrected item- total correlation	Alpha if item deleted
Lost sleep over worry	1.05	4.26	0.53	0.86
Under strain	1.03	4.17	0.55	0.87
Losing confidence	1.13	4.52	0.52	0.86
Face up to problems	1.15	4.58	0.54	0.86
Can't overcome difficulties	1.12	4.44	0.55	0.85
Unhappy and depressed	1.10	4.28	0.64	0.85
Enjoy activities	1.12	4.39	0.59	0.85
Reasonably happy	1.12	4.37	0.62	0.85
Able to concentrate	1.09	4.37	0.53	0.86
Playing useful part	1.13	4.52	0.53	0.86
Worthless person	1.11	4.43	0.54	0.86
Capable of decisions	1.15	4.67	0.48	0.86

psychological measures for population with a rural background. The GHQ is likely to be able to fill this gap. Previous studies have shown that the GHQ is a versatile assessment tool, with good reliability and validity in many urban populations. The present psychometric

evaluation extends the favourable findings to China's rural population.

Our data showed that the Chinese GHQ had good criterion validity and internal consistency. Both cut-offs of 1/2 and 2/3 had good sensitivity and specificity (~80%) in screening for

psychiatric disorders among rural people, but the cut-off of 1/2 is preferable if a higher sensitivity is desired. This is the first report on the sensitivity and specificity of GHQ-12 in identifying psychiatric morbidity among the Chinese population; hence we are not able to compare our findings with other Chinese studies. However, a study in the UK has reported similar sensitivity and specificity for cut-offs of 1/2 and 2/3 (Banks, 1983). We found the PPV of the GHQ comparatively low (29%) in the study population. This shortcoming, however, would disappear if the scale is applied to a population with a higher prevalence of psychiatric disorders (e.g. the PPV would reach 50% if the prevalence is ~15%).

Our ethno-epidemiological findings show that the rural respondents generally prefer interviewer-administered scales to self-report rating scales. Adaptation is hence needed if the latter are to be used in a rural population. The ethnography also shows that household survey is culturally acceptable in the villages, but data confidentiality and researcher independence must be assured.

Our ethnographic respondents had difficulties with the depression and anxiety rating scales. Future studies should examine why rural people have difficulties in responding to specific affective constructs. Do rural people lack ability to introspect their depression and anxiety? Do they lack the language to articulate their inner affective state? Are they just unfamiliar with the Western-derived affective terms? Or do they have a totally different emotive experience and vocabulary that urban researchers missed?

This study has several limitations that warrant discussion. First, we evaluated the GHQ in only one province. Replication studies in other provinces are needed. Second, test-retest reliability was not examined in the present study. Third, around 30% of eligible residents could not be contacted during the daytime hours when the survey team visited the villages. The non-respondents were more likely to be men, younger, better educated, and in a non-agricultural occupation. Whilst our analysis showed that the psychometric performance of the GHQ was not affected by gender, age, and education, the sampling biases should nonetheless be considered in data interpretation. Future studies can reduce the sampling bias by

extending the recruitment into the evenings. It is also important to note that a modified CIDI was used as the gold-standard comparison. As some diagnoses, albeit rare, were not included, this might have caused a slight increase of false-positive rates (or an underestimation of PPV). Further, since uncommon psychiatric disorders were not assessed in the present study, the 1-month prevalence of all DSM-IV diagnoses could be higher than 6.6%.

Last, notwithstanding that the GHQ is a well-tested research tool worldwide, it was originally designed for the British population. The experience and expression of affective distress, however, are shaped and patterned by local norms and culture (Cheng, 2001). A culturally sensitive research tool should hence include items that capture local socio-moral experiences as well as idioms of distress. In the Chinese rural context, this may involve adding items like tiredness (*henlei*), fear of losing face (*diulian*), and distressed by the hardship of life (*xinku*). Only with the incorporation of these native socio-emotive expressions would researchers be able to assess psychological well-being with a local view (*vis-à-vis* a Western view).

We hope this ethno-psychometric evaluation of the GHQ will stimulate more research on rural mental health, especially in middle- and low-income countries (Saraceno & Saxena, 2004). Recent epidemiological studies, including those of Chinese societies, have largely focused on urban populations (see e.g. WHO World Mental Health Survey Consortium, 2004). The present study showed that as many as 6.6% of China's rural population suffer from some form of psychiatric disorder. This estimate is several-fold higher than the 1.2% point prevalence reported by Zhang *et al.* (1998) in the last major rural epidemiological survey. Given that the last major epidemiological study in rural China was conducted 12 years ago (Zhang *et al.* 1998), a new wave of surveys are needed to fathom mental health conditions in the contemporary Chinese rural population.

ACKNOWLEDGEMENTS

Thanks are due to Sarah Chie, Yanlin He, Heidi Tsang, and Sheng Li. The study was supported by the Harvard Asia Center Research Fund.

DECLARATION OF INTEREST

None.

REFERENCES

- Banks, M. H.** (1983). Validation of General Health Questionnaire in a young community sample. *Psychological Medicine* **13**, 349–353.
- Boey, K. W.** (1999). Reliability and validity of GDS and GHQ short form for the aged. *Chinese Journal of Psychiatry* **1**, 41–43.
- Cheng, A. T. A.** (2001). Case definition and culture: are people all the same? *British Journal of Psychiatry* **179**, 1–3.
- Cheung, Y. B.** (2002). A confirmatory factor analysis of the 12-item General Health Questionnaire among older people. *International Journal of Geriatric Psychiatry* **17**, 739–744.
- Diener, E., Emmons, R. A., Larsen, R. J. & Griffin, S.** (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment* **49**, 71–75.
- Furukawa, T. A., Andrews, G. & Goldberg, D. P.** (2002). Stratum-specific likelihood ratios of the General Health Questionnaire in the community: help-seeking and physical co-morbidity affect the test characteristics. *Psychological Medicine* **32**, 743–748.
- Goldberg, D. P. & Williams, P.** (1988). *The User's Guide to the General Health Questionnaire*. NFER-Nelson: Windsor.
- Kleinman, A.** (2000). Social and cultural anthropology: salience for psychiatry. In *New Oxford Textbook of Psychiatry* (ed. M. G. Gelder, J. J. Lopez-Ibor and N. C. Andreasen), pp. 300–304. Oxford University Press: Oxford.
- Leung, J. P. & Leung, K.** (1992). Life satisfaction, self-concept, and relationship with parents in adolescence. *Journal of Youth and Adolescence* **21**, 653–665.
- McDowell, I. & Newell, C.** (1996). Psychological well-being: The General Health Questionnaire. In *Measuring Health: A Guide to Rating Scales and Questionnaires* (2nd edn), pp. 225–236. Oxford University Press: New York.
- Murphy, R. S.** (1998). Rural psychiatry in developing countries. *Psychiatric Services* **49**, 967–969.
- Pan, P. C. & Goldberg, D. P.** (1990). A comparison of the validity of GHQ-12 and CHQ-12 in Chinese primary care patients in Manchester. *Psychological Medicine* **20**, 931–940.
- Robins, L. N., Wing, J., Wittchen, H. U., Helzer, J. E., Babor, T. F., Burke, J., Farmer, A., Jablenski, A., Pickens, R., Regier, D. A., Sartorius, N. & Towle, L. H.** (1988). The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry* **45**, 1069–1077.
- Rubio-Stipec, M., Peters, L. & Andrews, G.** (2004). Test–retest reliability of the computerized CIDI (CIDI-Auto): substance abuse modules. *Substance Abuse* **20**, 263–272.
- Saraceno, B. & Saxena, S.** (2004). Bridging the mental health research gap in low- and middle-income countries. *Acta Psychiatrica Scandinavica* **110**, 1–3.
- WHO World Mental Health Survey Consortium** (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization Mental Surveys. *Journal of the American Medical Association* **291**, 2581–2590.
- Wittchen, H. U., Robins, L. N., Cottler, L. B., Sartorius, N., Burke, J. D. & Regier, D.** (1991). Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI). *British Journal of Psychiatry* **159**, 645–653.
- United Nations Population Division** (2002). The prospects for world urbanization and rural population growth. In *World Urbanization Prospects: The 2001 Revision*. United Nations: USA (<http://www.un.org/esa/population/publications/wup2001/WUP2001report.htm>). Accessed 25 August 2005.
- Zhang, W., Shen, Y. & Li, S.** (1998). Epidemiological investigation on mental disorders in 7 areas of China. *Chinese Journal of Psychiatry* **31**, 69–71.