

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): YONGJIAN HU, SARAH STEWART-BROWN, LIZ TWIGG  
and SCOTT WEICH

Article Title: Can the 12-item General Health Questionnaire be used to measure positive mental health?

Year of publication: 2007

Link to published version: [http://dx.doi.org/  
10.1017/S0033291707009993](http://dx.doi.org/10.1017/S0033291707009993)

Publisher statement: None

# Can the 12-item General Health Questionnaire be used to measure positive mental health?

YONGJIAN HU<sup>1</sup>, SARAH STEWART-BROWN<sup>1</sup>, LIZ TWIGG<sup>2</sup> AND SCOTT WEICH<sup>1\*</sup>

<sup>1</sup> *Health Sciences Research Institute, Warwick Medical School, University of Warwick, Coventry, UK;*

<sup>2</sup> *Institute for the Geography of Health, Department of Geography, University of Portsmouth, Portsmouth, UK*

## ABSTRACT

**Background.** Well-being is an important determinant of health and social outcomes. Measures of positive mental health states are needed for population-based research. The 12-item General Health Questionnaire (GHQ-12) has been widely used in many settings and languages, and includes positively and negatively worded items. Our aim was to test the hypothesis that the GHQ-12 assesses both positive and negative mental health and that these domains are independent of one another.

**Method.** Exploratory (EFA) and confirmatory (CFA) factor analyses were conducted using data from the British Household Panel Survey (BHPS) and the Health Survey for England (HSE). Regression models were used to assess whether associations with individual and household characteristics varied across positive and negative mental health dimensions. We also explored higher-level variance in these measures, between electoral wards.

**Results.** We found a consistent, replicable factor structure in both datasets. EFA results indicated a two-factor solution, and CFA demonstrated that this was superior to a one-factor model. These factors correspond to ‘symptoms of mental disorder’ and ‘positive mental health’. Further analyses demonstrated independence of these factors in associations with age, gender, employment status, poor housing and household composition. Statistically significant ward-level variance was found for symptoms of mental disorder but not positive mental health.

**Conclusions.** The GHQ-12 measures both positive and negative aspects of mental health, and although correlated, these dimensions have some independence. The GHQ-12 could be used to measure positive mental health in population-based research.

## INTRODUCTION

Well-being, and not just the absence of undesirable states such as depression or anxiety, may have consequences for physical health, cognition, relationships and even survival (Danner *et al.* 2001; Huppert & Whittington, 2003; Keyes, 2005). However, evidence remains sparse (Ostir *et al.* 2000; Strandberg *et al.* 2006). This is partly due to the absence of appropriate population-based measures, which in turn reflects the lack of consensus about the most

salient underlying constructs or dimensions (Ryff & Singer, 1998; Ryan & Deci, 2001; Huppert & Whittington, 2003). There are two broad (and complementary) traditions in conceptualizing well-being: the ‘hedonic’ and ‘eudaimonic’ approaches. The hedonic approach emphasizes happiness (pleasant affect, life satisfaction) whereas the eudaimonic approach (with origins in Platonic philosophy) emphasizes optimal psychological and social functioning (or ‘flourishing’).

The 12-item General Health Questionnaire (GHQ-12), a brief self-report measure, has excellent psychometric properties as a screening instrument for psychiatric disorders in non-clinical settings (Goldberg & Williams, 1988). It

\* Address for correspondence: Professor Scott Weich, Health Sciences Research Institute, Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK.  
(Email: s.weich@warwick.ac.uk)

has been used extensively in epidemiological research, with scores taken to indicate the severity of symptoms of the most common mental disorders, anxiety and depression (Weich & Lewis, 1998; Weich *et al.* 2001, 2003). Factor analyses have reported one-, two- and three-factor models (Campbell *et al.* 2003). Many two- and three-factor models describe a pathological dimension (for example ‘anxiety-depression’ or ‘loss of confidence’; Shevlin & Adamson, 2005). Lewis (1992), examining the 30-item GHQ, noted that positive and negative GHQ items had different patterns of factor loading. He found a first (unrotated) general factor on which all items were loaded and a second, bipolar factor with positive and negative mental health at opposite ends. Huppert & Whittington (2003) and others have also suggested that positive mental health and symptoms of mental disorder are independent of one another, to a degree (Watson *et al.* 1988; Keyes, 2005).

Our aim was to test the hypothesis that the GHQ-12 assesses both positive and negative mental health, and that these domains are independent of one another. We set out to achieve this using two large, nationally representative samples of the British population.

## METHOD

### Data sources and measures

Data were taken from the first wave of the British Household Panel Survey (BHPS), recruited in 1991, and the combined 1991 and 1992 samples of the Health Survey for England (HSE). The BHPS is an annual survey of each adult member of a representative sample of private households in England, Wales and Scotland (Taylor *et al.* 2005). The HSE is a series of annual surveys about the health of people in England (OPCS, 1994). Like the BHPS, the HSE samples households from the Postcode Address File (PAF) using a stratified sampling frame. In both surveys, interviews were sought with all adult members of sampled households. We restricted both samples to those aged 16–74.

The 12-item GHQ-12 comprises six ‘positive’ and six ‘negative’ items concerning the past few weeks. Positive items included ‘Have you recently felt capable of making decisions about things?’, while negative items included ‘Have

you recently felt constantly under strain?’ Items were classified in this way according to wording, with positively worded items having responses ‘Better than usual’, ‘Same as usual’, ‘Less than usual’ and ‘Much less than usual’. Responses to negatively worded items are ‘Not at all’, ‘No more than usual’, ‘Rather more than usual’ and ‘Much more than usual’. We regarded questions 1, 3, 4, 7, 8 and 12 as positively worded items. The remainder are negatively worded. Responses were coded using an unweighted four-point Likert scale (0, 1, 2, 3). Positively worded items were later rescored so that a high score was indicative of endorsement of these items (e.g. ‘better than usual’). Higher scores on negative items indicate greater distress and/or difficulty.

### Individual- and household-level risk factors

Age, gender, marital status, ethnicity, education, employment status, financial strain and number of current physical health problems were included as individual-level characteristics. Household-level characteristics were: social class of the head of household, structural housing problems (any major problem, or two or more minor problems such as damp, condensation, leaking roof or rot in wood), low income (household income below half the sample median), household access to a car, housing tenure, overcrowding (more than two household members per bedroom) and household composition.

### Statistical analyses

Exploratory factor analysis (EFA) was carried out to assess whether the GHQ-12 has one or two underlying factors. We used the principal component analysis method to extract factors. The number of factors was determined by the examination of scree plots and the size of eigenvalues. An orthogonal (varimax) rotation was then made to achieve a more readily interpretable factor structure. We chose 0·4 as a cut-off for size of loading to be interpreted. Factor analysis was undertaken using STATA (Stata Corporation, 2004).

We verified the factor structure of the GHQ-12 by creating several split samples from the study dataset (with replacement) (Diamantopoulos & Siguaw, 2000). These randomly drawn datasets were: (1) a sample of 500 observations including 250 women and 250 men; (2) and (3) samples of

500 observations for each sex, separately; (4) a sample of 500 observations from participants aged 40 years and over; and (5) a sample of 500 observations from those below 40 years of age. We first conducted separate EFAs and compared results with those derived from the original sample. We then conducted confirmatory factor analysis (CFA) in each sample. CFA is a modelling technique to investigate whether a particular factor structure is consistent with the correlations or covariance of a set of observed variables (Jacob *et al.* 1998). Hence, the hypothesized factor structure has to be specified *a priori*. The AMOS Graphics (AMOS 6, AMOS Development Corp. SPSS Inc., Chicago, IL, USA) was used. Fit indices of different models were compared. The  $\chi^2$  statistic assesses an overall model fit in covariance structure models, but it is sensitive to sample size. Using a sufficiently large sample tends to yield a significant value of  $\chi^2$  (Jöreskog, 1969; Bentler & Bonett, 1980), which indicates lack of satisfactory fit. Hence, the  $\chi^2$  statistic has limited value (Kaplan, 1990) and was not used. Instead, we used the goodness-of-fit index (GFI), the adjusted GFI (AGFI), the root-mean-square error of approximation (RMSEA) values, and the expected cross-validation index (ECVI). GFI is based on the ratio of the sum of the squared differences between the observed and reproduced matrices to the observed variances and should be  $\geq 0.90$ . AGFI is a variant of GFI that adjusts GFI for degrees of freedom. RMSEA is also known as the discrepancy (between the observed and produced covariance matrices) per degree of freedom. ECVI is used to identify whether a model is likely to cross-validate samples of the same size from the same population (Diamantopoulos & Siguaw, 2000); lower ECVI indicates better fit.

Because extracted factors are based on latent variables, loadings were used to derive factor scores. These factor scores are estimates of values for unobserved random variables (Johnson & Wichern, 1998). Factor scores are obtained using a regression approach, producing the highest correlations between factors and factor scores. Each factor score had a standardized normal distribution with zero mean and unit variance.

Finally, we tested the hypothesis that scores for 'negative' and 'positive' GHQ-12 factors

would show different patterns of association with demographic, socio-economic and contextual risk factors. These analyses were conducted in the full BHPS data set. To reflect the hierarchical structure of these data (individuals nested within household; households within regions, etc.) and to avoid underestimating standard errors, we used multilevel analysis using MLwiN (Snijders & Bosker, 1999; Rasbash *et al.* 2004). We specified three levels, comprising individuals, households and electoral wards (mean population 5222, s.d.=899; the smallest geographical area at which BHPS data were available). Random intercept models were fitted. A null model was first estimated and individual and household's risk factors were added subsequently. Parameters were estimated using Iterative Generalized Least Square (IGLS). Wald tests on both fixed and random parameters were performed. We also reported the 95% interval estimates for each random parameter based on normality assumption and the MCMC estimation. The MCMC method is a Bayesian formulation that puts priors on parameters and generates a correlated chain of sample draws from the posterior distribution of the parameters. When comparing two non-nested models, for example, the model with age being used as a continuous variable and the model with age being used as a categorical variable, we used the Akaike Information Criterion (AIC). The AIC is calculated based on the formula  $l + 2p$ , where  $l$  is the value of  $-2(\log \text{likelihood})$  for each model,  $p$  is the number of parameters fitted in the model. The model with the smallest AIC is deemed to fit best.

## RESULTS

The BHPS sample comprised 9204 participants, of whom 8978 (97.5%) completed the GHQ. Among these, mean age was 41.7 years (s.d.=16.0) and 52.7% ( $n=4735$ ) were female. The HSE sample comprised 6451 participants with a completed GHQ, of whom 52.9% ( $n=3412$ ) were women. Mean age of the HSE sample was 43.0 years (s.d.=16.4). The sample characteristics are shown in Table 1.

EFAs produced almost identical results in the two samples (Table 2). Unrotated factor analysis revealed a first factor on which all negatively worded items loaded at above +0.6. The first

Table 1. Demographic, socio-economic and health-related characteristics of BHPS and HSE samples

	BHPS sample (n=8978)	HSE sample (n=6451)
Age group (years)		
16–35	3683 (41·0)	2467 (38·2)
36–55	3218 (35·8)	2282 (35·4)
≥56	2077 (23·1)	1702 (26·4)
Gender		
Female	4735 (52·7)	3412 (52·9)
Ethnic group		
White	8618 (96·1)	6121 (96·2)
Marital status		
Married/cohabiting	6082 (67·8)	4379 (67·9)
Divorce/separated/widow	1015 (11·3)	779 (12·1)
Single	1870 (20·9)	1293 (20·0)
Education		
No qualifications	2690 (30·1)	2118 (34·6)
Employment		
Employed	5620 (62·6)	3948 (61·2)
Unemployed	521 (5·8)	360 (5·6)
Inactive		
Self-rated health		
Very good/good	6761 (75·4)	5122 (79·5)
Fair	1558 (17·4)	1074 (16·7)
Bad/very bad	654 (7·3)	248 (3·9)
No. of physical health problems		
0	4108 (46·5)	
1	2837 (32·1)	
≥2	1896 (21·5)	
Financial strain		
Comfortable/all right	4943 (55·2)	
Getting by	2777 (31·0)	
Difficult	1237 (13·8)	

Values are n (%).

BHPS, British Household Panel Survey; HSE, Health Survey for England.

two unrotated factors explained 40·3% and 10·9% of the variance in GHQ-12 score in the BHPS sample, and 42·1% and 11·6% of variance in the HSE sample. Inspection of eigenvalues and scree plots indicated two-factor solutions in both samples. Factor loadings after rotation are shown in Table 2.

Table 2 shows that all negatively worded items had a high loading on the first rotated factor, with a smallest value of 0·54. For positively worded items, negative loading values on factor 1 were found, but were much lower than the loadings of negatively worded items. Loadings for item 1 ('able to concentrate on whatever you are doing'), item 7 ('able to enjoy your normal day-to-day activities') and item 12 ('feeling reasonably happy, all things

considered?') on the first factor were all  $\leq -0·40$ , with the exception of item 7 in the HSE sample (where loading on factor 1 was  $-0·38$ ). The first factor therefore appears to measure symptoms of mental disorder, incorporating negative affect (feeling unhappy or depressed; worthlessness, loss of confidence, unable to enjoy normal activities), anxiety (feeling under strain, losing sleep over worry), and impaired mental functioning (poor concentration).

There was also a clear pattern to loadings on the second factor: loadings for all positively worded items had a positive sign, and all negatively worded items had a negative sign. Correlations between factor 2 and some positively worded items were very high: the loadings of items 3, 4 and 8 were above 0·6. By contrast, all negatively worded items had much lower loadings, and all were below 0·4. We judged factor 2 to represent positive mental health, incorporating positive mental functioning (good concentration; playing a useful part in things; facing up to problems; making decisions) and positive affect (feeling happy; enjoying normal daily activities).

Table 3 reports measures of fit for CFA models. The findings from both samples were again highly similar, and we have therefore only reported results from the BHPS sample (HSE results available from the authors on request). Model I has a one-factor structure and assumes that the GHQ-12 is unidimensional. Model II postulates a two-factor solution in which each negatively worded item is specified to load only on the first factor, while each positively worded item depends only on the second factor. The two factors are constrained to have zero covariance. The factor structure in model III is the same as model II, except that the covariance between two factors is free to be estimated. In keeping with the findings of our EFA, model IV was specified in which all six negatively worded items only loaded on the first factor. Among positively worded items, items 1, 7 and 12 were allowed to load on both first and second factors, as EFA showed that the correlation between these three items and the first factor was  $>0·4$ . Likewise, covariance between the two factors was not constrained in model IV. Model IV was superior on all four fit statistics for all samples (Table 3). The values of RMSEA for both

Table 2. Principal components factor analysis of the GHQ-12 using BHPS and HSE data: rotated factor loadings (negatively worded items in bold)

	BHPS		HSE	
	Factor 1	Factor 2	Factor 1	Factor 2
1. Have you recently been able to concentrate on whatever you're doing?	<b>-0.42</b>	<b>0.44</b>	<b>-0.40</b>	<b>0.46</b>
2. <b>Have you recently lost much sleep over worry?</b>	<b>0.74</b>	-0.03	<b>0.73</b>	-0.04
3. Have you recently felt that you were playing a useful part in things?	-0.12	<b>0.68</b>	-0.06	<b>0.71</b>
4. Have you recently felt capable of making decisions about things?	-0.04	<b>0.77</b>	-0.08	<b>0.77</b>
5. <b>Have you recently felt constantly under strain?</b>	<b>0.80</b>	-0.06	<b>0.80</b>	-0.03
6. <b>Have you recently felt you couldn't overcome your difficulties?</b>	<b>0.71</b>	-0.16	<b>0.74</b>	-0.20
7. Have you recently been able to enjoy your normal day-to-day activities?	<b>-0.43</b>	<b>0.49</b>	-0.38	<b>0.54</b>
8. Have you recently been able to face up to problems?	-0.23	<b>0.67</b>	-0.29	<b>0.65</b>
9. <b>Have you recently been feeling unhappy or depressed?</b>	<b>0.77</b>	-0.25	<b>0.80</b>	-0.20
10. <b>Have you recently been losing confidence in yourself?</b>	<b>0.63</b>	-0.38	<b>0.71</b>	-0.33
11. <b>Have you recently been thinking of yourself as a worthless person?</b>	<b>0.54</b>	-0.36	<b>0.61</b>	-0.36
12. Have you recently been feeling reasonably happy, all things considered?	<b>-0.43</b>	<b>0.50</b>	<b>-0.47</b>	<b>0.50</b>

GHQ, General Health Questionnaire; BHPS, British Household Panel Survey; HSE, Health Survey for England.

This form may only be reproduced for use within the purchasing institution within the terms stated in the permission agreement from the publisher. GHQ-12 copyright © David Goldberg, 1978. GHQ-28 copyright © David Goldberg and the Institute of Psychiatry, 1981. GHQ-30 copyright © David Goldberg, 1978. GHQ-60 copyright © David Goldberg and the Institute of Psychiatry, 1978.

Published by nferNelson Publishing Company Ltd, The Chiswick Centre, 414 Chiswick High Road, London W4 5TF, UK. All rights reserved. nferNelson is a division of Granada Learning Ltd, part of Granada plc. To obtain permission to reproduce the GHQ, please email the publisher at permissions@nfer-nelson.co.uk.

models I and II exceeded 0.1, the point at which models should not be accepted. Comparison of GFI across the models demonstrates that the proportion of the sample matrix predicted by the reproduced matrix for models III and IV was greater than 0.90. Although there were marginal differences in the values of GFI between models III and IV, ECVI shows that model IV has smaller values, and therefore had a slightly better fit. Where variance between the two factors was not constrained, correlation coefficients between these were slightly larger in model III (range -0.71 to -0.83) than model IV (-0.56 to -0.73) across the subsamples specified in Table 3.

### Multilevel analyses

In multilevel null models, individual-level variance was similar for scores on symptoms of mental disorder (factor 1) (0.96, s.e.=0.02,  $p<0.001$ ) and positive mental health (factor 2) (0.95, s.e.=0.02,  $p<0.001$ ). Approximately 4% of variance in these scores (0.04, s.e.=0.02,  $p=0.007$ ) was found at household level for symptoms and 5% (0.05, s.e.=0.02,  $p=0.002$ ) for positive mental health, which was statistically significant in both cases. A different pattern was observed for variances at ward level: the estimated variance in scores between wards

was statistically significant for the symptoms factor (0.013, s.e.=0.005,  $p=0.003$ ) but not for positive mental health (0.004, s.e.=0.004,  $p=0.24$ ).

Table 4 shows associations between individual and household characteristics and symptoms (factor 1) and positive mental health (factor 2) scores. Different patterns of association were observed, particularly with age (Fig. 1), gender, ethnicity, employment, financial strain, poor housing and household composition. Female gender was associated with a higher symptoms (F1) score to a statistically significant degree, but not with lower positive mental health (F2). The same was true for non-white ethnicity, moderate financial strain, housing problems, living alone and being a single parent. Even where the association with positive mental health score mirrored that with symptoms, the effect size of the latter was substantially larger. Being retired was associated with significantly lower positive mental health and lower symptoms scores.

### DISCUSSION

We found a consistent, replicable factor structure to the GHQ-12, using EFA and CFA in two large, nationally representative datasets.

**Table 3. Measures of fit from confirmatory factor analysis (CFA), using five split samples within the BHPS dataset**

Model specification	Sample <sup>a</sup>	Fit indices			
		GFI	AGFI	RMSEA	ECVI
Model I (one factor)	1	0.905	0.863	0.094	0.682
	2	0.893	0.846	0.097	0.718
	3	0.881	0.829	0.103	0.783
	4	0.875	0.819	0.106	0.815
	5	0.894	0.847	0.097	0.723
Model II (two factors) <sup>b</sup>	1	0.870	0.813	0.128	1.086
	2	0.876	0.821	0.121	1.003
	3	0.858	0.795	0.137	1.219
	4	0.881	0.828	0.115	0.919
	5	0.875	0.819	0.123	1.026
Model III (two factors) <sup>c</sup>	1	0.926	0.891	0.079	0.541
	2	0.924	0.888	0.078	0.537
	3	0.905	0.861	0.091	0.650
	4	0.916	0.877	0.084	0.580
	5	0.923	0.886	0.080	0.553
Model IV (two factors) <sup>d</sup>	1	0.930	0.890	0.079	0.531
	2	0.935	0.899	0.074	0.493
	3	0.916	0.869	0.088	0.603
	4	0.922	0.878	0.084	0.570
	5	0.927	0.886	0.081	0.546

BHPS, British Household Panel Survey; GFI, goodness-of fit index; AGFI, adjusted goodness-of fit index; RMSEA, root-mean-square error of approximation; ECVI, expected cross-validation index.

<sup>a</sup> Sample 1: 250 men + 250 women; Sample 2: 500 men; Sample 3: 500 women; Sample 4: 500 individuals  $\leq 40$  years of age; Sample 5: 500 individuals  $> 40$  years of age.

<sup>b</sup> Model II: two factors: negatively worded items load only on first factor, positively worded items only on second factor; zero covariance between factors (constrained).

<sup>c</sup> Model III: same as Model II, but covariance free to be estimated.

<sup>d</sup> Model IV: All six negatively worded items only loaded on the first factor. Positively worded items all load on factor 2. In addition, items 1, 7 and 12 (positively worded) allowed to load on both factors. Covariance not constrained.

Although all studies have shown that the first, general factor explains over 40% of the variance in GHQ-12 scores, findings about the precise structure of this measure differ (Campbell *et al.* 2003; Shevlin & Adamson, 2005). Previous studies have reported one- (Lewis, 1992), two- (Politi *et al.* 1994; Kalliath *et al.* 2004) and three-factor solutions (Cheung, 2002; Campbell *et al.* 2003; Shevlin & Adamson, 2005). Some of this apparent inconsistency may be methodological in origin, including differences in setting, sample size and composition, weighting of item scores and methods of analysis. Studies using non-Likert scoring and/or EFA are prone to divergent findings. The largest number of studies (including Graetz, 1991; Politi *et al.* 1994;

Martin, 1999; Campbell *et al.* 2003; Shevlin & Adamson, 2005) report similar factor structures to ours, namely one factor on which all or nearly all of the positively worded items have high loadings (often referred to as ‘social dysfunction’), and either one or two factors on which the negatively worded items have high loadings (labelled ‘anxiety-depression’, ‘dysphoria’ or ‘loss of confidence’). As in our findings, previous studies typically find inter-factor correlations of  $\geq 0.6$  (Cheung, 2002; Kalliath *et al.* 2004; Shevlin & Adamson, 2005).

Campbell *et al.* (2003) used CFA to compare seven models using data from 500 primary care attendees in Tasmania and original investigators’ scoring methods. No model achieved ‘good’ or ‘better than good’ fit, but all fitted equally well. When Likert scoring was used, the best fit was achieved by a three-factor model in which many items loaded on two factors. These factors were described as ‘dysphoria’ (or ‘anxiety and depression’), ‘social functioning’ and ‘loss of confidence’. A similar factor structure was reported by Shevlin and Adamson (2005) using data from a general population survey in Northern Ireland. In the model preferred by Shevlin and Adamson (2005), ‘difficulty with concentration’ (item 1), ‘(not) enjoying normal activities’ (item 7) and ‘(not) feeling reasonably happy’ (item 12) all have high positive loadings on the second factor, entitled (perhaps erroneously) ‘social dysfunction’. These are the same items that had high loadings on both of the factors identified in the present study, and were allowed to load on both in our CFAs. Our factor descriptions may have greater face validity, but our two-factor model is not substantially different to the three-factor model proposed by Shevlin and Adamson (2005).

We did not test a three-factor solution as this was not supported by the results of our EFAs. CFAs demonstrated the superiority of two-factor models compared with a one-factor solution. Goodness of fit improved markedly when the two factors were permitted to covary, and correlations were of the order of 0.6 or greater. Although the model with the best goodness of fit allowed three positively worded items (concentration, enjoyment of normal activities, and feeling happy) to load on both factors, this had little advantage compared to the model in which

Table 4. Associations between individual- and household-level characteristics and symptoms of mental disorder (F1) and positive mental health (F2) scores (BHPS data)

	n	Symptoms (F1)		Positive mental health (F2)	
		B (s.e.)	p	B (s.e.)	p
Age (v. > 55 years)					
16–35 years	3683	0.20 (0.04)	<0.001	0.20 (0.04)	<0.001
36–55 years	3218	0.24 (0.04)	<0.001	0.08 (0.04)	0.04
Females	4735	0.18 (0.02)	<0.001	-0.02 (0.02)	0.45
Non-white ethnicity	350	0.12 (0.05)	0.02	0.09 (0.05)	0.09
Marital status (v. married/cohabited)					
Divorced/separated/widow	1015	0.07 (0.04)	0.14	0.06 (0.05)	0.23
Single (never married)	1870	-0.07 (0.03)	0.02	0.12 (0.04)	<0.001
No educational qualifications	2690	-0.05 (0.02)	0.04	-0.10 (0.03)	<0.001
Employment status (v. employed)					
Unemployed	521	0.06 (0.05)	0.18	-0.38 (0.05)	<0.001
Retired	1148	-0.18 (0.04)	<0.001	-0.19 (0.04)	<0.001
Economically inactive	2837	0.06 (0.03)	0.04	-0.16 (0.03)	<0.001
Financial strain (v. comfortable)					
Doing all right	2546	0.08 (0.03)	0.002	0.01 (0.03)	0.62
Just about getting by	2777	0.35 (0.03)	<0.001	-0.06 (0.03)	0.05
Finding it difficult	1237	0.76 (0.04)	<0.001	-0.23 (0.04)	<0.001
Physical health problems (v. 0)					
1	2837	0.20 (0.02)	<0.001	-0.04 (0.02)	0.10
2	1234	0.27 (0.03)	<0.001	-0.13 (0.03)	<0.001
3	463	0.47 (0.05)	<0.001	-0.28 (0.05)	<0.001
4 or more	199	0.69 (0.07)	<0.001	-0.53 (0.07)	<0.001
Low household income	1707	0.05 (0.04)	0.22	0.05 (0.04)	0.21
Poor housing	2430	0.12 (0.02)	<0.001	-0.004 (0.02)	0.86
No car access	531	-0.05 (0.03)	0.10	0.03 (0.03)	0.37
Living alone	973	0.15 (0.05)	<0.001	-0.08 (0.05)	0.11
Single parent	344	0.21 (0.06)	<0.001	-0.006 (0.06)	0.92

BHPS, British Household Panel Survey; s.e., standard error.

these items were loaded only on the second factor. We suggest that these two factors are best described as symptoms of mental disorder (encompassing negative affect, anxiety and impaired mental functioning) and positive mental health (covering positive mental functioning and positive affect).

Although the utility of correlated GHQ factors as subscales is potentially limited (Shevlin & Adamson, 2005), our analyses show that the two factors studied here differ from one another to a significant degree in their associations with socio-economic and demographic variables. Like Huppert and Whittington (2003), we found that associations with age were different for positive mental health and symptoms of mental disorder. Positive mental health showed a consistent decline with age for men and women, while the association with symptoms of mental disorder defied simple description. Unemployment was associated with positive mental health

to a greater extent than with symptoms of mental disorder, also in keeping with previous findings (Huppert & Whittington, 2003). Positive mental health was most strongly associated with younger age, being in work, being single, not having any financial strain and having no or few physical health problems. A different picture emerged when considering those with fewer symptoms of mental disorder, who were more likely to be male, older, white, neither a single parent nor living alone, with fewer physical health problems and not experiencing financial strain. This supports the view that 'symptoms' and 'positive mental health' scores are not merely mirror images of each other.

This is the first study of which we are aware to estimate variance in positive mental health at distinct levels within a multilevel model. Around 95% of variance in both positive mental health and symptoms of mental disorder occurred at the individual level, with a further 4% occurring

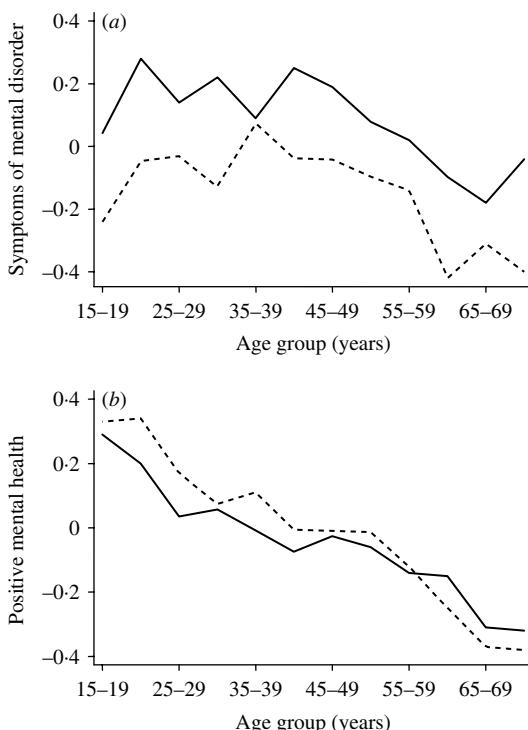


Fig. 1. Factor scores of (a) symptoms of mental disorder (F1) and (b) positive mental health (F2) scores, by gender [British Household Panel Survey (BHPS) data]. —, Men; —, women.

between households. Previous findings based on 1991 BHPS data estimated that 0.9% (95% CI 0.01–1.8,  $p=0.05$ ) of variance in total GHQ-12 score occurred between wards (Weich *et al.* 2003). We found statistically significant between-ward variance in symptoms score (1.3% of total variance) but not in positive mental health score (0.4% of variance).

### Strengths and weaknesses

An important strength was use of both BHPS and HSE datasets: large, representative population-based samples with high response rates. These were large enough to demonstrate consistent findings in CFAs across subsamples.

A further strength was scoring the GHQ-12 using an unweighted Likert scale. We avoided *a priori* assumptions about the relationship between individual responses and underlying GHQ-12 structure (Huppert & Whittington, 2003). In keeping with Huppert and Whittington (2003), we reversed the scoring of positively

worded items such that ‘more than usual’ was scored more highly than ‘less than usual’. By contrast, traditional scoring rates the absence of positive feelings as equivalent to symptoms of mental disorder. We repeated our analyses using ‘original’ GHQ-12 scores; that is, positively worded items scored 0, 1, 2, 3 (from ‘more than usual’ to ‘less than usual’). EFAs indicated that these remained identical numerically to those shown in Table 1, but all became positive in sign. We found no evidence of any substantial effect on results shown in Tables 3 or 4 or Fig. 1. Direction of association with the ‘new’ F2 (positive mental health) score was reversed, such that higher scores reflected a reduction in positive mental health (because ‘less than usual’ attracted a higher score than ‘more than usual’ on items such as ‘able to enjoy normal daily activities’). Our findings were not due to artefact arising from the chosen scoring method. Finally, standard EFA implicitly assumes that the observed variables represent measurements that are, at least approximately (like the GHQ), on an interval scale (Everitt, 1984).

One weakness lies in the GHQ-12 itself. All self-report questionnaires are prone to method variance, namely the tendency for people to respond the same way to similarly worded items. This may contribute to the aggregation of responses to positively and negatively worded items. Although statistical methods exist for distinguishing between this and ‘true’ variance due to latent constructs, these methods are not applicable where there are only one or two construct factors. While a one-factor solution with positive and negative methods factors might have worked equally well (or better), it was not possible to test this empirically.

We have not formally established the concurrent validity of the proposed GHQ positive mental health scale. This would require testing whether the ‘positive’ items on the GHQ-12 do indeed correlate with items of positive mental health or positive emotionality on other personality scales, which was beyond the scope of the present study. Kallith *et al.* (2004) reported similar positive and negative factors (although restricted to four items each) and found that scores on both were correlated to a statistically significant, and similar, degree with measures of employment and family well-being.

## CONCLUSIONS

The best description of the factor structure of the GHQ-12 is given by a variant of a two-factor solution, corresponding to positive mental health and symptoms of mental disorder. A more precise specification reveals that the former encompasses both positive affect and positive mental functioning. This dimension has been shown to exhibit substantial, clinically meaningful independence from measures of symptoms (the more commonly measured, pathological, 'negative mental health' dimension). The present study provides support for the view that valid measures of positive mental health should include items that assess both the hedonic and eudaimonic domains – positive affect as well as positive social and mental functioning. While a two-dimensional model may not offer much of an advantage over a one-factor model when screening for psychiatric disorders, the former approach may come into its own in studying the determinants and consequences of well-being.

## DECLARATION OF INTEREST

None.

## REFERENCES

- Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* **88**, 588–606.
- Campbell, A., Walker, J. & Farrell, G. (2003). Confirmatory factor analysis of the GHQ-12: can I see that again? *Australian and New Zealand Journal of Psychiatry* **37**, 475–483.
- 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.
- Danner, D. D., Snowdon, D. A. & Friesen, W. V. (2001). Positive emotions in early life and longevity: findings from the nun study. *Journal of Personality and Social Psychology* **80**, 804–813.
- Diamantopoulos, A. & Siguaw, J. A. (2000). *Introducing Lisrel: A Guide for the Uninitiated*. Sage: London.
- Everitt, B. S. (1984). *An Introduction to Latent Variable Models*. Chapman & Hall: London.
- Goldberg, D. P. & Williams, P. (1988). *The User's Guide to the General Health Questionnaire*. NFER-Nelson: Windsor.
- Graetz, B. (1991). Multidimensional properties of the General Health Questionnaire. *Social Psychiatry and Psychiatric Epidemiology* **26**, 132–138.
- Huppert, F. A. & Whittington, J. E. (2003). Evidence for the independence of positive and negative well-being: implications for quality of life assessment. *British Journal of Health Psychology* **8**, 107–122.
- Jacob, K. S., Everitt, B. S., Patel, V., Weich, S., Araya, R. & Lewis, G. H. (1998). The comparison of latent variable models of non-psychotic psychiatric morbidity in four culturally diverse populations. *Psychological Medicine* **28**, 145–152.
- Johnson, R. A. & Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*. Prentice-Hall International: London.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* **34**, 183–202.
- Kallith, T. J., O'Driscoll, M. P. & Brough, P. (2004). A confirmatory factor analysis of the General Health Questionnaire-12. *Stress and Health* **20**, 11–20.
- Kaplan, D. (1990). Evaluating and modifying covariance structure models: a review and recommendation. *Multivariate Behavioral Research* **25**, 137–155.
- Keyes, C. L. M. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology* **73**, 539–548.
- Lewis, G. (1992). Dimensions of neurosis. *Psychological Medicine* **22**, 1011–1018.
- Martin, A. J. (1999). Assessing the multidimensionality of the 12-item General Health Questionnaire. *Psychological Reports* **84**, 927–935.
- OPCS (1994). *Health Survey for England. A Survey Carried Out by the Social Survey Division of OPCS on Behalf of the Department of Health*. TSO: London.
- Ostir, G. V., Markides, K. S., Black, S. A. & Goodwin, J. S. (2000). Emotional well-being predicts subsequent functional independence and survival. *Journal of the American Geriatrics Society* **48**, 473–478.
- Politi, P. L., Piccinelli, M. & Wilkinson, G. (1994). Reliability, validity and factor structure of the 12-item General Health Questionnaire among young males in Italy. *Acta Psychiatrica Scandinavica* **90**, 432–437.
- Rasbash, J., Steele, F., Browne, W. & Prosser, B. (2004). *A User's Guide to MLwiN*. Institute of Education: London.
- Ryan, R. M. & Deci, E. L. (2001). On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology* **52**, 141–166.
- Ryff, C. D. & Singer, B. (1998). The contours of positive mental health. *Psychological Inquiry* **9**, 1–28.
- Shevlin, M. & Adamson, G. (2005). Alternative factor models and factorial invariance of the GHQ-12: a large sample analysis using confirmatory factor analysis. *Psychological Assessment* **17**, 231–236.
- Snijders, T. A. B. & Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage: London.
- Stata Corporation. (2004). *Stata Statistical Software: Release 8.0*. Stata Corporation: College Station, Texas.
- Strandberg, T. E., Strandberg, A. Y., Pitkala, K. H., Salomaa, V. V., Tilvis, R. S. & Miettinen, T. A. (2006). Cardiovascular risk in midlife and psychological well-being among older men. *Archives of Internal Medicine* **166**, 2266–2271.
- Taylor, M., Brice, J., Buck, N. & Prentice-Lane, E. (2005). *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices*. University of Essex: Colchester.
- Watson, D., Clark, L. A. & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology* **97**, 346–353.
- Weich, S., Holt, G., Twigg, L., Lewis, G. & Jones, K. (2003). Geographical variation in the prevalence of common mental disorders in Britain: a multilevel investigation. *American Journal of Epidemiology* **157**, 730–737.
- Weich, S. & Lewis, G. (1998). Poverty, unemployment and common mental disorders: population-based cohort study. *British Medical Journal* **317**, 115–119.
- Weich, S., Lewis, G. & Jenkins, S. P. (2001). Income inequality and the prevalence of common mental disorders in Britain. *British Journal of Psychiatry* **178**, 222–227.