Article:
Webb, EJD orcid.org/0000-0001-7918-839X (2023) An Item-Response Mapping from General Health Questionnaire Responses to EQ-5D-3L Using a General Population Sample from England. Applied Health Economics and Health Policy, 21. pp. 327-346. ISSN 1175-5652

https://doi.org/10.1007/s40258-022-00767-4

An item-response mapping from General Health Questionnaire responses to EQ-5D-3L using an England general population sample

Edward J.D. Webb*†

November 22, 2022

Abstract

Background The 12-item General Health Questionnaire (GHQ-12) is widely used to measure mental health and well-being. However, it is not possible to estimate values on the full-health=1, dead=0 scale used to construct quality-adjusted life-years (QALYs) from GHQ-12 responses as it is not preference-based.

Aim Creating an item-response mapping between GHQ-12 and EQ-5D-3L health states, for which several valuesets exist.

Methods Data from the 2012 Health Survey for England with complete GHQ-12 and EQ-5D-3L descriptive system responses were used for analysis. Data were split 70/30 into estimation/ test samples. Four modelling approaches, with EQ-5D-3L levels on each dimension as dependent variables, and GHQ-12 responses as independent variables were assessed: non-parametric, simple ordered logit (OL), extended OL, and least absolute shrinkage and selection operator (LASSO). Approaches were assessed using: Akaike and Bayesian information criteria, predictive accuracy measured using root mean squared error (RMSE) and simplicity.

Results A total of 8,114 responses became 6,924 after discarding missing values, with 4,847 used in estimation and 2,077 for testing. LASSO had better model fit on the pain/discomfort dimension, but no model had markedly superior predictive accuracy. The non-parametric approach was chosen for the mapping algorithm based on simplicity. Predicted and observed EQ-5D-3L values for the test sample had a correlation of 0.488. Prediction accuracy was better for GHQ-12 scores below 20 than above 20.

Conclusion The mapping allows EQ-5D-3L responses to be predicted using GHQ-12 responses, which may be useful in estimating utility values and QALYs. An R script and Excel spreadsheet are provided to facilitate calculations.

*Corresponding author: e.j.d.webb@leeds.ac.uk, Worsley Building, Clarendon Way, Leeds, LS2 9NL, UK
†Academic Unit of Health Economics, Leeds Institute of Health Sciences, University of Leeds

Keywords: General Health Questionnaire-12; EQ-5D-3L; mapping; LASSO

Key points for decision makers:

- The paper constructs a mapping algorithm between 12 item General Health Questionnaire (GHQ-12) responses and EQ-5D-3L
- The mapping algorithm can be used to predict EQ-5D-3L responses from GHQ-12 responses, which may be useful in estimating values on the full health=1, dead=0 scale and hence quality-adjusted life years.

1 Introduction

The General Health Questionnaire (GHQ) is a survey which assesses mental well-being, and was originally designed as a screening tool in mental health diagnosis [1, 2]. There are different versions, of which the 12-item (GHQ-12) is most common [3, 4, 5]. It is widely used in both diagnostic and clinical settings, and as a measure of population health [6, 7].

Participants answering the GHQ-12 are shown 12 items about their experiences over the past few weeks, including "Have you been able to concentrate well on what you were doing?"; "Have you felt that you were playing a useful role in life?"; and "Have you been able to enjoy your normal daily activities?". They are asked how frequently they have experienced each item on a four-point scale. The full GHQ-12 questionnaire is given in the appendix. Responses may be scored in one of two ways. GHQ-12 caseness is calculated by scoring the two responses indicating few problems as 0 and the two responses indicating many problems as 1, and summing the score for all items, giving a score from 0 to 12. With GHQ-12 score (or Likert), each response is given a score from 0 to 3, with higher scores indicating more severe problems. The score for each item is then summed, giving a total score from 0 to 36 [8].

GHQ-12 responses can be compared between individuals and groups of individuals. However, in health economics it is often desirable to compare individuals' and groups' health-related quality of life using preference-based generic instruments not specific to any particular condition. Such generic instruments can have health utilities attached on a scale with full health defined to have a value of 1 and dead defined to be 0. Health utilities derived from preference-based measures can also be used to construct quality-adjusted life-years (QALYs) by multiplying the value of a health state by the time spent in it. QALYs can be used in health technology assessment [9] as well as assessing the impact of interventions.

It is not possible to estimate health utilities from GHQ-12. In such circumstances, a popular approach is to use a mapping to a preference-based generic health survey instrument for which relevant valuesets are

available [10, 11, 12]. This study creates a mapping between GHQ-12 and EQ-5D-3L using an England general population sample.

EQ-5D-3L is a widely used health-related quality of life instrument [13], and a copy is given in the appendix. It measures respondents' health on five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. On each dimension, respondents indicate if they have no (level 1), some (level 2), or severe (level 3) problems. The responses on each dimension can then be combined to form an EQ-5D-3L health state e.g. 11111 represents full health and 33333 represents severe problems on all dimensions. Several EQ-5D-3L valuesets exist [14], including the one currently recommended by the National Institute of Health and Care Excellence (NICE) to assess whether the NHS should provide new health technologies [15]. It thus represents an ideal target measure for a GHQ-12 mapping.

The Health Economics Research Centre database of mapping studies [10] was used to identify two other studies mapping GHQ-12 responses to EQ-5D-3L values: Lindkvist and Feldman [16] who used a Swedish population, and Serrano-Aguilar et al. [17] who used a population from the Canary Islands. Neither is hence ideal for use in England or the rest of the UK as health-related views and values may differ in different populations. The GHQ-12-EQ-5D-3L mapping constructed in this study will thus enable utility values and QALYs to be estimated for the English population. We comment on its potential use in other populations in section 4.

2 Methods

When creating a mapping, it is necessary to decide which measure from the surveys to use as input/output. For example, when mapping another survey instrument to EQ-5D-3L it would be possible to map to health states, or to the utilities associated with the health states. Here, the aim is to create an item-response mapping, i.e. the goal is estimating what EQ-5D-3L health state people occupy, rather than utilities. The reason is that health-state utilities are dependent on the valueset used to calculate them, so utility mappings are limited to estimating only values from a particular valueset. With an item-response mapping, once the health state an individual occupies is predicted, any desired valueset can be used to calculate the utility associated with that state.

2.1 Data

Data came from the 2012 Health Survey for England (HSE) survey [18]. HSE is a large annual cross-sectional survey of representative sample of England's population. Participants answer questions about their circumstances and health, with different question modules included in different years. Individual weights are

included in the dataset which correct for non-response and improve the sample's representativeness.[1]

Data from 2012 was used as it is the only year in which participants self-completed GHQ-12 and EQ-5D-3L. Responses from participants aged 18 and over with complete responses to GHQ-12 and the EQ-5D-3L descriptive system were included in the analysis.[2] The data was randomly divided in a 70/30 split, with the larger sample used for estimation and the smaller sample used to test prediction accuracy.

2.2 Analysis

Analysis was carried out using R version 4.2.0 and significance was judged at the 5% level throughout.

Four different models were created using different approaches.

Model 1: non-parametric

Separate non-parametric transition matrices were created between total GHQ-12 score and each EQ-5D-3L dimension. The transition probability from GHQ-12 score n to EQ-5D-3L level m was calculated as the proportion of respondents with score n in level m, weighted using HSE weights.

Explicitly, if $\hat{\boldsymbol{\ell}}_{di}$ is a vector of length 3 giving the predicted probability that individual i is in each EQ-5D-3L level of dimension d, then

$$\hat{\boldsymbol{\ell}}_{di} = T_d \mathbf{s}_i \tag{1}$$

where T_d is a 3×36 transition probability matrix and $\mathbf{s}_i = (s_{i1}, \ldots, s_{36})$ is a vector with $s_{in} = 1$ if i's GHQ-12 score is n and 0 otherwise. Let $\boldsymbol{\ell}_{di} = (\ell_{di1}, \ell_{di2}, \ell_{di3})$ represent i's EQ-5D-3L level for dimension d, with $\ell_{dim} = 1$ if i is in level m and 0 otherwise. The transition matrix elements are then

$$T_{dmn} = \sum_{\ell_{dim}=1, s_{in}=1} \frac{w_i}{N_w} \tag{2}$$

where $N_w = \sum_i w_i$ is the effective sample size.

Model 2: Simple ordered logit

Five weighted ordered logit (OL) models were created with individual EQ-5D-3L dimensions as dependent variables and scores from 0 to 3 for each GHQ-12 item as (continuous) independent variables.

Explicitly, let $P_{dim} = \Pr(\ell_{dim} = 1 | \boldsymbol{\beta}_d)$ be the probability that individual i is in level m on EQ-5D-3L dimension d, conditional on a vector of model parameters $\boldsymbol{\beta} = (\beta_{d1}, \ldots, \beta_{d12})$. In the simple OL models, this

[1] For more information see section 7.2.6. of Bridges et al. [19].
[2] I.e. responses with missing EQ-VAS were not discarded.

probability is

$$P_{dim} = \Lambda^{-1}\left(\sum_{q=1}^{12}\beta_{dq}g_{qi} - \tau_{m-1}\right) - \Lambda^{-1}\left(\sum_{q=1}^{12}\beta_{dq}g_{qi} - \tau_{m}\right) \tag{3}$$

where $g_{qi} \in \{0, 1, 2, 3\}$ gives i's response to GHQ-12 item q, $\Lambda^{-1}(\cdot)$ is the inverse logistic function and the τ_js are a set of four thresholds with $\tau_{m+1} > \tau_m$, $\tau_0 = -\infty$, $\tau_3 = +\infty$.

If $\mathbf{P}_{di} = (P_{di1}, P_{di2}, P_{di3})$ is a vector of i's probabilities, then the likelihood function may be written $\mathcal{L}_d = \prod_i w_i \mathbf{P}_i \boldsymbol{\ell}_{di}$. Maximum likelihood estimation was performed for five models, one for each dimension, using the polr command from the MASS package [20].

Model 3: Extended ordered logit

Model 3 allowed a more flexible relationship between GHQ-12 items and EQ-5D-3L dimensions. For each GHQ-12 item, dummy variables were included for each response from 1 to 3, with 0 used as the baseline. Each model thus had $3 \times 12 = 36$ parameters. The probability that individual i is in EQ-5D-3L level m on dimension d is now

$$P_{idm} = \Lambda^{-1}\left(\sum_{q=1}^{12}\sum_{n=1}^{3}\beta_{dqn}g_{qni} - \tau_{m-1}\right) - \Lambda^{-1}\left(\sum_{q=1}^{12}\sum_{n=1}^{3}\beta_{dqn}g_{qni} - \tau_{m}\right) \tag{4}$$

where $q_{qni} = 1$ if i responded n to GHQ-12 item q and 0 otherwise. The likelihood function follows as above. Five models, one for each dimension were estimated via maximum likelihood using the polr command from the MASS package.

Model 4: LASSO

Models 2 and 3 assume additive separability of GHQ-12 items. A further set of models was created which relaxed this assumption. A parameter was included for each GHQ-12 item, treated as continuous as in model 2, and also for every two-way interaction between items (again, with the items treated as continuous). There are 66 two-way interactions (11 for item 1, 10 for item 2, etc.), which added to the 12 main effects means 78 parameters per model. The probability that individual i is in EQ-5D-3L level m on dimension d is now

$$P_{dim} = \Lambda^{-1}\left(\sum_{q=1}^{12}\beta_{dq}g_{qi} + \sum_{q=1}^{11}\sum_{r=q+1}^{12}\beta_{dqr}g_{qi}q_{ri} - \tau_{m-1}\right) - \Lambda^{-1}\left(\sum_{q=1}^{12}\beta_{dq}g_{qi} + \sum_{q=1}^{11}\sum_{r=q+1}^{12}\beta_{dqr}g_{qi}q_{ri} - \tau_{m}\right) \tag{5}$$

from which the likelihood follows as previously.

Including many parameters can result in overfitting, in that the model predicts estimation data responses well, but performs poorly outside that [21]. To mitigate this possibility, a machine learning technique termed

the Least Absolute Shrinkage and Selection Operator (LASSO) was used [22]. LASSO is a penalised regression approach, i.e. a penalty term is added to the likelihood function. The penalty is $-\lambda_d \sum |\boldsymbol{\beta}_d|$, where $\lambda_d \geq 0$ is a constant determining how severe the penalty is. As the penalty increases in the number of non-zero model coefficients, including more parameters does not necessarily improve likelihood, so LASSO selects the most relevant independent variables, while setting the others to 0. LASSO can result in non-zero parameters of very small magnitude which would be impractical to use in a mapping algorithm. Thus any parameter with an absolute magnitude below 1×10^{-6} was set to 0 post estimation.

The ordinalNet package [23] was used to estimate LASSO models. For each model, a sequence of 100 candidate λs were tried, decreasing in a uniform log-linear fashion from the largest λ such that all coefficients were 0 to 0.01 times that value. The final model for each EQ-5D-3L dimension was selected as the one minimising the Akaike information criterion (AIC) [24].

For models 2-4, AIC and the Bayesian information criterion (BIC) [25] were calculated. For all models, the root mean squared error (RMSE) of predicted values was calculated for both the analysis and test samples.

The final model to use for mapping was chosen pragmatically using the following criteria:

- Model fit, measured by AIC and BIC.
- Predictive accuracy in the estimation and test samples, measured by RMSE.
- Practicality of use. A more complex mapping algorithm would be more difficult to use in practice, thus favour was shown towards simpler models. The ranking of simplicity went from model 1 being the simplest, followed by model 2, then model 3, with model 4 the most complex.

No particular priority was assigned a priori to the selection criteria. While many alternative measures of model fit and predictive accuracy exist, the above were chosen due to being standard measures commonly used in the literature.

After selecting the final model to use for the mapping, EQ-5D-3L values were calculated for the estimation and test samples. Values were calculated using their observed EQ-5D-3L responses as well as responses predicted from GHQ-12 responses using the mapping algorithm. Observed and predicted values were compared, including mean values for participants above and below a clinically relevant GHQ-12 threshold. The threshold is based on GHQ-12 caseness, where a score of 4 or more is often used to indicate possible mental health problems [26, 8]. Values were calculated using the Measuring and Valuing Health Group (MVH) valueset [27] which is recommended for use by NICE [15].

3 Results

Table 1 summarises all respondents, as well as those included in the analysis, estimation and test samples. There were 8,114 responses, of which 1,190 (14.7%) had a missing response to at least one GHQ-12 or EQ-5D-3L descriptive system question and were excluded. Few differences were observed between the full and analysis samples, and differences were also minimal between the estimation sample of 4,847 individuals, and the test sample of 2,077 individuals.

Table 2 summarises GHQ-12 responses. The modal score for every item was 1, with the exception of losing confidence, and thinking of self as worthless, where the most common was 0. Few respondents gave a score of 3 for any item, with the highest percentage seen for lost much sleep, at 3.0%. Average total scores were 11.0 for the estimation and 10.9 for the test samples, and few differences were seen for any item or total score between either sample, or all responses.

Table 3 summarises EQ-5D-3L responses. As is common [28], there were a large proportion of level 1 responses, and over half of respondents were in state 11111. Few participants reported being in level 3. EQ-5D-3L level sum is useful for summarising individuals' EQ-5D-3L without using a specific valueset [29]. It is constructed as the sum of the levels on each dimension minus 5 so that it ranges from 0 for state 11111, to 10 for state 33333. Mean level sum was 1.02 in the estimation and 0.952 in the test samples.

Figure 1 shows bubble plots of GHQ-12 score against EQ-5D-3L dimensions and a bar plot of mean level sum against GHQ-12 score. A correlation matrix of EQ-5D-3L dimensions and GHQ-12 items is given in the appendix. Many respondents reported a high GHQ-12 score, but level 1 on any given EQ-5D-3L dimension.

Table 4 presents the results of model 1. Higher GHQ-12 scores imply greater transition probabilities to more severe EQ-5D-3L levels, though the effect is not always linear. For GHQ-12 scores below 5, transition to level 1 is almost certain for all dimensions. Even with high GHQ-12 scores there are relatively high transition probabilities to level 1 for all dimensions except anxiety/depression. The transition probability to mobility level 3 is never greater than 0.1. RMSEs were highest for pain/discomfort (0.64 estimation, 0.63 test samples) and lowest for self-care (0.25 estimation, 0.24 test samples).

Table 5 shows results for model 2. Contrary to expectations, not all coefficients were positive, although negative coefficients tended to be insignificant. RMSEs were similar to model 1.

Table 6 presents results for model 3. Most parameters were positive, as expected, and also had the expected ordering (i.e. greater coefficients for higher item scores). However, there are statistically significant exceptions. RMSEs are similar to models 1 and 2.

Table 7 gives the results from model 4. Plots of AIC and the number of non-zero parameters against the log of the λ_d shrinkage parameters are given in the appendix. The signs of all non-zero main effects but one

were positive, in line with expectations. Only three GHQ-12 items had non-zero main effects for all EQ-5D-3L dimensions: playing a useful part, could not overcome difficulties, and feeling unhappy and depressed. The item feeling reasonably happy had no non-zero main effects, though there were several non-zero interactions. RMSEs were similar to models 1-3.

Figure 2 compares AIC and BIC for OL and LASSO models. LASSO was superior to the simple and extended OL models for pain/discomfort on both measures and had marginally higher AIC/BIC for mobility, usual activities and anxiety/depression. LASSO had higher AIC/BIC for self-care, but those values were similar to the OL models' performance for mobility, usual activities and anxiety/depression.

Figure 3 shows the RMSEs for all models for the estimation and test samples, as well as splitting the samples by gender and above/below median age. For all models, the pattern of results is remarkably similar. The RMSEs are similar for the estimation and test samples, with somewhat larger differences for males. RMSEs are smallest for self-care and, highest for pain/discomfort. Few differences were seen between female and male subgroups, apart from the aforementioned larger gaps between estimation and test RMSEs for males. RMSEs were lower for younger participants, and higher for older participants. No model showed a clear advantage in terms of RMSE.

To summarise how the different approaches performed against the model selection criteria:-

- LASSO provided better model fit than simple or extended OL on the pain/discomfort dimension.
- All models demonstrated similar prediction accuracy.
- Non-parametric was considered the simplest approach LASSO the most complex.

The pragmatic decision was taken to select the non-parametric model as the basis for the mapping. This was primarily due to its greater simplicity, and that no other model outperformed it in terms of prediction accuracy. While LASSO had an advantage over the OL approaches model fit, this did not justify adopting a much more complex model, especially as better model fit did not translate into better predictive accuracy.

Figure 4 shows bubble plots of observed EQ-5D-3L responses on each dimension against the most probable level predicted by the final non-parametric mapping. It also shows a scatter plot of observed EQ-5D-3L level sum. Figure 5 shows scatter plots of predicted against observed EQ-5D-3L values calculated using the MVH valueset. While there is some individual level variation, mean values are generally close to the 45° line for GHQ-12 score up to the low 20s. Similar patterns were seen for both female/male and younger/older subgroups.

Figure 6 compares average predicted and observed EQ-5D-3L values for test sample participants above and below the clinically significant GHQ-12 caseness threshold of 4. Differences between observed and predicted

values were -0.008 for below threshold participants and -0.0004 for above threshold participants. Larger discrepancies were seen when looking at female and male subgroups, the largest difference between observed and predicted values being 0.046 for above threshold males. Some underestimation of younger people's values was seen, with differences between observed and predicted values of -0.054 for those below threshold and -1.04 for above threshold. Conversely, overestimation was seen for older people, with differences between observed and predicted values of 0.040 and 0.109 for above and below the threshold respectively.

4 Discussion

The final mapping is based on a non-parametric approach, similar to that used in e.g. van Hout et al.'s EQ-5D-3L/5L crosswalk [30]. An R script and an Excel spreadsheet have been provided as supplementary material to help facilitate using the GHQ-12-EQ-5D-3L mapping in practice.

Examining the mapping's properties, there is a tendency to underestimate EQ-5D-3L state severity. This is related to the higher correlations between GHQ-12 items and the anxiety/depression and self-care dimensions, and low transition probabilities to mobility and self-care level 3. This may lead to a lower range of predicted utilities predicted than are present in an EQ-5D-3L valueset. However, on a population level, it may not be as much of a problem, since few people indicate being in mobility or self-care level 3 (e.g. only 0.14% and 0.36% respectively in the present study).

It has been demonstrated that it is possible to estimate EQ-5D-3L values using predicted responses from the mapping. While there may be discrepancies between individual observed and predicted values, average values can be much closer together. Mean observed and predicted values tended to be closer for lower GHQ-12 scores, with less accurate estimates for higher scores. However, relatively few people report high GHQ-12 scores, for example in the analysis sample only 6.1% of respondents reported a score greater than 20 out of 36. The relationship between observed and predicted values was similar amongst female/male and younger/older subgroups.

Mean predicted EQ-5D-3L values for groups above and below the GHQ-12 caseness threshold of 4 were very close to observed values. This is encouraging, as comparing QALYs for groups with and without a possible mental health problem where only GHQ-12 data is available, is an area where the mapping might be useful. If using the mapping in such way, the tendency to estimate younger people's health-related quality of life as worse, and older people's as better than it was, should be noted.

The GHQ-12-EQ-5D-3L mapping will be most useful in calculating QALYs when no general or specific measure with an associated valueset is available. An example is the Understanding Society COVID-19 survey [31]. While all nine waves included GHQ-12, participants did not complete measures such as EQ-5D or SF-

6D, so the GHQ-12-EQ-5D-3L mapping will allow QALYs to be estimated from this data. GHQ-12 is also used as a diagnostic tool, so the mapping may be useful with routine clinical data, where measures such as EQ-5D-3L may not be collected.

In line with several other EQ-5D-3L mapping studies (e.g. [30, 32, 33, 34]) covariates such as age, gender, other health measures or the presence of long-term conditions were not added. While covariates could have improved model fit and predicted accuracy, the justification for excluding them is that it allows the mapping to be more widely applicable. While some demographic variables are commonly collected, covariate data is not always available, or may not be in the format required. (For example, age may be recorded as single year, five year categories, 10 year categories, etc.)

Comparing this study's results with previous GHQ-12-EQ-5D-3L mapping studies, Lindkvist and Feldman [16] report a Pearson correlation of 0.65 between observed and predicted values, compared to 0.49 in the test sample here. The higher correlation is probably due to their inclusion of age, gender, and self-reported health which seems to provide a lot of explanatory power.[3] Covariates were not included here for reasons discussed above. Similarly, Serrano-Aguilar at al. [17] reported good predictive accuracy from their model, but also included age, gender, education level and number of co-morbidities. Both the aforementioned studies map to EQ-5D-3L values, whereas this study maps to EQ-5D-3L health states. The advantage of the latter approach is that values can be calculated using any valueset. Thus, for example, the mapping could be used in a country outside the UK with a valueset local to that country.

The GHQ-12-EQ-5D-3L mapping algorithm was calculated for England, since the data came from HSE. It could still be used in other areas of the UK in the absence of better alternative, in line with guidance to use valuesets from a similar country if no values are available for a given country [35]. It could also be used in non-UK countries, and, as noted above, a local valueset could be used.

This study has several advantages. Several approaches were evaluated. A number of different factors were considered in the decision as to which to use for the final mapping, taking account of both theoretical and practical concerns. The predictive accuracy of models was also tested on data not included in the estimation.

The study also has several disadvantages. Although it used a large sample, data was not collected prospectively. The data was collected in 2012, and people's attitudes towards health, particularly mental health, may have changed since then. For example, Webb et al. [36] present evidence that the COVID-19 pandemic may have changed people's EQ-5D valuations.

Around 15% of participants were discarded due to missing data, presenting a risk of bias. However, the characteristics of the full and analysis samples were similar. Although models were tested using observations

[3] For example, see page 6 of Lindkvist and Feldman [16]: "Inclusion of self-reported health together with GHQ-12 was essential for acceptable values of R^2."

not included in estimation, this is not a true test of external validity, since both estimation and test data came from the same sample.

This study presented an illustration of how values on the full health=1, dead=0 scale could be estimated using the mapping algorithm. However, this is just one example, and further investigation is required to examine the validity of values and/or QALYs calculated using the algorithm. This includes using different datasets and/or other valuesets.

Another limitation is the is poor conceptual overlap between GHQ-12 and EQ-5D-3L, with the former measuring mental health, and the latter measuring general health-related quality of life with only one of five dimensions devoted to mental health. The resulting EQ-5D-3L states are only estimated, and any values calculated from them, will not be as accurate as values calculated from EQ-5D-3L, SF-6D, or other general health measure. It would always be preferable to use such measures where they are available. Yet mental health will likely have an impact on other EQ-5D-3L dimensions, for example poor mental health could impact individuals' ability to self-care or perform their usual activities. In the other direction, mobility problems or pain could have an adverse effect on mental health.

5 Conclusion

This paper presents a mapping algorithm to estimate EQ-5D-3L states from GHQ-12 responses. Analysis tools have also been provided as supplementary material to help researchers wishing to use this mapping in practical applications.

Future research could usefully assess the validity and predictive accuracy of the mapping in other data sets, and in particuar clinical populations.

References

[1] Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. Psychological Medicine. 1979;9(1):139-45.

[2] Banks MH, Clegg CW, Jackson PR, Kemp NJ, Stafford EM, Wall TD. The use of the General Health Questionnaire as an indicator of mental health in occupational studies. Journal of Occupational Psychology. 1980;53(3):187-94.

[3] Jackson C. The General Health Questionnaire. Occupational Medicine. 2007;57(1):79-9.

[4] Kalliath TJ, O'Driscoll MP, Brough P. A confirmatory factor analysis of the General Health Questionnaire-12. Stress and Health: Journal of the International Society for the Investigation of Stress. 2004;20(1):11-20.

[5] Romppel M, Braehler E, Roth M, Glaesmer H. What is the General Health Questionnaire-12 assessing?: Dimensionality and psychometric properties of the General Health Questionnaire-12 in a large scale German population sample. Comprehensive Psychiatry. 2013;54(4):406-13.

[6] Aalto AM, Elovainio M, Kivimäki M, Uutela A, Pirkola S. The Beck Depression Inventory and General Health Questionnaire as measures of depression in the general population: a validation study using the Composite International Diagnostic Interview as the gold standard. Psychiatry Research. 2012;197(1-2):163-71.

[7] Lundin A, Hallgren M, Theobald H, Hellgren C, Torgén M. Validity of the 12-item version of the General Health Questionnaire in detecting depression in the general population. Public Health. 2016;136:66-74.

[8] Goldberg D, Williams P. A user's guide to the General Health Questionnaire. Windsor: NFER-Nelson; 1988.

[9] National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal; 2013.

[10] Dakin H, Abel L, Burns R, Yang Y. Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: An online database and application of the MAPS statement. Health and Quality of Life Outcomes. 2018;16(1):1-9.

[11] Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: An online database. Health and Quality of Life Outcomes. 2013;11(1):1-6.

[12] Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. The European journal of health economics. 2010;11(2):215-25.

[13] Devlin NJ, Brooks R. EQ-5D and the EuroQol group: Past, present and future. Applied Health Economics and Health Policy. 2017;15(2):127-37.

[14] EuroQol Group. EQ-5D-3L valuation;. Accessed 25/03/22. Available from: https://euroqol.org/eq-5d-instruments/eq-5d-3l-about/valuation/.

[15] National Institute for Health and Clinical Excellence (NICE). Position statement on use of the EQ-5D-5L value set for England (updated October 2019);. Accessed 25/03/22. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l.

[16] Lindkvist M, Feldman I. Assessing outcomes for cost-utility analysis in mental health interventions: mapping mental health specific outcome measure GHQ-12 onto EQ-5D-3L. Health and Quality of Life Outcomes. 2016;14(1):1-9.

[17] Serrano-Aguilar P, Ramallo-Fariña Y, Trujillo-Martín MDM, Muñoz-Navarro SR, Perestelo-Perez L, De Las Cuevas-Castresana C. The relationship among mental health status (GHQ-12), health related quality of life (EQ-5D) and health-state utilities in a general population. Epidemiology and Psychiatric Sciences. 2009;18(3):229-39.

[18] NatCen Social Research, University College London, Department of Epidemiology and Public Health. Health Survey for England, 2012. UK Data Archive Colchester, UK.; 2014.

[19] Bridges S, Doyle M, Fuller E, Knott C, Mindell J, Moody A, et al. Health Survey for England 2012 Methods and Documentation. Craig R, Mindell J, editors. NatCen Social Research, Department of Epidemiology and Public Health, UCL (University College London); 2012. Available from: https://files.digital.nhs.uk/publicationimport/pub13xxx/pub13218/hse2012-methods-and-docs.pdf.

[20] Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A, Firth D, et al. Package 'MASS'. CRAN R. 2013;538:113-20.

[21] Lever J, Krzywinski M, Altman N. Points of significance: Model selection and overfitting. Nature methods. 2016;13(9):703-5.

[22] Tibshirani R. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society: Series B (Methodological). 1996;58(1):267-88.

[23] Wurm MJ, Rathouz PJ, Hanlon BM. Regularized ordinal regression and the ordinalNet R package. arXiv preprint arXiv:170605003. 2017.

[24] Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974;19(6):716-23.

[25] Schwarz G. Estimating the dimension of a model. The Annals of Statistics. 1978:461-4.

[26] Thomson RM, Niedzwiedz CL, Katikireddi SV. Trends in gender and socioeconomic inequalities in mental health following the Great Recession and subsequent austerity policies: A repeat cross-sectional analysis of the Health Surveys for England. BMJ Open. 2018;8(8):e022924.

[27] Dolan P. Modeling valuations for EuroQol health states. Medical Care. 1997;35(11):1095-108.

[28] Janssen M, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: A multi-country study. Quality of Life Research. 2013;22(7):1717-27.

[29] Devlin N, Parkin D, Janssen B. Methods for analysing and reporting EQ-5D data. Springer Nature; 2020.

[30] Van Hout B, Janssen M, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health. 2012;15(5):708-15.

[31] University of Essex, Institute for Social and Economic Research. Understanding Society: COVID-19 Study, 2020-2021. [data collection]. UK Data Service; 2021.

[32] Jang RW, Isogai PK, Mittmann N, Bradbury PA, Shepherd FA, Feld R, et al. Derivation of utility values from European Organization for Research and Treatment of Cancer Quality of Life-Core 30 questionnaire values in lung cancer. Journal of Thoracic Oncology. 2010;5(12):1953-7.

[33] Lawrence WF, Fleishman JA. Predicting EuroQoL EQ-5D preference scores from the SF-12 Health Survey in a nationally representative sample. Medical Decision Making. 2004;24(2):160-9.

[34] Van Exel N, Scholte op Reimer W, Koopmanschap M. Assessment of post-stroke quality of life in cost-effectiveness studies: The usefulness of the Barthel Index and the EuroQoL-5D. Quality of Life Research. 2004;13(2):427-33.

[35] EuroQol Group. Choosing a value set;. Accessed 25/03/22. Available from: https://euroqol.org/eq-5d-instruments/eq-5d-3l-about/valuation/choosing-a-value-set/.

[36] Webb EJ, Kind P, Meads D, Martin A. Does a health crisis change how we value health? Health Economics. 2021;30(10):2547-60.

Declarations

The author received no funding in association with this work and has no competing interests.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication (from patients/participants)

Not applicable.

Availability of data and material

Health Survey for England data is available from the UK Data Service.

Code availability

Code is available on request to the corresponding author.

Table 1: Respondent demographics

		All		Missing		Analysis sample		Estimation sample		Test sample	
		N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Age	mean (sd)	50.8	(18.4)	0	0	50.3	(18.0)	50.6	(18.0)	49.8	(18.1)
Female		4530	(55.8)	0	0	3870	(55.9)	2720	(56.1)	1150	(55.5)
Ethnicity	white	7240	(89.6)	25	0.308	6260	(90.4)	4390	(90.6)	1870	(89.9)
	mixed	95	(1.17)			76	(1.10)	45	(0.929)	31	(1.49)
	asian	501	(6.19)			399	(5.77)	286	(5.91)	113	(5.44)
	black	200	(2.47)			156	(2.25)	102	(2.11)	54	(2.60)
	other	49	(0.606)			32	(0.463)	20	(0.413)	12	(0.578)
Household size	mean (sd)	2.62	(1.35)	0	0	2.62	(1.34)	2.63	(1.36)	2.62	(1.29)
Marital status	single	1440	(17.7)	3	0.037	1230	(17.8)	854	(17.6)	377	(18.2)
	married/civil partnership	4390	(54.1)			3780	(54.6)	2660	(55.0)	1120	(53.8)
	cohabiting	861	(10.6)			739	(10.7)	511	(10.5)	228	(11.0)
	separated/divorced	811	(10.0)			699	(10.1)	495	(10.2)	204	(9.83)
	widowed	618	(7.62)			474	(6.85)	323	(6.67)	151	(7.27)
Highest qualification	degree/equivalent	2050	(25.3)	26	0.32	1830	(26.5)	1250	(25.8)	586	(28.2)
	higher education below degree	892	(11.0)			791	(11.4)	552	(11.4)	239	(11.5)
	A-level/equivalent	1240	(15.3)			1100	(15.9)	769	(15.9)	333	(16.0)
	GCSE/O-level/equivalent	1560	(19.2)			1380	(19.9)	975	(20.1)	400	(19.3)
	other	456	(5.64)			382	(5.52)	257	(5.31)	125	(6.02)
	none	1900	(23.5)			1430	(20.7)	1040	(21.5)	394	(19.0)
Occupation	employed	3610	(44.6)	24	0.296	3190	(46.2)	2240	(46.2)	954	(46.0)
	self-employed	699	(8.64)			607	(8.78)	407	(8.41)	200	(9.64)
	unemployed	384	(4.75)			334	(4.83)	221	(4.57)	113	(5.45)
	retired	2240	(27.7)			1840	(26.5)	1300	(26.8)	538	(25.9)
	other	1160	(14.3)			946	(13.7)	676	(14.0)	270	(13.0)
Rurality	urban	6320	(77.9)	0	0	5380	(77.8)	3750	(77.4)	1630	(78.5)
	town and fringe	868	(10.7)			753	(10.9)	548	(11.3)	205	(9.87)
	village, hamlet and isolated dwellings	923	(11.4)			787	(11.4)	546	(11.3)	241	(11.6)
Self-assessed health	very good	2580	(31.8)	2	0.0246	2280	(33.0)	1600	(33.0)	686	(33.0)
	good	3370	(41.6)			2920	(42.1)	2000	(41.3)	917	(44.2)
	fair	1540	(18.9)			1250	(18.0)	898	(18.5)	348	(16.8)
	bad	454	(5.60)			356	(5.14)	261	(5.38)	95	(4.58)
	very.bad	169	(2.08)			118	(1.70)	88	(1.82)	30	(1.45)
N		8114				6924		4847		2077	

Note. sd=standard deviation; GCSE=General Certificate of Higher Education

Table 2: General health questionnaire responses

Item	Score	All		Missing		Analysis sample		Estimation sample		Test sample	
		N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Able to concentrate	0	240	(3.31)	873	(10.8)	223	(3.22)	161	(3.32)	62	(2.99)
	1	6138	(84.8)			5889	(85.1)	4100	(84.6)	1789	(86.1)
	2	735	(10.2)			695	(10.0)	500	(10.3)	195	(9.39)
	3	128	(1.77)			117	(1.69)	86	(1.77)	31	(1.49)
Lost much sleep	0	2264	(31.3)	873	(10.8)	2167	(31.3)	1516	(31.3)	651	(31.3)
	1	3699	(51.1)			3544	(51.2)	2478	(51.1)	1066	(51.3)
	2	1058	(14.6)			1007	(14.5)	704	(14.5)	303	(14.6)
	3	220	(3.04)			206	(2.98)	149	(3.07)	57	(2.74)
Playing a useful part	0	595	(8.27)	919	(11.3)	569	(8.22)	387	(7.98)	182	(8.76)
	1	5820	(80.9)			5616	(81.1)	3926	(81.0)	1690	(81.4)
	2	613	(8.52)			584	(8.43)	413	(8.52)	171	(8.23)
	3	167	(2.32)			155	(2.24)	121	(2.50)	34	(1.64)
Capable of making decisions	0	447	(6.17)	870	(10.7)	419	(6.05)	293	(6.04)	126	(6.07)
	1	6260	(86.4)			6002	(86.7)	4200	(86.7)	1802	(86.8)
	2	444	(6.13)			418	(6.04)	289	(5.96)	129	(6.21)
	3	93	(1.28)			85	(1.23)	65	(1.34)	20	(0.963)
Under stress	0	1721	(23.8)	879	(10.8)	1636	(23.6)	1159	(23.9)	477	(23.0)
	1	4148	(57.3)			3982	(57.5)	2762	(57.0)	1220	(58.7)
	2	1151	(15.9)			1102	(15.9)	772	(15.9)	330	(15.9)
	3	215	(2.97)			204	(2.95)	154	(3.18)	50	(2.41)
Could not overcome difficulties	0	2540	(35.1)	879	(10.8)	2432	(35.1)	1699	(35.1)	733	(35.3)
	1	3929	(54.3)			3778	(54.6)	2628	(54.2)	1150	(55.4)
	2	597	(8.25)			565	(8.16)	410	(8.46)	155	(7.46)
	3	169	(2.34)			149	(2.15)	110	(2.27)	39	(1.88)
Enjoy your day-to-day activities	0	358	(4.92)	843	(10.4)	330	(4.77)	232	(4.79)	98	(4.72)
	1	5772	(79.4)			5540	(80.0)	3885	(80.2)	1655	(79.7)
	2	945	(13.0)			877	(12.7)	601	(12.4)	276	(13.3)
	3	196	(2.70)			177	(2.56)	129	(2.66)	48	(2.31)
Face up to problems	0	301	(4.15)	853	(10.5)	277	(4.00)	203	(4.19)	74	(3.56)
	1	6302	(86.8)			6041	(87.2)	4215	(87.0)	1826	(87.9)
	2	548	(7.55)			509	(7.35)	355	(7.32)	154	(7.41)
	3	110	(1.51)			97	(1.40)	74	(1.53)	23	(1.11)
Feeling unhappy and depressed	0	2889	(39.8)	856	(10.5)	2759	(39.8)	1931	(39.8)	828	(39.9)
	1	3261	(44.9)			3128	(45.2)	2181	(45.0)	947	(45.6)
	2	903	(12.4)			847	(12.2)	597	(12.3)	250	(12.0)
	3	205	(2.82)			190	(2.74)	138	(2.85)	52	(2.50)
Losing confidence	0	3245	(44.7)	854	(10.5)	3090	(44.6)	2181	(45.0)	909	(43.8)
	1	3057	(42.1)			2933	(42.4)	2028	(41.8)	905	(43.6)
	2	760	(10.5)			720	(10.4)	509	(10.5)	211	(10.2)
	3	198	(2.73)			181	(2.61)	129	(2.66)	52	(2.50)
Thinking of self as worthless	0	4761	(65.6)	855	(10.5)	4542	(65.6)	3148	(64.9)	1394	(67.1)
	1	1955	(26.9)			1877	(27.1)	1334	(27.5)	543	(26.1)
	2	407	(5.61)			386	(5.57)	276	(5.69)	110	(5.30)
	3	136	(1.87)			119	(1.72)	89	(1.84)	30	(1.44)
Feeling reasonably happy	0	787	(10.8)	856	(10.5)	745	(10.8)	504	(10.4)	241	(11.6)
	1	5695	(78.5)			5449	(78.7)	3821	(78.8)	1628	(78.4)
	2	643	(8.86)			603	(8.71)	427	(8.81)	176	(8.47)
	3	133	(1.83)			127	(1.83)	95	(1.96)	32	(1.54)
Caseness score	mean (sd)	1.47	(2.75)	1005	(12.4)	1.46	(2.74)	1.48	(2.77)	1.41	(2.67)
Total score	mean (sd)	11	(5.12)	1005	(12.4)	11	(5.08)	11	(5.17)	10.9	(4.87)
N		8114				6924		4847		2077	

Note. sd=standard deviation

Table 3: EQ-5D-3L responses

		All		Missing		Analysis sample		Estimation sample		Test sample	
		N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Mobility	level 1	5934	(81.6)	844	(10.4)	5717	(82.6)	3980	(82.1)	1737	(83.6)
	level 2	1324	(18.2)			1197	(17.3)	859	(17.7)	338	(16.3)
	level 3	12	(0.165)			10	(0.144)	8	(0.165)	2	(0.0963)
Self-care	level 1	6826	(94.3)	876	(10.8)	6551	(94.6)	4579	(94.5)	1972	(94.9)
	level 2	386	(5.33)			352	(5.08)	251	(5.18)	101	(4.86)
	level 3	26	(0.359)			21	(0.303)	17	(0.351)	4	(0.193)
Usual activities	level 1	6059	(83.6)	865	(10.7)	5824	(84.1)	4050	(83.6)	1774	(85.4)
	level 2	1077	(14.9)			999	(14.4)	721	(14.9)	278	(13.4)
	level 3	113	(1.56)			101	(1.46)	76	(1.57)	25	(1.20)
Pain/discomfort	level 1	4758	(65.5)	851	(10.5)	4597	(66.4)	3198	(66.0)	1399	(67.4)
	level 2	2203	(30.3)			2060	(29.8)	1457	(30.1)	603	(29.0)
	level 3	302	(4.16)			267	(3.86)	192	(3.96)	75	(3.61)
Anxiety/depression	level 1	5771	(79.7)	875	(10.8)	5544	(80.1)	3871	(79.9)	1673	(80.5)
	level 2	1301	(18.0)			1226	(17.7)	864	(17.8)	362	(17.4)
	level 3	167	(2.31)			154	(2.22)	112	(2.31)	42	(2.02)
VAS	mean (sd)	77.4	(18.5)	1742	(21.5)	77.9	(18.2)	77.5	(18.6)	78.6	(17.1)
In state 11111		3987	(55.8)	970	(12.0)	3884	(56.1)	2701	(55.7)	1183	(57.0)
Level sum	mean (sd)	1.02	(1.56)	970	(12.0)	1	(1.55)	1.02	(1.57)	0.952	(1.50)
N		8114				6924		4847		2077	

Note. VAS=visual analogue scale; sd=standard deviation

Table 4: Non-parametric mapping transition probability matrices

GHQ score	Mobility			Self-care			Usual activities			Pain/discomfort			Anxiety/depression		
	level 1	level 2	level 3	level 1	level 2	level 3	level 1	level 2	level 3	level 1	level 2	level 3	level 1	level 2	level 3
0	1	0	0	1	0	0	0.969	0.0313	0	0.937	0.0625	0	1	0	0
1	1	0	0	1	0	0	1	0	0	0.913	0.0867	0	1	0	0
2	1	0	0	1	0	0	1	0	0	0.806	0.194	0	1	0	0
3	0.968	0.0317	0	1	0	0	1	0	0	0.918	0.0816	0	1	0	0
4	0.952	0.048	0	1	0	0	0.965	0.0352	0	0.93	0.0702	0	0.986	0.0142	0
5	0.936	0.0636	0	1	0	0	0.954	0.046	0	0.866	0.112	0.0219	0.97	0.0299	0
6	0.921	0.0777	0.00163	0.989	0.0099	0.000994	0.944	0.0531	0.00259	0.81	0.188	0.00196	0.992	0.00833	0
7	0.899	0.101	0	0.982	0.0166	0.00188	0.921	0.0755	0.00371	0.772	0.215	0.0131	0.971	0.0285	0
8	0.912	0.0863	0.00144	0.985	0.0151	0	0.925	0.0733	0.00204	0.756	0.238	0.0056	0.949	0.0514	0
9	0.877	0.123	0	0.977	0.0217	0.0017	0.924	0.0687	0.00762	0.729	0.253	0.0178	0.949	0.0509	0
10	0.891	0.109	0	0.983	0.0167	0	0.925	0.0724	0.00216	0.716	0.271	0.0127	0.912	0.0878	0
11	0.863	0.135	0.00234	0.978	0.0204	0.00193	0.9	0.0886	0.0119	0.685	0.298	0.0173	0.883	0.113	0.00353
12	0.839	0.161	0	0.955	0.0436	0.00158	0.865	0.132	0.00267	0.694	0.279	0.0267	0.812	0.18	0.00815
13	0.785	0.215	0	0.921	0.0789	0	0.788	0.198	0.0144	0.554	0.402	0.0435	0.664	0.314	0.0224
14	0.776	0.219	0.00559	0.912	0.0786	0.00936	0.796	0.19	0.0137	0.58	0.365	0.0549	0.586	0.395	0.0199
15	0.762	0.238	0	0.964	0.0365	0	0.728	0.256	0.0158	0.542	0.409	0.0486	0.525	0.458	0.0173
16	0.696	0.304	0	0.92	0.0804	0	0.675	0.311	0.014	0.474	0.423	0.103	0.421	0.539	0.04
17	0.666	0.334	0	0.869	0.131	0	0.648	0.308	0.0449	0.529	0.441	0.0301	0.448	0.481	0.0708
18	0.723	0.277	0	0.864	0.136	0	0.655	0.345	0	0.532	0.358	0.11	0.479	0.493	0.028
19	0.703	0.297	0	0.877	0.111	0.0112	0.653	0.31	0.0379	0.46	0.472	0.0681	0.324	0.59	0.0857
20	0.604	0.379	0.0169	0.763	0.22	0.0169	0.532	0.381	0.0874	0.404	0.451	0.145	0.31	0.595	0.0951
21	0.625	0.375	0	0.886	0.114	0	0.481	0.466	0.0527	0.235	0.59	0.175	0.293	0.614	0.0931
22	0.53	0.47	0	0.932	0.0683	0	0.518	0.482	0	0.307	0.489	0.203	0.174	0.744	0.0822
23	0.603	0.397	0	0.702	0.247	0.0517	0.603	0.23	0.167	0.382	0.491	0.126	0.283	0.683	0.0345
24	0.582	0.418	0	0.798	0.202	0	0.437	0.497	0.0659	0.434	0.463	0.103	0.0782	0.791	0.131
25	0.54	0.46	0	0.722	0.278	0	0.343	0.545	0.112	0.301	0.584	0.115	0.144	0.751	0.105
26	0.325	0.675	0	0.589	0.411	0	0.228	0.613	0.159	0.254	0.514	0.232	0.0593	0.672	0.269
27	0.616	0.333	0.0512	0.732	0.149	0.119	0.497	0.329	0.173	0.336	0.39	0.273	0.0519	0.609	0.339
28	0.402	0.598	0	0.838	0.0808	0.0813	0.361	0.473	0.167	0.322	0.504	0.174	0	0.592	0.408
29	0.464	0.536	0	0.759	0.241	0	0.231	0.596	0.173	0.314	0.422	0.264	0.0597	0.631	0.309
30	0.295	0.705	0	0.604	0.396	0	0.362	0.638	0	0.221	0.442	0.337	0	0.708	0.292
31	0.726	0.274	0	0.735	0.176	0.0886	0.21	0.702	0.0886	0.353	0.558	0.0886	0.112	0.52	0.368
32	0.395	0.495	0.11	0.449	0.44	0.11	0.337	0.552	0.11	0.344	0.25	0.406	0	0.259	0.741
33	0.389	0.611	0	0.389	0.511	0.1	0.292	0.521	0.187	0.172	0.292	0.536	0.0755	0.199	0.726
34	0.593	0.407	0	0.508	0.492	0	0.388	0.492	0.12	0.424	0.169	0.407	0	0.507	0.493
35	0.28	0.72	0	1	0	0	0	0.838	0.162	0.119	0.249	0.633	0	0.249	0.751
36	0.329	0.551	0.12	0.52	0.36	0.12	0.112	0.429	0.459	0.233	0.349	0.418	0	0.262	0.738
RMSE estimation	0.421			0.255			0.431			0.636			0.413		
RMSE test	0.411			0.239			0.403			0.630			0.399		

Note. N=4,847; GHQ=General health questionnaire; RMSE=Root mean squared error

Table 5: Simple ordered logit model results

	MO	s.e.	SC	s.e.	UA	s.e.	PD	s.e.	AD	s.e.
Able to concentrate	0.465*	(0.111)	0.245	(0.161)	0.543*	(0.112)	0.456*	(0.0912)	0.261*	(0.108)
Lost much sleep	-0.0228	(0.0713)	-0.0929	(0.119)	-0.0176	(0.0744)	0.217*	(0.0543)	0.315*	(0.0702)
Playing a useful part	0.491*	(0.0925)	0.890*	(0.137)	0.542*	(0.0950)	0.292*	(0.0759)	0.359*	(0.0922)
Capable of making decisions	-0.191	(0.126)	0.242	(0.181)	-0.0459	(0.130)	-0.074	(0.102)	-0.176	(0.124)
Under stress	-0.346*	(0.0841)	-0.259	(0.139)	-0.277*	(0.0885)	-0.0244	(0.0636)	0.398*	(0.0845)
Could not overcome difficulties	0.351*	(0.0873)	0.278	(0.144)	0.350*	(0.0913)	0.108	(0.0686)	0.318*	(0.0909)
Enjoy your day-to-day activities	1.04*	(0.0960)	1.00*	(0.142)	1.32*	(0.0978)	0.917*	(0.0804)	0.0896	(0.0969)
Face up to problems	0.0121	(0.119)	-0.282	(0.168)	-0.0481	(0.118)	-0.0861	(0.0988)	-0.0889	(0.113)
Feeling unhappy and depressed	0.0035	(0.0873)	-0.198	(0.150)	-0.0406	(0.0919)	0.0636	(0.0672)	1.12*	(0.0890)
Losing confidence	0.134	(0.0871)	0.405*	(0.142)	0.187*	(0.0909)	-0.0191	(0.0690)	0.200*	(0.0864)
Thinking of self as worthless	0.107	(0.0831)	0.216	(0.126)	0.156	(0.0844)	0.0737	(0.0674)	0.503*	(0.0779)
Feeling reasonably happy	-0.251*	(0.0998)	-0.374*	(0.153)	-0.127	(0.103)	-0.0335	(0.0792)	-0.141	(0.0991)
τ_1	3.65*	(0.149)	5.54*	(0.222)	4.61*	(0.157)	2.76*	(0.122)	4.54*	(0.154)
τ_2	8.89*	(0.419)	8.55*	(0.353)	7.83*	(0.241)	5.64*	(0.161)	8.30*	(0.247)
AIC	3826.1		1595.1		3671.7		6501.3		3658.2	
BIC	3916.8		1685.8		3762.4		6592.1		3748.9	
RMSE estimation	0.408		0.249		0.412		0.611		0.416	
RMSE test	0.402		0.236		0.402		0.614		0.416	

Note. N=4,847; *=significant at 5% level; MO=mobility; SC=self-care; UA=usual activities; PD=pain/discomfort; AD=anxiety/depression; s.e.=standard error; AIC=Akaike information criterion; BIC=Bayesian information criterion; RMSE=root mean squared error

Table 6: Extended ordered logit model results

Item	Score	MO	s.e.	SC	s.e.	UA	s.e.	PD	s.e.	AD	s.e.
Able to concentrate	1	0.355	(0.334)	-0.187	(0.588)	-0.0527	(0.338)	0.604*	(0.238)	-0.129	(0.301)
	2	0.891*	(0.358)	-0.0502	(0.622)	0.655	(0.361)	1.10*	(0.262)	0.353	(0.324)
	3	1.42*	(0.464)	0.625	(0.681)	1.06*	(0.456)	1.29*	(0.377)	0.162	(0.431)
Lost much sleep	1	0.151	(0.120)	0.181	(0.239)	0.0474	(0.132)	0.419*	(0.0885)	0.340*	(0.147)
	2	-0.176	(0.170)	0.00582	(0.302)	-0.216	(0.178)	0.320*	(0.126)	0.600*	(0.170)
	3	0.158	(0.268)	-0.527	(0.422)	0.207	(0.270)	0.869*	(0.225)	0.899*	(0.258)
Playing a useful part	1	0.266	(0.202)	0.758	(0.524)	0.318	(0.222)	0.181	(0.138)	0.381	(0.207)
	2	0.907*	(0.235)	1.76*	(0.542)	0.910*	(0.251)	0.632*	(0.179)	0.855*	(0.237)
	3	1.40*	(0.327)	2.55*	(0.593)	1.77*	(0.336)	0.899*	(0.277)	0.946*	(0.334)
Capable of making decisions	1	0.106	(0.244)	0.145	(0.513)	-0.0261	(0.264)	-0.0927	(0.164)	-0.124	(0.244)
	2	-0.195	(0.307)	0.434	(0.571)	-0.0916	(0.320)	-0.113	(0.230)	-0.345	(0.300)
	3	-0.188	(0.457)	1.32*	(0.652)	0.623	(0.453)	0.0192	(0.387)	-0.202	(0.441)
Under stress	1	-0.640*	(0.142)	-0.745*	(0.296)	-0.413*	(0.162)	-0.0515	(0.104)	0.242	(0.214)
	2	-0.786*	(0.190)	-0.924*	(0.358)	-0.523*	(0.206)	0.0329	(0.142)	0.610*	(0.234)
	3	-1.08*	(0.301)	-1.11*	(0.457)	-1.29*	(0.316)	-0.415	(0.253)	1.01*	(0.315)
Could not overcome difficulties	1	0.417*	(0.133)	0.670*	(0.286)	0.652*	(0.150)	0.135	(0.0937)	0.276	(0.157)
	2	0.840*	(0.204)	1.20*	(0.360)	0.766*	(0.216)	0.234	(0.164)	0.557*	(0.206)
	3	0.752*	(0.343)	0.406	(0.512)	0.672	(0.345)	0.0505	(0.309)	0.867*	(0.331)
Enjoy your day-to-day activities	1	0.456	(0.295)	0.0154	(0.530)	0.675	(0.358)	0.418*	(0.200)	0.0468	(0.263)
	2	1.71*	(0.312)	1.25*	(0.552)	2.32*	(0.371)	1.62*	(0.221)	0.22	(0.284)
	3	2.36*	(0.395)	2.28*	(0.603)	3.08*	(0.439)	1.96*	(0.321)	0.0352	(0.382)
Face up to problems	1	0.364	(0.286)	1.56*	(0.711)	0.375	(0.304)	0.149	(0.197)	-0.848*	(0.235)
	2	0.525	(0.317)	0.852	(0.725)	0.581	(0.330)	-0.00501	(0.237)	-0.585*	(0.268)
	3	0.413	(0.460)	1.03	(0.788)	0.0864	(0.459)	0.49	(0.403)	-0.00594	(0.424)
Feeling unhappy and depressed	1	0.169	(0.135)	0.0752	(0.279)	-0.153	(0.150)	0.0939	(0.0965)	1.83*	(0.209)
	2	0.0421	(0.199)	-0.477	(0.374)	-0.145	(0.209)	0.193	(0.151)	2.64*	(0.237)
	3	-0.419	(0.357)	-0.627	(0.515)	-0.416	(0.351)	-0.126	(0.314)	3.47*	(0.361)
Losing confidence	1	0.101	(0.134)	0.388	(0.275)	0.126	(0.149)	-0.14	(0.0982)	0.26	(0.147)
	2	0.207	(0.193)	0.603	(0.335)	0.213	(0.203)	-0.112	(0.155)	0.351	(0.189)
	3	0.208	(0.353)	1.21*	(0.482)	0.627	(0.352)	0.296	(0.318)	0.294	(0.343)
Thinking of self as worthless	1	0.268*	(0.117)	0.621*	(0.214)	0.485*	(0.125)	0.180*	(0.0906)	0.509*	(0.110)
	2	0.186	(0.205)	0.663*	(0.301)	0.214	(0.209)	0.132	(0.176)	0.841*	(0.185)
	3	0.31	(0.378)	0.0368	(0.491)	0.0957	(0.365)	-0.133	(0.338)	1.44*	(0.356)
Feeling reasonably happy	1	0.0456	(0.172)	-0.42	(0.311)	0.0325	(0.191)	0.183	(0.125)	-0.157	(0.194)
	2	-0.358	(0.239)	-0.273	(0.379)	-0.0088	(0.248)	0.00763	(0.187)	-0.119	(0.234)
	3	-0.42	(0.379)	-0.807	(0.520)	-0.173	(0.376)	0.0779	(0.328)	-0.239	(0.364)
τ_1		3.76*	(0.436)	6.20*	(1.00)	4.00*	(0.475)	2.85*	(0.293)	3.95*	(0.396)
τ_2		8.96*	(0.579)	9.29*	(1.04)	7.17*	(0.498)	5.74*	(0.306)	7.65*	(0.422)
AIC		3824.6		1581.8		3646.5		6495.2		3650.9	
BIC		4070.9		1828.1		3892.8		6741.5		3897.2	
RMSE estimation		0.409		0.241		0.408		0.601		0.412	
RMSE test		0.4		0.234		0.402		0.6		0.408	

Note. N=4,847; *=significant at 5% level; MO=mobility; SC=self-care; UA=usual activities; PD=pain/discomfort; AD=anxiety/depression; s.e.=standard error; AIC=Akaike information criterion; BIC=Bayesian information criterion; RMSE=root mean squared error

Table 7: Least absolute shrinkage and selection operator results

Item		Interaction	MO	SC	UA	PD	AD
Able to concentrate			0.383	-	0.454	0.462	0.141
Lost much sleep			-	-	-	0.327	0.248
Playing a useful part			0.378	0.896	0.563	0.0681	0.27
Capable of making decisions			0.179	0.164	-	0.076	-
Under stress			-	-	-	0.0508	0.363
Could not overcome difficulties			0.386	1.06	0.354	0.108	0.278
Enjoy your day-to-day activities			1.28	0.829	1.68	1.21	-
Face up to problems			0.134	-0.0656	0.0266	-	-
Feeling unhappy and depressed			0.219	0.158	0.106	0.372	1.28
Losing confidence			0.0781	-	0.151	-	0.345
Thinking of self as worthless			0.263	0.804	0.339	-	0.407
Able to concentrate	×	Lost much sleep	-	0.107	-	-	0.0431
	×	Capable of making decisions	-	0.0718	0.0382	-	-
	×	Under stress	-	-	-	-0.018	-
	×	Thinking of self as worthless	-	3.57×10^{-3}	-	-	-
Lost much sleep	×	Playing a useful part	-0.116	-	-	-	-
	×	Under stress	0.0613	-	-	-0.0111	-
	×	Could not overcome difficulties	-	-0.191	-9.26×10^{-3}	-	-
	×	Enjoy your day-to-day activities	-	-	-	-0.0263	0.0183
	×	Face up to problems	-	-	-	-0.105	-
	×	Feeling unhappy and depressed	-	-	-	-0.0898	-
	×	Losing confidence	-	-	-	-	-0.019
	×	Thinking of self as worthless	-	-	-	-0.0294	-
	×	Feeling reasonably happy	-	-	-	0.088	-
Playing a useful part	×	Capable of making decisions	0.0238	-	-	0.0163	-
	×	Could not overcome difficulties	0.0545	-	-	0.17	-
	×	Enjoy your day-to-day activities	0.0932	-	-	-	-
	×	Face up to problems	-	-	-	9.08×10^{-3}	-
	×	Thinking of self as worthless	-	-0.197	-	-	-
	×	Feeling reasonably happy	-	0.109	-	-	-
Capable of making decisions	×	Under stress	-0.13	-	-	-	-
	×	Could not overcome difficulties	-	-0.0439	-	-0.0959	-
	×	Enjoy your day-to-day activities	-	0.0835	-	-	-
	×	Face up to problems	-	-	8.67×10^{-3}	0.0962	-
	×	Feeling reasonably happy	-0.0478	-	-	-0.0995	-
Under stress	×	Could not overcome difficulties	-	-0.0516	-	-	-
	×	Enjoy your day-to-day activities	-0.197	-	-0.146	-0.104	-
	×	Feeling unhappy and depressed	-	0.162	-	-	-
	×	Losing confidence	0.0353	-	-	-	-
	×	Thinking of self as worthless	-	-	-	0.0809	0.0705
	×	Feeling reasonably happy	-	-0.359	-	-	-0.0648
Could not overcome difficulties	×	Enjoy your day-to-day activities	-	-0.0972	-	-	-
	×	Feeling unhappy and depressed	-0.0879	-0.0433	-	-0.107	-
	×	Losing confidence	-	-0.102	-	-	-
	×	Thinking of self as worthless	-	-	-0.0494	-	-
Enjoy your day-to-day activities	×	Feeling unhappy and depressed	-0.102	-0.27	-0.144	-0.0831	-
	×	Losing confidence	-	0.256	-	0.0267	0.0204
	×	Feeling reasonably happy	-	0.127	-	-	-
Face up to problems	×	Feeling unhappy and depressed	-	-4.91×10^{-3}	-	-	-
	×	Thinking of self as worthless	-	-	-	0.0248	-
Feeling unhappy and depressed	×	Losing confidence	-	-	-	-	-0.0919
	×	Thinking of self as worthless	-0.0561	-0.0912	-0.0623	-0.0161	-0.0767
	×	Feeling reasonably happy	-	-	-	-	-0.0164
Losing confidence	×	Thinking of self as worthless	-0.0411	-0.0415	-	-	-
Thinking of self as worthless	×	Feeling reasonably happy	-	-	-	-	0.0554
τ_1			4.43	6.34	5.22	3.12	4.58
τ_2			9.69	9.49	8.35	5.97	8.23
AIC			4095.7	4099.7	4100.2	4099	4099
BIC			4251.4	4313.7	4255.9	4326	4326
RMSE estimation			0.132	0.0491	0.113	0.24	0.0904
RMSE test			0.126	0.0472	0.104	0.223	0.0929

Note. N=4,847; MO=mobility; SC=self-care; UA=usual activities; PD=pain/discomfort; AD=anxiety/depression; AIC=Akaike information criterion; BIC=Bayesian information criterion; RMSE=root mean squared error

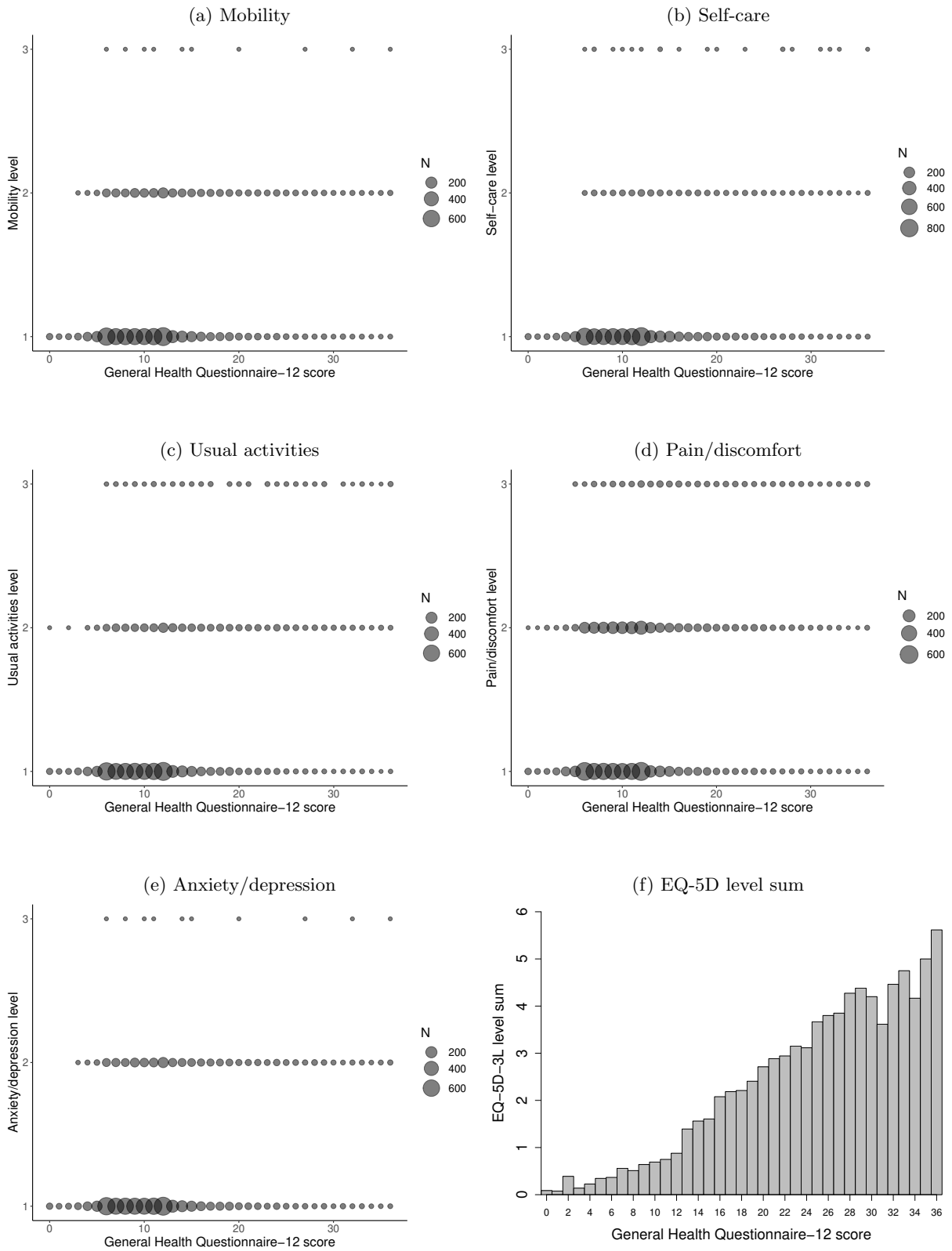


Figure 1: Bubble plots of General Health Questionnaire responses against EQ-5D-3L responses. N=6,924

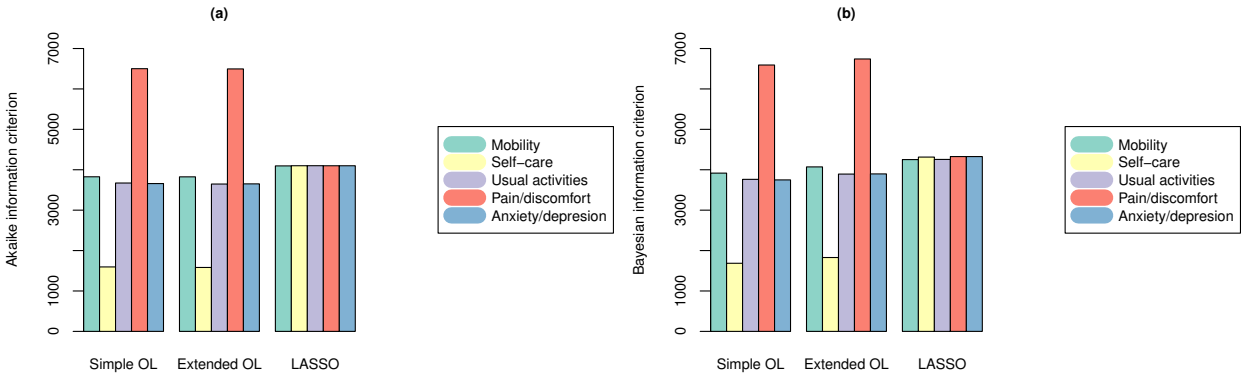


Figure 2: (a) Akaike information criteria and (b) Bayesian information criteria for OL and LASSO models. Estimation sample N=4,847, test sample N=2,077, OL=ordered logit, LASSO=least absolute shrinkage and selection operator

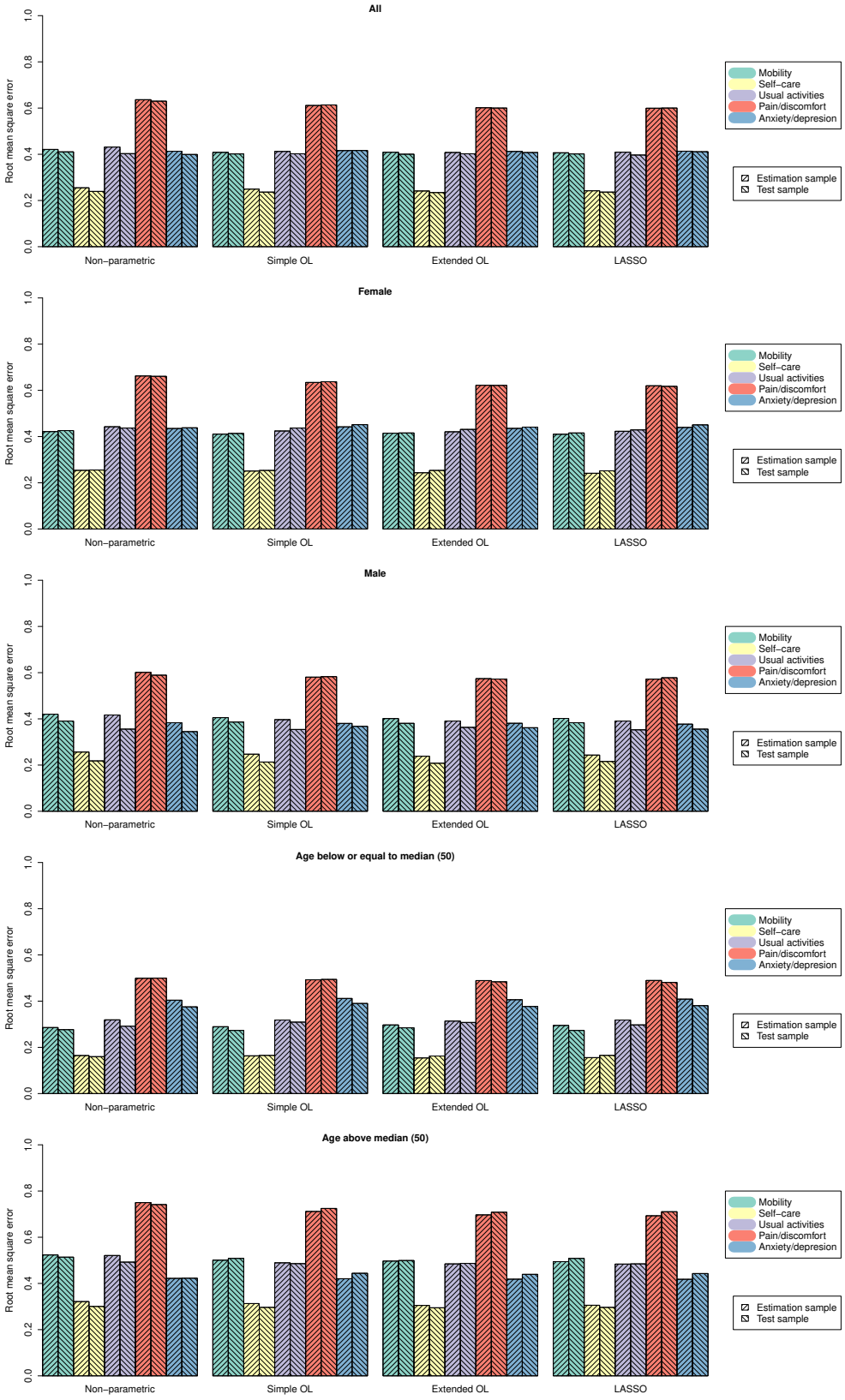


Figure 3: Root mean square errors for all models. Estimation sample N=4,847, test sample N=2,077, OL=ordered logit, LASSO=least absolute shrinkage and selection operator

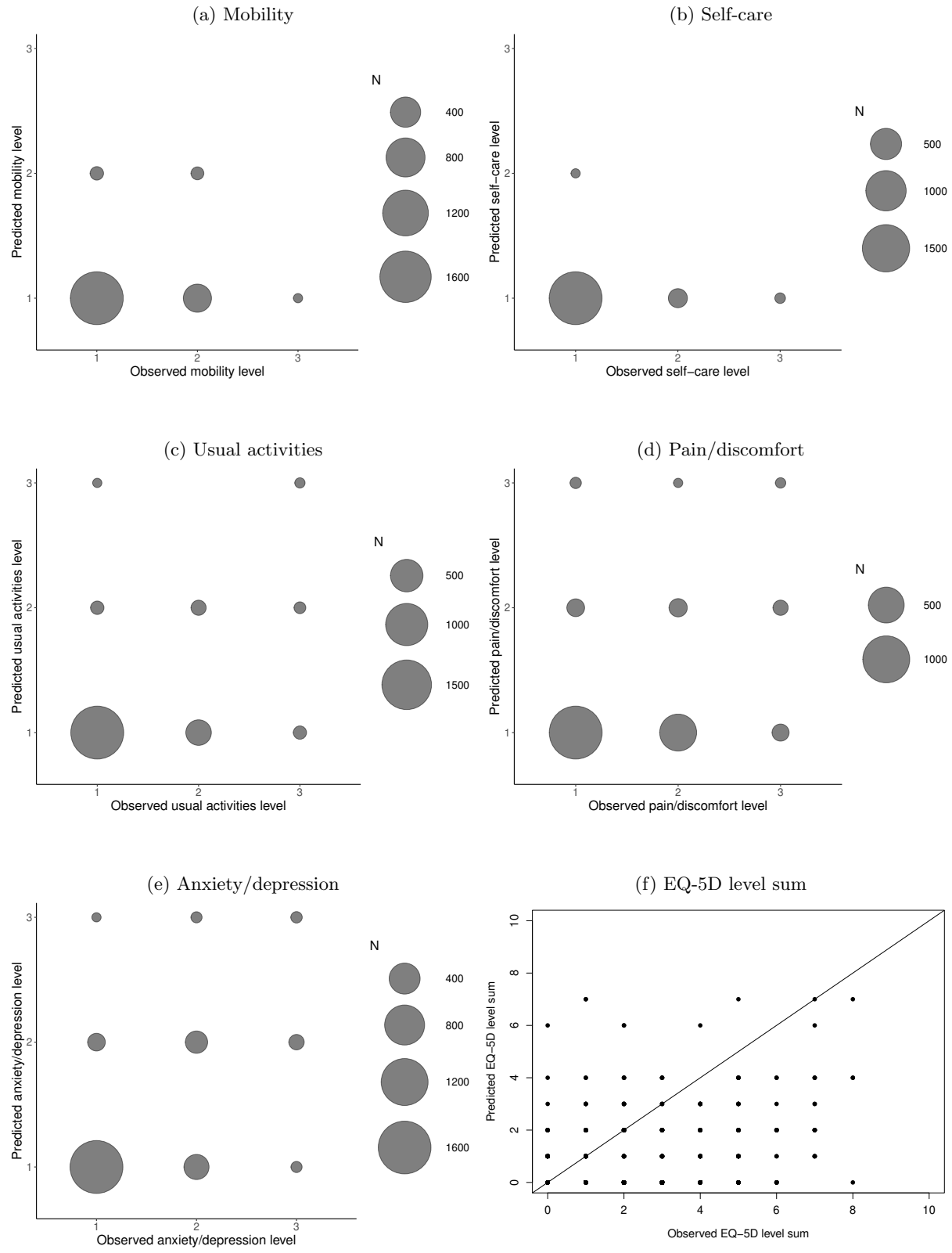
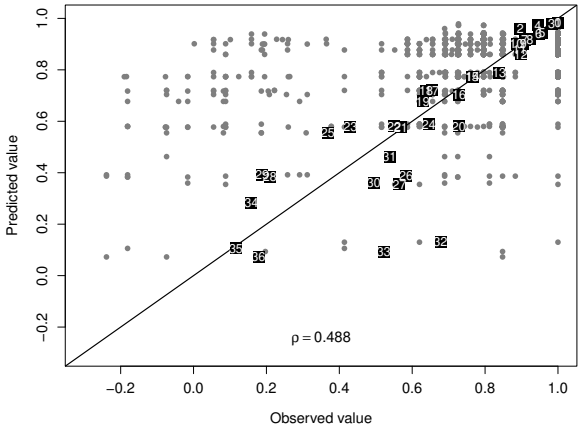
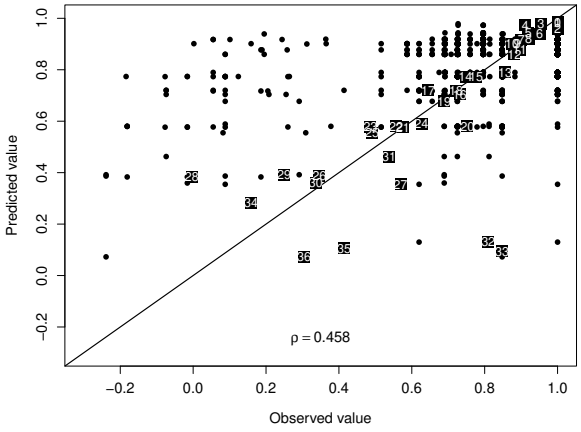


Figure 4: Observed against predicted EQ-5D-3L responses in test sample. N=2,077

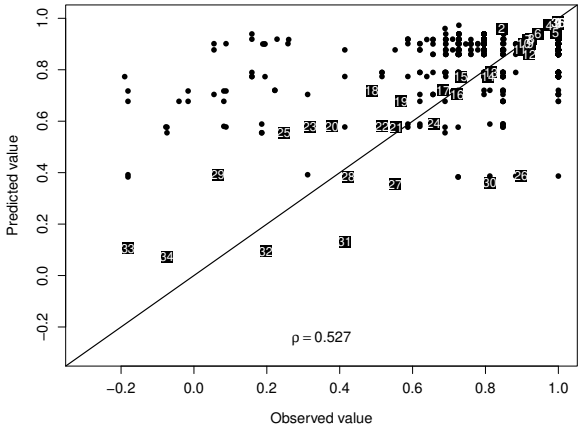
(a) All



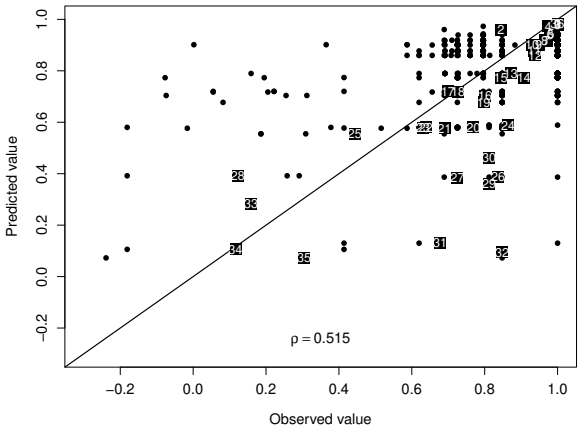
(b) Female



(c) Male



(d) Age below or equal to median (50)



(e) Age above median (50)

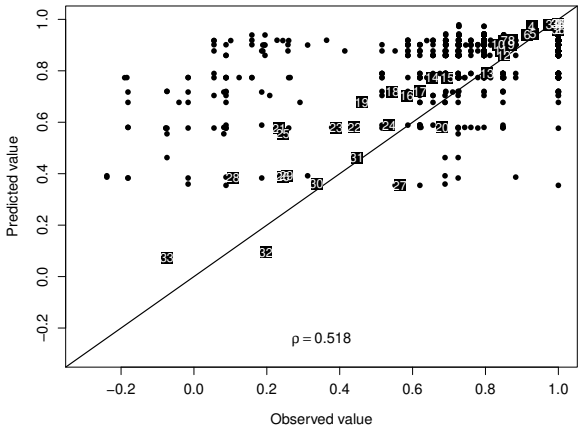


Figure 5: Observed against predicted EQ-5D-3L values in test sample. Numbers represent means for participants with that General Health Questionnaire-12 score. N=2,077

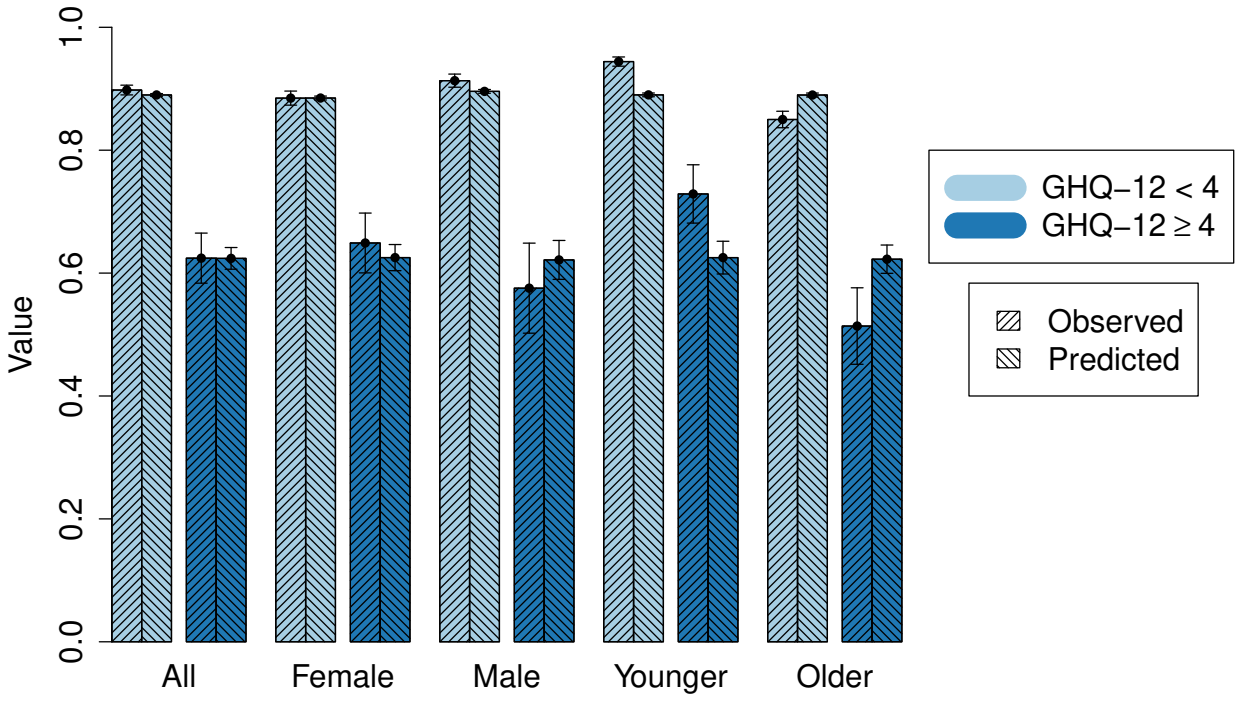


Figure 6: Mean observed and predicted EQ-5D-3L values for participants above and below GHQ-12 caseness threshold of 4. Estimation sample N=4,847, test sample N=2,077, GHQ-12=General Health Questionnaire-12

Appendix

Under each heading, please tick the ONE box that best describes your health TODAY

MOBILITY
I have no problems in walking about ☐
I have some problems in walking about ☐
I am confined to bed ☐

SELF-CARE
I have no problems with self-care ☐
I have some problems washing or dressing myself ☐
I am unable to wash or dress myself ☐

USUAL ACTIVITIES *(e.g. work, study, housework, family or leisure activities)*
I have no problems with performing my usual activities ☐
I have some problems with performing my usual activities ☐
I am unable to perform my usual activities ☐

PAIN / DISCOMFORT
I have no pain or discomfort ☐
I have some pain or discomfort ☐
I have extreme pain or discomfort ☐

ANXIETY / DEPRESSION
I am not anxious or depressed ☐
I am moderately anxious or depressed ☐
I am extremely anxious or depressed ☐

Table A.1: EQ-5D-3L classification system

Have you recently?

1. Been able to concentrate on what you're doing?	Better than usual	☐	Same as usual	☐	Less than usual	☐	Much less than usual	☐
2. Lost much sleep over worry?	Not at all	☐	No more than usual	☐	Rather more than usual	☐	Much more than usual	☐
3. Felt you were playing a useful part in things?	More so than usual	☐	Same as usual	☐	Less useful than usual	☐	Much less useful	☐
4. Felt capable of making decisions about things?	More so than usual	☐	Same as usual	☐	Less useful than usual	☐	Much less capable	☐
5. Felt constantly under strain?	Not at all	☐	No more than usual	☐	Rather more than usual	☐	Much more than usual	☐
6. Felt you couldn't overcome your difficulties?	Not at all	☐	No more than usual	☐	Rather more than usual	☐	Much more than usual	☐
7. Been able to enjoy your normal day-to-day activities?	More so than usual	☐	same as usual	☐	Less so than usual	☐	Much less than usual	☐
8. Been able to face up to your problems?	More so than usual	☐	same as usual	☐	Less so than usual	☐	Much less able	☐
9. Been feeling unhappy and depressed?	Not at all	☐	No more than usual	☐	Rather more than usual	☐	Much more than usual	☐
10. Been losing confidence in yourself?	Not at all	☐	No more than usual	☐	Rather more than usual	☐	Much more than usual	☐
11. Been thinking of yourself as a worthless person?	Not at all	☐	No more than usual	☐	Rather more than usual	☐	Much more than usual	☐
12. Been feeling reasonably happy, all things considered	More so than usual	☐	About same as usual	☐	Less so than usual	☐	Much less than usual	☐

Table A.2: General Health Questionnaire-12

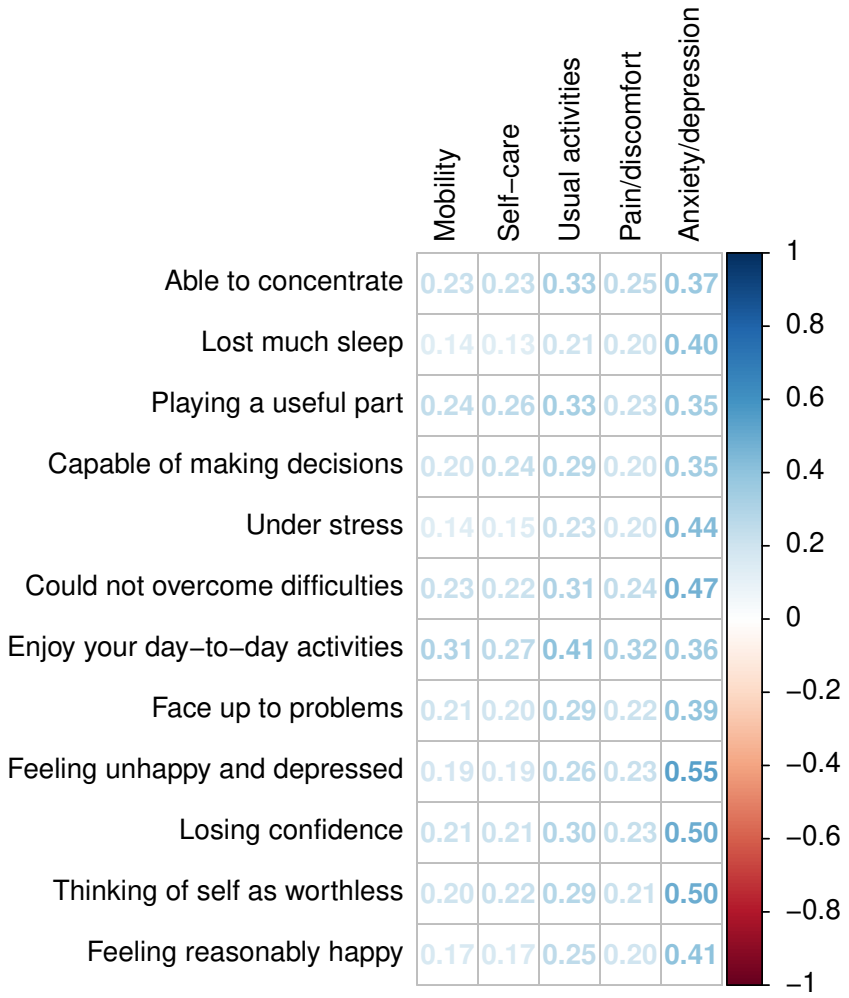
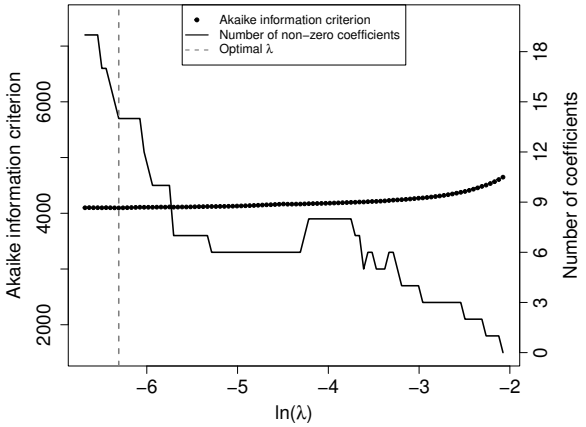
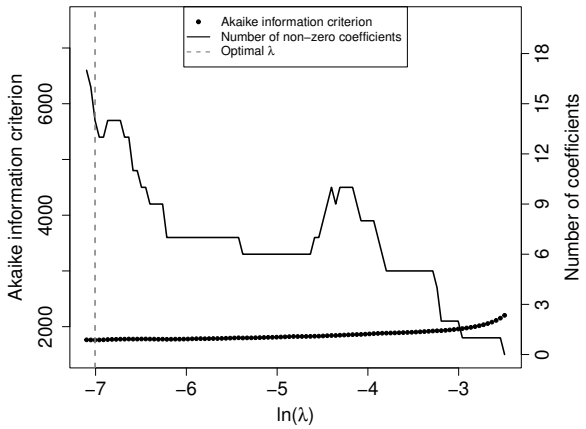


Figure A.1: Correlations between EQ-5D-3L dimensions and General Health Questionnaire-12 items in analysis sample. N=6,924

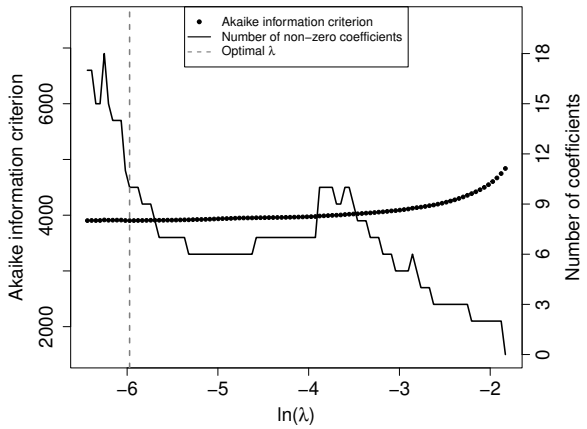
(a) All



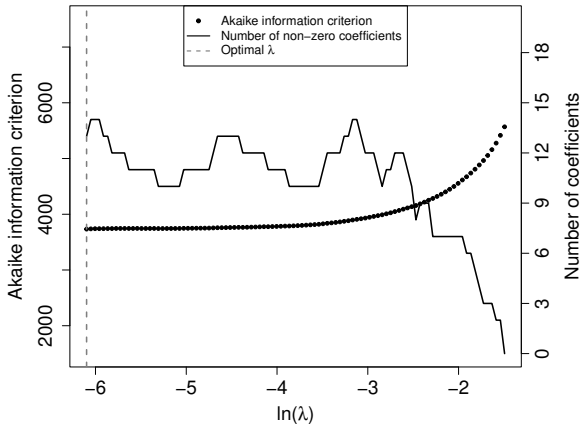
(b) Female



(c) Male



(d) Age below or equal to median (50)



(e) Age above median (50)

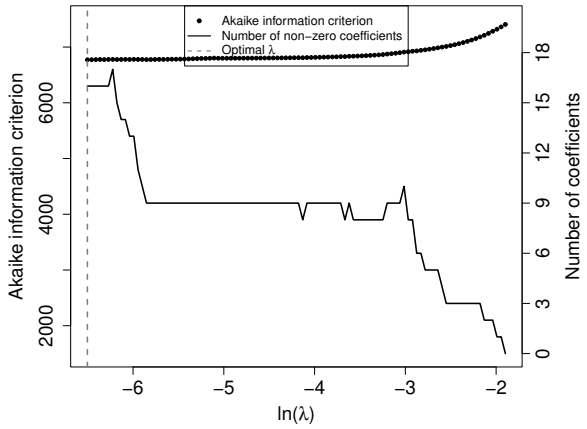


Figure A.2: Convergence of least absolute shrinkage and selection operator models