WORKING PAPERS IN ECONOMICS

No.08/11

LENE LUNDE AND KATRINE V. LØKEN

"HOW ARE YOU FEELING"? ASSESSING REPORTING BIAS IN A SUBJECTIVE MEASURE OF HEALTH BY QUANTILE REGRESSION

Department of Economics
UNIVERSITY OF BERGEN

"How are you feeling?" Assessing reporting bias in a subjective measure of health by quantile regression

By

Lene Lunde[1]
Department of Economics
University of Bergen
Box 7802, 5020 Bergen, Norway
lene.lunde@econ.uib.no

Katrine V. Løken
Department of Economics
University of Bergen
Box 7802, 5020 Bergen, Norway
katrine.loken@econ.uib.no

[1] Corresponding author: Lene Lunde, Department of Economics, University of Bergen, Box 7802, 5020 Bergen, Norway, lene.lunde@econ.uib.no, phone: +4755589238, fax:+475589210

Abstract

In this paper we investigate reporting heterogeneity in the Visual Analogue Scale (VAS) when it is used to measure current health status in cardiovascular patients. We provide a new framework to identify reporting heterogeneity using quantile regressions. EQ-5D responses are used as a proxy to control for objective health. The objectiveness of this generic measure is supported by other measures of actual health. The data comes from a Norwegian, health-related quality of life study. We find substantial evidence of reporting bias in VAS related to gender and education. For some quantiles we observe reporting heterogeneity related to age and weight.

Keywords: Reporting bias, quantile regression, health measurement

JEL Codes: I10, I12,C21

1. Introduction

The Visual Analogue Scale (VAS) is a well-known health measure which has a long history in psychometrics. In economics, it has been commonly applied to derive preference weights for hypothetical health states. However, more recently it has been used in health service research as a simple measure of individuals' self-reported rating of their own health state, or as a single health measure in large population or household surveys (Shumeli 2003). VAS is practical and straightforward, easy to understand and easy to use, but is nevertheless considered to be highly subjective. The respondents indicate, on a scale similar to a thermometer, how they rate their overall health. Answers to subjective measures like these may not mirror the individuals' latent health but rather the individuals' opinion on what ideal health is (Thomas and Frankenberg, 2000). The answer may also be influenced by the health status of persons with whom the individual identifies herself. Age, gender, income, religion and education are variables that are likely to influence the notion of ideal health or the health of an individuals' reference group. Systematic differences between an individual's actual health and their own indication of health in a subjective health measure are referred to as reporting bias, or reporting heterogeneity.

In this paper, we investigate reporting heterogeneity in VAS among different groups of cardiovascular patients. For the same level of health, general heterogeneity in responses to health measures should not vary by, for example, gender or socioeconomic status. Previous studies confirm the existence of reporting bias in subjective measures, primarily in the commonly used five-category self-assessed health (SAH) classification (Groot 2000; Humphris and Van Doorslaer 2000; Lindeboom and Van Doorslaer 2004). A previous study also indicates that diverse health measures have different reporting heterogeneity properties (Shmueli 2003). Even though VAS is a common health measure there has not been much emphasis on reporting heterogeneity in this measure. It is important to estimate whether there is any potential reporting bias, for instance, when VAS is used as an outcome measure in an economic evaluation. We may under- or overestimate the true effect of the intervention because individuals with the same level of latent health perceive their health differently depending on what kind of person they are, and thus report their health differently. For instance, if a young person reports higher VAS scores than an older person at any given health state, there are equity and efficiency issues, given that

the reported scores are used for resource allocation decisions. Further, we need to know about reporting bias when VAS is used as a health measure for a single population. Suppose we want to compare the health of two different populations using VAS. Assume that higher educated individuals report higher VAS scores than individuals with less education, even though they have the same level of latent health. Without any adjustment for reporting heterogeneity, we can easily conclude that the better health level is found among the population with the highest frequency of highly educated individuals. Similarly the socioeconomic gradient cannot be entirely related to differences in health when reporting heterogeneity is present. Without any adjustment, we may over- or underestimate inequity related to socioeconomic status when there is reporting heterogeneity.

Shmueli's study from 2003 gave evidence that VAS is sensitive to reporting bias related to economic status in the Israeli population. We want to further explore reporting heterogeneity in VAS by using a novel Norwegian data set on health-related quality of life in cardiovascular patients, and by employing some new techniques to identify reporting heterogeneity using quantile regressions. A better understanding of reporting bias will help us to obtain better estimates of an individual's latent health. Two main approaches have been used previously to identify reporting heterogeneity in subjective measures of health. One of the approaches uses a second measure of health to identify objective health and model reporting behaviour by hierarchically ordered probit models with cut-points depending on observed characteristics (Kerkhofs and Lindeboom 1995, 2004; Groot 2000; Lindeboom and Van Dorslaer 2004). The other approach identifies reporting heterogeneity by analysing variation in evaluation of given health states represented by vignettes (Jones at al 2007; King et al 2004; Salomon et al 2004). These identification strategies have been applied mainly to categorical health measures. We follow the main idea behind identification of reporting heterogeneity in a subjective self–reported health measures in Kerkhof and Lindeboom (1995, 2004), where the focus was on differential reporting behaviour across different labour market states. However, the VAS is a continuous variable and ordered probit models are therefore not a suitable approach in our cases, nor are vignettes available. We therefore use a standard ordinary least squares approach and quantile regression techniques. A conditional mean regression is a useful tool for summarizing the average relationship between a variable of interest and a set of regressors. However, it gives us

only a partial view of this relationship. Reporting behaviour might differ with respect to latent health; therefore it is essential to investigate different points on the health measure distribution. By using quantile regression, we obtain a better picture of the conditional mean regression relationship, but at different points in the conditional distribution of interest. We will show that with a quantile regression we acquire a better understanding of reporting differences across the whole distribution of the VAS, and we demonstrate that this method is suitable for analysing reporting heterogeneity in non- categorical subjective measures of health.

Finding a good measure of latent health is non-trivial. As already mentioned, in studies not using vignettes to identify latent health, a second measure of health is used, assumed to be an objective indicator of true health. For this analysis we use EQ-5D responses as a proxy for objective health, and group individuals according to these responses. The objectivity of the second measure used in previous studies to identify true health has often been taken for granted, and has not been much discussed or questioned (Groot, 2000; Humphris and Van Doorslaer, 2000; Lindeboom and Van Doorslaer, 2004; Van Doorslaer and Jones 2003). We stress the importance of having a real objective measure of health to identify reporting heterogeneity in VAS. Therefore, several other objective measures of health are used to support the objectivity of the EQ-5D responses, and we control for specific health indicators to verify that reporting heterogeneity is due to actual differences in reporting and not differences in latent health. We do not use the utility score derived by EQ-5D as a proxy for objective health, but rather responses to each of the attributes, due to potential heterogeneity in the valuation of health states. The EQ-5D score is valued with respect to a population tariff[2], as a proxy for mean preference at different levels on the different dimensions in EQ-5D. This preference element is removed by using responses instead of the score, and we obtain information about an individual's level of objective health.

Three sources of reporting heterogeneity can be distinguished: differences in socioeconomic status, differences in culture and differences related to specific health conditions (Shmueli 2003). In this paper, we test whether there is any reporting heterogeneity in VAS mainly related to the

[2] An EQ-5D health state may be converted to a single summary index by applying a formula that essentially attaches weights to each of the levels in each dimension. This formula is based on the valuation of EQ-5D health states from general population samples (www.euroqol.org)

first source. The dataset is not suited to measure reporting bias across cultures. Nevertheless, there is neither information nor reason to believe that there is much variation in culture/religion/immigration status in Norway. Even though we are dealing with a homogenous sample of patients we find it hard to identify individuals with different diagnoses with exactly equal objective health. Instead of testing whether there are reporting differences in relation to diagnoses, we use the main diagnoses of the respondents as health controls.

The paper is organised as follows: Section 2 discusses the data and variables used; section 3 presents the model and the econometric specification. Section 4 presents the results, while section 5 provides us with robustness tests, and finally in Section 6 we discuss our results and draw conclusions.

2. Data and variables

We use a novel data set obtained from a Norwegian cardiovascular quality of life survey conducted between October 2007 and June 2009 (Lunde and Wisløff 2011). The data set consists of three subsamples; i) all types of heart patients, ii) post stroke patients iii) uncomplicated hypertension patients. The recruitment of heart and stroke patients was done in co-operation with the Heart and Neurology divisions at Haukeland University Hospital, while the hypertension patients were recruited via five randomly selected GP centres. All data are based on postal questionnaires. A total of 1174 patients participated in the appraisal (47.3% of all invited patients). The patients' health status was measured by VAS, 15D and EQ-5D. The survey also contained a questionnaire covering general information about the respondents' socioeconomic status and health background.

The main variable investigated, the VAS score, is a continuous variable varying between 0 and 100[3]. Responses to the attributes of EQ-5D are used as a proxy of the individual's objective health. EQ-5D is a generic multi attribute utility instrument that is a widely used (Barton et al. 2008; Brazier et al. 2004; Holland et al.2004; Lamers et al. 2006; McNamee et al. 2005; Seymour

[3] For the analysis we have for simplicity transformed it into a 0-10 scale

et al.2009) and may be considered one of the most valid preference based measures used in economic evaluations[4]. The patient population is grouped according to EQ-5D responses. These health groups are used as controls for sorting out reporting heterogeneity that can potentially stem from real differences in objective health.

EQ-5D consists of five very general questions covering mobility, self-care, pain/discomfort, usual activities, and anxiety/depression. The questionnaire is found in the appendix, part A. Each of these questions has three different, predefined answers or statements. For example in the question about mobility, respondents are asked to indicate whether they have "no problems walking around", which is considered as a level 1 answer , "some problems in walking about", considered as a level 2 answer, or are "confined to bed", which is considered as a level 3 answer. We assume the EQ-5D responses to be objective because they consist of very specific and predefined statements. Since this health classification system is straightforward, indicators of health are easily verified in this system. An examination of the relationship between self-reported and physical assessments by Thomas and Frankenberg (2000) suggests that self-reports, and particularly those involving general questions, contain a good deal of information about the respondents' actual health.

When we divide individuals into health groups according to EQ-5D responses, the preferred approach would be to have a very fine split of individuals into groups that had exactly the same response to the same questions. The small sample size limits such a fine split of groups. To make reliable econometric analysis we had to go for a rougher definition of groups. We have chosen to split the individuals purely after response level, independent of question. Table 1 gives us an overview of how the split of individuals was done. Group 1 consists of those who had a level 1 response to all questions. Group 2 consists of those who have one answer at level 2, and the remaining answers are level 1. Group 3 consists of those with two answers at level 2, and three answers at level 1. Group 4 and 5 are necessarily more mixed, due to the small number of observations. Group 4 consists of those who have answered with a level 2 three, four or five times, while group 5 consists of those with one or two level 3 answers. Individuals with three or

[4] EQ-5D is for instance recommended for use in cost effectiveness analyses by institutions such as the National Institute of Health and Clinical Excellence (NICE) in the UK, and the health care insurance board in the Netherlands.

more level 3 answers are omitted (8 individuals), because the variance in objective health for these are too large, and there are too few individuals to constitute a group.

[*Table 1 about here*]

We verify that individuals within the same health group have similar levels of objective health by using two other measures of objective health. These two measures are also based on self-report; we believe, however, that these measures are more objective than a health measure such as VAS. The first measure is concerned with how bothered patients are by chronic diseases. The second concerns diseases other than their primary diagnosis and any chronic diseases. In both of the measures, the patients are first asked whether they suffer from any chronic disease or non-chronic disease other than their main diagnosis. If the patients answered yes to any of these questions, they were given four different alternatives to describe how bothered they are in daily activities by the chronic disease or the other diseases. We categorised the patients as having a substantial chronic disease, or a substantial other disease if they answered that the disease bothered them highly or to some extent.

Even though we verify that the individuals in each of the defined health groups have the same level of objective health there can still be differences in health that have an impact on the reporting behaviour. In our regressions we therefore also control for diagnostic group (stroke, heart or hypertension patient) to capture the differences in the experience of these diseases. We also include an extensive set of control dummies based on responses to certain attributes in the EQ-5D to make sure that we capture differences in reporting and not real differences in health. We constructed a dummy variable for those who have reported that they have experienced some or major problems related to mobility, performance of usual activities, anxiety/depression and pain respectively.

We investigate whether there are any reporting heterogeneity in gender, age, educational level and body mass index (BMI). These variables are therefore considered as explanatory variables in

our regressions. Individuals aged 67 or above, are denoted old[5], and individuals below 67, are denoted young. We label those who have a university or college degree as highly or well educated, while those without any degree are labeled as poorly or low-educated. Individuals are defined to be overweight or obese if they have a Body Mass Index (BMI) of 25 or more.[6] A BMI between 20 and 25 is defined here as normal. Note that we have focused on individual characteristics that are plausibly exogenous to the specific diagnosis. Gender and age are deterministic and educational attainment is completed long before these individuals fall ill. BMI is the only variable that might have changed after the illness and we have tested whether excluding this variable changes the main results. We will not consider income and marital status, as they are potentially endogenous with respect to the diagnoses.

3. Model and econometric specification

3.1. Underlying model

We will follow the approach introduced by Lindeboom and Kerkhofs (1995 and 2004).

The patients' subjective response to a question regarding their overall health is denoted H^S. A latent variable of true health generates the response to the subjective measure of health. We denote the true health H^*.

$$H^S = g(H^*, X, \varepsilon; \beta) \qquad (1)$$

The relationship between H^s and H^* may vary across individuals or subgroups of individuals according to socioeconomic and demographic variables within a population. There is reporting bias if people with the same level of true health (H^*) report the subjective measure H^S systematically differently from each other. In equation 1 the individual characteristics are denoted by *X*. β captures the effect of X on the subjective health measure. Controlling for true health may give us an idea of how much of the variation in subjective health that comes from reporting bias. β is then interpreted as a parameter of pure reporting bias.

[5] The dividing age between young and old is set to 67 because this is the compulsory retirement age in Norway.
[6] This classification is in line with the World Health Organization's classification, where a BMI $\geq$ 25 is categorised as overweight and a BMI $\geq$30 is categorised as obesity

In reality H^* is unobserved and difficult to measure. We therefore proxy true health by the responses to the different EQ-5D questions, which we assume represent an objective measure of health that captures the underlying true health, H^o.

$$H^* = f(H^o, z, \eta; \alpha) \quad (2)$$

If the objective health measure fully captures true health H^o, equation 2 will be an identity. Normally z will correct for the dissimilarity between H^* and H^o. In z we include diagnoses and other health variables and individual characteristics. We will assume that the objective measure of health, H^o, will capture the true underlying health conditional on the z. We will test this assumption later in the paper.

3.2. Empirical implementation

As a benchmark, we run the conditional mean-regression as

$$VAS_i = \beta_0 + \beta_2 X_i + \beta_1 1(H^o = k) + \beta H_i + \varepsilon_i, k = 1,...,5 \quad (3)$$

The VAS score is the subjective measure of health. The Xs presents the different background characteristics (gender, age, educational attainment and weight).

The patients are divided into five groups according to the distribution of answers to different questions on objective health, H^o. We create dummy variables for belonging to each of the groups. In the absence of reporting bias, there should not be any systematic differences in reporting behaviour related to different socioeconomic groups, age or gender, since all respondents have the same level of objective health. In addition we control for diagnoses and other objective health measures, H_i. ε_i is the error term assumed to have zero conditional mean.

As the main approach, we will use the quantile regression method. The reason for focusing on quantile regressions is to capture heterogeneity in responses across the distribution of the VAS. An absence of reporting bias in the mean VAS score could stem from reporting bias at one end of the distribution and opposite reporting bias at the other end of the distribution, thereby cancelling

each other out. We would then need quantile regressions to capture the complete picture of reporting bias.

We follow Koenker and Bassett (1978) and run the quantile regression

$$VAS_i = \beta_0^{(p)} + \beta_1^{(p)} X_i + \beta_2^{(p)} 1(H^o = k) + \beta_3^{(p)} H_i + \varepsilon_i^{(p)}, k = 1,...,5 \qquad (4)$$

Where i denotes individuals, $0 < p < 1$ indicates the proportion of the population with VAS scores below quantile p. The p[th] conditional quantile is then given by

$$Q^{(p)}(VAS) = \beta_0^{(p)} + \beta_1^{(p)} X_i + \beta_2^{(p)} 1(H^o = k) + \beta_3^{(p)} H_i + Q^{(p)}(\varepsilon), k = 1,...,5$$

where p = 25th, 50th and 75th quantile.

We assume that $Q^{(p)}(\varepsilon) = 0$, however we allow $\varepsilon_i^{(p)}$ to vary across quantiles. We will focus on the 25th, 50th and 75th percentiles. The different betas will then be the estimated difference in VAS between the different explanatory variables at quantile p. We will also perform the analysis for the 20th, 40th, 60th and 80th quantiles.

There are relatively small differences between patients within a group; they all have some kind of cardiovascular disease. We expect the degree of heterogeneity in the data to be low. This gives us an advantage and allows us to keep the econometric framework rather simple. When reporting bias has been investigated in earlier studies, cross-section data or large household surveys have been used. In such large samples, the respondents may for instance be either perfectly healthy or very ill and we therefore expect the degree of heterogeneity in such data to be larger. Heterogeneity needs special econometric considerations and such samples need much finer sub-samples than we do. We run robustness tests to verify that the sample is indeed homogenous and that we can argue that the results represent reporting heterogeneity and not differences in objective health.

The last empirical step we will conduct is to produce a cleansed VAS score. We will compare the actual distribution of VAS to the distribution of VAS after subtracting the coefficients for differences in reporting behaviour across subgroups. The procedure is as follows: for the mean regression we run (3) and store the β_2 coefficients which represent the reporting bias for females,

older individuals, well-educated and overweight individuals. From the individuals' VAS scores we subtract the set of coefficients pertaining to these individual characteristics. For example for a well-educated young woman with normal weight we would subtract the coefficients for female and well-educated from the VAS score, however not for old and overweight. There are potentially 16 different combinations of the four explanatory dummy variables. We then compare the actual distribution of the VAS using density estimation with a standard kernel (epanechnikov) and bandwidth to the distribution of the cleansed VAS. It is important for our estimations to use the same kernel and bandwidth to make the plots comparable.

For the quantile regressions we store the estimates for each explanatory variable, for females, older individuals, higher educated and overweight for the 25th, 50th and 75th quantile. Then we separate the sample in three equally sized groups and subtract the coefficients from the individuals' VAS according to the individual statuses and whether the VAS is in the lower, middle or upper part of the distribution. We then compare the density of the actual VAS to the quantile cleansed VAS. The idea here is that the betas for each of the explanatory variables may differ in different parts of the distribution. Ideally we would have different estimates for each centile across the whole distribution. However, our sample limits us to focus on three subgroups and three quantiles.[7]

4. Results

4.1. Descriptive Statistics

Table 2 provides some descriptive statistics for our sample. The average age is 63 years; the youngest is 19 years old, while the oldest is 97. Nearly half of the sample is 67 or older (44%), and thus retired. 66% are men, and 34% have received a university or college degree. 64% are considered obese or overweight. 31% of the patients have suffered a stroke, 39% are diagnosed with hypertension and 30% have heart-related diseases.

[7] An alternative approach would be to compare the predicted distribution of the VAS to the cleansed predicted distribution.

[*Table 2 about here*]

Table 3 shows the composition of individuals in each of the defined health groups. The lowest mean age is found in the best health group, while the highest average age is found in group 4. Further, we find the highest proportion of males, those who are well educated, obese or overweight, and those with hypertension in health group 1. Health group 1 also has the lowest percentage of old individuals, stroke patients, and heart patients. Health group 5 (the group with poorest health) has a low percentage of males, the lowest proportion of well-educated, obese individuals and also hypertension patients. The same group has the highest proportion of stroke and heart patients. This gives us evidence that individuals with best health according to EQ-5D responses are males, highly educated, and those who are overweight /obese, and finally those who are diagnosed with "only" hypertension. It is not surprising that individuals with higher socioeconomic status typically have better health than those at low status. This is in line with the estimated relationship between socioeconomic status and health, indicating that lower socioeconomic status is associated with poorer health (Humphris and Van Doorslaer, 2000; Kunst and Mackenbach, 1994; Lynch et al., 1996; Townsend and Davidson, 1982). Further, it is well-known that women experience more illness than men, even though they live longer. Women respond to illness differently and use more health services than men (Langely, 2003). It is also reasonable that the health of hypertension patients is better than the health of those who have suffered a stroke or have had a heart attack.

[*Table 3 about here*]

4.2. The distribution of VAS

Table 4 gives us an overview of the distribution of VAS scores for the whole sample (the first column) and is divided by subgroups.

[*Table 4 about here*]

It is clear that the mean score of the VAS is following the defined health groups. Mean VAS score decreases with the health of each group. The difference in mean VAS score between group 1 and group 2 is much smaller than the differences between group 2 and 3, group 3 and 4 and group 4 and 5. The largest differences in mean score are found between group 3 and 4, and 4 and 5. This indicates that the variance in how the VAS is reported increases with lower levels of objective health. The standard error also increases by health groups, giving us the same indication. The mean values of VAS for each subgroup differ slightly within each health group; however, the largest variance in mean scores is found in the poorest health groups (group 4 and 5). For all sub groups, the mean VAS score decreases monotonically going from group 1 to 5. Females have a higher mean VAS score than males except for group 5, where males and females have the same scores. Well-educated individuals have higher scores than less educated, except for group 5. We see the same pattern for those who are overweight. They have the lowest means score compared to those who are overweight in all groups, except in the poorest. Older individuals have higher mean VAS scores than younger except in group 3 and 5. Hypertension patients always have a higher mean VAS score than the two other patient groups, while there is a more mixed picture between heart and stroke patients.

4.3. How objective is the health group definition?

The main assumption underlying the method by which to identify reporting heterogeneity is that the objective health measures indicate true underlying health. Individuals with the same level of objective health according to EQ-5D responses should not report their health in VAS systematically different from one other if there is no reporting heterogeneity. We use this section to verify that this assumption is likely to be corroborated. We first base our assumption on the amount of information we have about a person's underlying health. The health groups we use as controls are defined on the basis of a whole set of objective health questions. In addition, we run all the regressions with some extra health control variables. These include diagnoses and

dummies indicating major problems with mobility, doing usual activities, anxiety/depression and pain. We want to stress that we have a homogenous sample, since all the recruited individuals are patients having experienced stroke, hypertension or some kind of heart problems. It is therefore more plausible that we can capture objective health using the health information we have in our data, than if we had a population data set where health states have much more variance, and it is harder to evaluate differences of health for populations with many individuals reporting no problems in the areas of mobility, self-care, depression, pain etc.[8]

Secondly, we test the objectivity of the health groups by using other indicators of health that can be viewed as more direct measures of underlying health than VAS. The questions regarding chronic diseases and other diseases are quite specific and are assumed to reflect objective health because there is a lower likelihood that the respondents will over- or under-report their own health through answers to such questions. These measures are supposed to record the facts of an individual's current or past medical condition, rather than their perceptions of their physical state. Earlier research has indicated reporting errors in self- reports of specific aliments. These, however, are related to labour market status; people tend to justify their absences from work (Baker et al., 2004). Answers to questions related to specific aliments as general diseases and chronic diseases may vary substantially across different subgroups, but we find it unlikely that they are affected by socioeconomic status. If the health groups we have defined and the health controls capture objective health, they should be able to eliminate these differences between sub-groups. Table 5 shows the results for chronic diseases. The first column shows the mean regressions without including any of the control variables for underlying health. Here we see that there are differences in chronic diseases across sub-groups of the population. Women are more likely to have chronic diseases than are lower educated and normal weight (compared to overweight) individuals. In column two, we control for diagnoses and objective health groups, and we see that the earlier correlations are reduced substantially. There are no significant results and the magnitudes are significantly lower than column one. Column three adds the additional health controls, and we see that the results are almost unchanged compared to column two.

[8] There is potentially a lot of variation in underlying health also for individuals without problems with these major categories. Since we have individuals who are much more likely to have issues with one or more of these attributes, we obtain more information to identify underlying health.

[*Table 5 & 6 about here*]

We do the same exercise in Table 6 using other diseases as an outcome variable. Although not as pronounced as chronic diseases, women are also more likely to have other diseases, and the pattern for low educated and normal weight individuals is also the same. When adding health controls in column two and three, we see that these correlations disappear. There is no longer evidence that women, low educated and normal weight individuals have more diseases than the main diagnosis. This exercise indicates that we have a good proxy for objective health.

4.4. Quantile regressions

In this section we focus on the main results, reporting heterogeneity in VAS. The results from the mean regression are first presented in table 7. The first column displays the mean regression without any control variables for underlying health. We see a pattern, although not always statistically significant, that women report higher health scores in VAS than men; older patients report slightly lower scores than younger individuals; the highly educated report much higher VAS scores than lower educated, and the overweight report higher than normal weight individuals. These results do not represent reporting bias as they confound potential reporting bias with individuals' actual health. For example it is highly likely that higher educated individuals also have better underlying health than lower educated individuals and therefore report higher on the VAS. We also saw this from the results on chronic diseases; higher educated individuals were less likely to have chronic diseases. However since we saw that our objective health control variables were able to eliminate all these differences in health we can be more certain that columns two and three of table 7 give us indications of reporting bias. Here we see that females are much more likely than males to report higher values on VAS. There are no significant effects across age groups, while we still see that higher educated individuals report higher VAS than the lower educated. Nevertheless, this effect is more than half of the correlation which shows that some of the correlation stemmed from actual health. However, our results

indicate that reporting heterogeneity is likely to explain the difference between high and low educated. There is no significant effect with respect to BMI although we see that the results from the first column are reversed. When controlling for the different aspects of health, we see that individuals who are overweight are likely to report lower standards of VAS than those who are not overweight.

[*Table 7 & 8 about here*]

Mean regression may not provide us with information about heterogeneity in responses across the distribution of the VAS. In Table 8, we report results for the quantile regression. We discuss column two and three directly as we have already discussed the correlations in detail and seen that there are large differences between the correlations and the regressions with objective health control variables, and that it is difficult to disentangle variation in real health and variation in reporting without controlling for objective health. From column two and three, we see that females report higher VAS than males across all of the quantiles. There do not seem to be any reporting differences between males and females within the distribution of VAS. This is also confirmed with a Wald test giving an F-statistics of 0.69, so we cannot reject the null hypotheses that the coefficients are similar across the quantiles. For age, we remember that the mean regression was not able to detect any significant differences in reporting. However, we see that for the 75th quantile, older individuals report higher VAS scores than the younger. We can also see that there are no differences at all in reporting between older and younger individuals in the 25th and 50th quantiles.[9] Further we find that those with higher education report significantly higher VAS scores than lower educated individuals in the 50th and 75th quantiles. Nevertheless, we do not find large differences in reporting across the quantiles (F-statistics is only 0.13). When we look at weight, we remember that we could not find any significant differences in reporting in the mean regression. By using quantile regression we see that there are significant differences in the median; in this quantile, the overweight report a lower standard of VAS than those who are of normal weight (F-statistics of 1.21). Although the tests cannot detect statistically significant

[9] Again a joint Wald test with F-statistics of 1.27 cannot detect statistically significant results, although separate comparisons for the 75th quantile compared to the 25th and 50th indicate that the 75th quantile is different.

effects across the quantiles, the quantile regression is still extremely useful since it is able to detect effects that the mean regressions were not able to pick up. Finally we observe that there are not large differences by running the regressions with an extra set of health controls, which indicates that the health groups are well defined.[10]

4.5. Cleansed estimates

We see from Figure 1 that the actual distribution of the VAS has different peaks which suggest that the distribution differs across different subgroups. When we clean the score using the mean regression with the approach discussed in 3.2., we see a smoother density with the distribution shifted slightly to the left. In figure 2, we see that using the quantile regression does not give very different results, which corresponds with the earlier main results indicating that results do not vary significantly across quantiles. However, comparing Figure 2 and Figure 1, we see that the distribution is shifted even more to the left especially from very high values to middle-to-high values of the VAS.[11]

[*Figure 1 & 2 about here*]

5. Robustness tests

We wanted to study quantiles simultaneously to take into account unobserved heterogeneity across the whole distribution of VAS. The most desirable would have been to run the regressions per decile; however the sample size restricted our focus to fewer points in the distribution of the VAS. Therefore we have chosen to focus on the 25th, 50th and 75th quantiles. To verify that the results are not peculiar to the choice of these points, we additionally perform the analysis using the 20th, 40th, 60th and 80th quantiles. However, as we have already seen, there is a lot of heterogeneity in responses across the VAS, and we therefore could also expect the effects in other quantiles to differ slightly. We expect the 40th and 60th quantiles to be close to the median

[10] We have also tried to redefine the health groups, especially the last two groups which lump together more different responses to the EQ-5D questions without any significant changes to our results.
[11] Using the alternative approach discussed in 3.2. gives very similar results.

score, while the 20th and 80th should be closer to the two ends of the distribution. However, we expect to verify the results by finding a similar pattern when we move the quantiles slightly.

[*Table 9 about here*]

Table 9 shows that the results for the other quantiles are consistent with the earlier pattern and that our main conclusion holds. Females and the higher educated always report higher VAS across the whole distribution. Older individuals report higher only in the 80th quantile, which is consistent with the same effect found for the 75th quantile. We find those who are overweight to report lower standards of VAS than normal weighted in the 40th quantile, which is consistent with the effect we found at the median. Remember that we could not find any significant differences in reporting related to age and weight in the mean regression.

6. Discussion and concluding remarks

In this paper, we have demonstrated reporting heterogeneity in the mean score of VAS for gender and education level. The main contribution has been to analyse reporting behaviour within quantiles of the distribution of VAS. This has been an important task because it has shown us that there are reporting differences across the distribution of VAS which cannot be detected by using mean regression. We were unable to find any significant differences in reporting related to body weight and age in the mean regression; nevertheless, we found significant differences in some of the quantiles. Earlier studies did not demonstrate that reporting heterogeneity in subjective measures of health is dependent on the level of health reported. Nor has it been previously demonstrated that quantile regression is a suitable tool for measuring reporting heterogeneity in non-categorical subjective health measures. In general, women tend to report higher standards of VAS scores than men, and well educated also generally report higher standards of health than less educated individuals. Older individuals report higher standards of VAS in the upper part of

the distribution of VAS, while individuals who are overweight report lower standards of health than those who are normal weight around the median.

Our findings are not necessarily valid for the population in general, or for other patient groups. We have used a small sample of a very specific type of patient; we therefore must be careful in generalising the results. However, we have the advantage of a homogenous sample of patients, which is important for the internal validity of the results. Some of the findings are consistent with earlier research on other subjective measures of health. Lindeboom and Kerkhofs (2004) found clear evidence that males and females, as well as individuals in different age groups, report systematically different standards of health in the self- assessed health question when they have the same level of objective health. This study was undertaken in Canada, and objective health was measured with the Health Utility Index (HUI). However, they did not find any reporting heterogeneity in relation to education level. In a study where chronic conditions and disabilities worked as a proxy for true health, Groot (2000) found that women, younger individuals, and those with lower education reported systematically lower health standards in SAH for a given chronic condition or disability. Nevertheless, we found that women report higher VAS scores than men. This illustrates that we can expect different directions of reporting bias related to different socioeconomic groups depending on the subjective measure of health being used.

We found evidence that older patients report higher VAS scores than younger individuals, even though they have the same level of objective health. One possible explanation could be differences in the individuals' reference groups. It is likely that an old person will compare herself with persons of similar age, and that older individuals normally experience more health problems than younger. With age, one is expected to experience some problems related to mobility, and one's perception of health will then be affected by one's perception of how mobility should be. Younger individuals who suffer strokes and have problems walking are not expected to have mobility problems, and they identify themselves with individuals in quite good health. The result of this might be that they under-report their actual health in a subjective measure. Note that this paper has focused on the reporting differences across different subgroups; we have not attempted to predict the causes of reporting heterogeneity. The differences can be due to both unobserved and observed characteristics of the sub-groups. For example, there could be many reasons why females tend to report better health than males, even

after controlling for objective health. Subjective health measures do not have a natural reference point. It is likely that the reference point of well-being is determined by the individual's specific situation and characteristics, rather than latent health. We believe that individuals with the same level of objective health perceive their health to be different from others', and that the difference depends on their reference group. The type of reference group an individual has is unobserved; however, we do not find it likely that the types of reference groups are likely to vary systematically across different sub-groups. This gives us reason to believe that there are other factors within the various subgroups that influence how their subjective health is reported.

Whatever the story behind reporting heterogeneity, decisions concerning the allocation of resources and appropriate policies require unbiased indicators of health across different population groups. Much research in social science is based on self-reported surveys. Self-assessment of health or quality of life is used in many areas, such as in studies on labour supply and retirement decisions, studies on the effect of health on wages, and studies of equity in health. When using subjective measures of health, we need, therefore, to be aware of reporting bias. We have shown that the subjective health measure is sensitive to the patient population and the composition of individuals. This limits the ability to generalise results– and to make comparisons with other populations. However, we have demonstrated how to identify reporting heterogeneity, and how to cleanse the estimates of the subjective measures. The main message is that whenever the VAS is used as a measure of health, we need to adjust for the determinants of reporting heterogeneity. An adjustment towards the mean may not be sufficient.

Acknowledgements

The authors are very grateful to chief physician Dr. Halvor Næss, division of Neurology, chief physician Jan Erik Nordrehaug, and Kjersti Oterhals, heart division at Haukeland University Hospital for facilitating the data collection. We also thank all the GP's that took part in the data collection. We are especially grateful to Dr. Eivind Meland for helping in the start of survey and Professor Andrew Jones for taking part in early discussions about the paper. We are also in debt Professor Jan Erik Askildsen, Professor John Cairns and Professor Ivar Sønby Kristiansen for

advices and valuable comments. Finally, we would like to thank Health Economics Bergen and the Department of Economics, University of Bergen for financial support.

References

Baker M, Stabile M, Deri C. What do Self-Reported, Objective, Measures of Health Measure? *Journal of human resources, 39(4):1067–1093, 2004.*

Barton G.R, Sach T.H, Avery J.A, Jenkinson C, Doherty M, Whynes D.K, Muir K.R. A comparison of the performance of the EQ-5D and SF-6D for individuals aged $\geq$ 45 years. *Health Econonomics, 17(7):815–832, 2008*

Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics 13:873–884, 2004*

Groot W. Adaption and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics 19: 403–420, 2004*

Holland R, Smith RD, Harvey I, Swift L, Lenaghan E. Assessing quality of life in the elderly: a direct comparison of the EQ-5D and AQoL. *Health Econ., 13:793–805, 2004*

Humphris K, Van Doorslaer E. Income related health inequality in Canada. *Social Science and Medicine 50: 663–671, 2000*

Jones A.M, Rice N, Bago d'Uva T, Balia S. Applied Health Economics, first edition. *Routledge, Abingdon, 2007*

Kerkhofs M.J, Lindeboom, M. Subjective health measure and state dependent reporting errors. *Health Economics, 4 : 221–235,1995*

Kerkhofs M.J, Lindeboom, M. Subjective health measures, reporting errors and endogeneity in the relationship between health and work. *CEIS Tor Vergata-Research Paper Series, Vol 16, NO. 46 January 2004*

King G, Murray C.J.L., Salomon J, Tandon, A. Enhancing the validity and cross- cultural comparability of measurement in survey research. *American Political Science Review, 98:184–191, 2004*

Koenker R and Basset G. Regression Quantiles. *Econometrica 46:1, 1978*

Kunst A.E, Mackenbach J.P. The size of mortality differences associated with educational level in nine industrial countries. *American Journal of Public Health 84: 932–937, 1994*

Langely R. Sex and Gender differences in health and disease. *Carolina Academic Press. Durham, North Carolina 2003*

Lamers L.M, Bouwmans C.A.M, van Straten A, Donker M.C.H, Hakkart, L. Comparision of EQ-5D and SF-6D utilities in mental health patients. *Health Econonomics 15:1229–1236, 2006*

Lindeboom M, Van Doorslaer E. Cut-point shift and index shift in self-reported health. *Journal of Health Economics 23: 1083–1099, 2004*

Lunde L, Wisløff T. "Does choice of outcome measure matter in a cost effectiveness analysis of antihypertensive treatment??" *Working paper, Department of Economics, University of Bergen, 2011*

Lynch J.W, Kaplan G.A, Cohen R.D, Tuomilehto J, Salonen J.T. Do cardiovascular risk factors explain the relationship between socioeconomic status, risk of all-cause mortality, and acute myocardial infarction? *American Journal of Epidemiology 144: 934–942, 1996*

McNamee P, Seymour J. Comparing generic preference-based health-related quality of life measures: advancing the research agenda. *Expert Rev. Pharmacoeconomics Outcomes Res., 5(5), 567–581, 2005*

Salomon J, Tandon A, Murray C.J.L., and WHSPSC Group. Comparability of self-related health: cross sectional multi-country survey using anchoring vignettes. *British Medical Journal, 328:258, 2004*

Seymour J, McNamme P, Scott A, Tinelli M. Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. *Health Econonomics, 19(6):1229–1236, 2010*

Shmueli A. Socio-economic and demographic variation in health and in its measures: the issue of reporting heterogeneity. *Social Science and Medicine* 57: 125–134, *2003*

Thomas D, Frankenberg E. The measurement and Interpretation of Health in Social Surveys. *Rand, Labor and Population Program, Working Paper Series 01–06, 2000*

Townsend P, Davidson N. Inequalities in Health: The Black Report. *Penguin Books. Harmondswort, Middelsex 1982*

Van Doorslaer E, Jones A.M. Inequalities in self-reported health: validation of a new approach to the measurement. *Journal of Health Economics 22: 61–87, 2003*

www.euroqol.org/eq-5d/valuation-of-eq-5d.html Bergen, *July 2011*

FIGURES

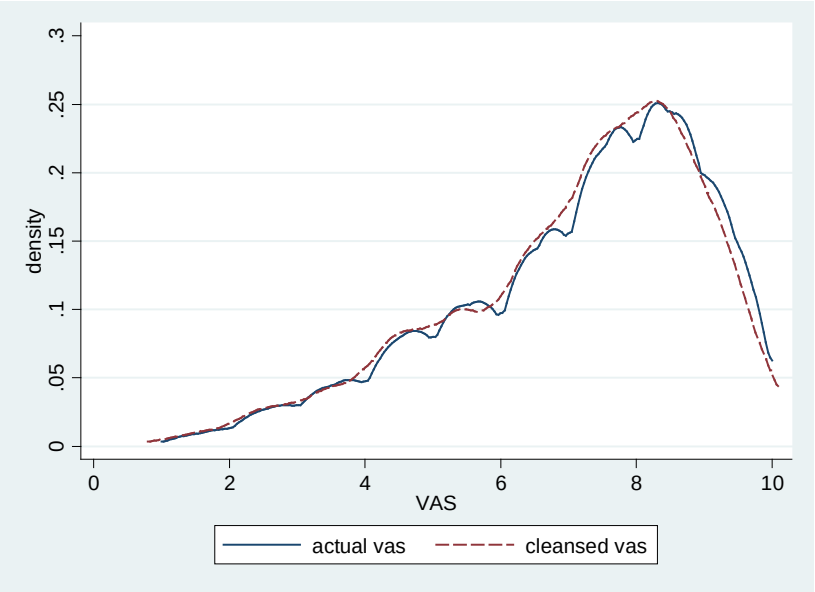

Figure 1: Density of actual VAS compared to cleansed VAS using coefficients on female, old, higher educated and overweight from the mean regression.

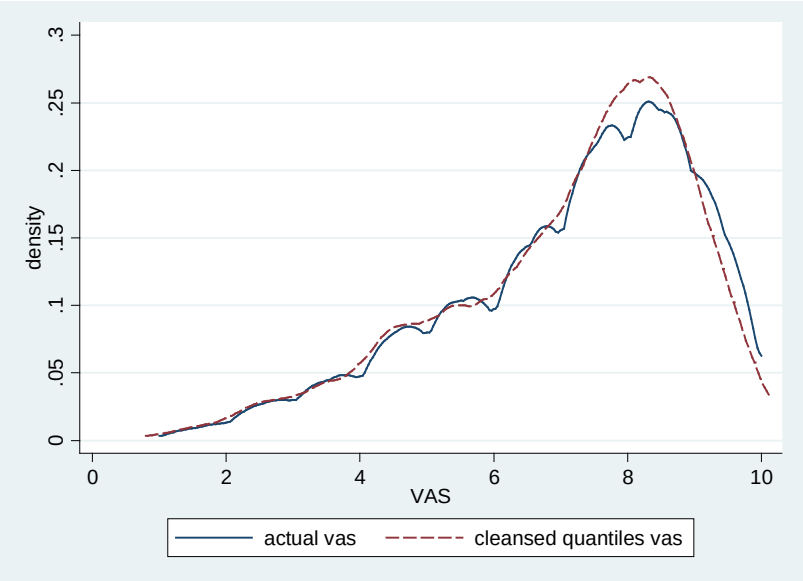

Figure 2:Density of actual VAS compared to cleansed VAS using coefficients on each of the quantiles, 25th, 50th and 75th on female, old, higher educated and overweight and distributed to three equal groups defined by the distribution of the VAS.

TABLES

EQ-5D LEVEL	1	2	3
Group 1	5	-	-
Group 2	4	1	-
Group 3	3	2	-
Group 4	2	3	-
	1	4	-
	-	5	-
Group 5	4	-	1
	3	1	1
	2	2	1
	1	3	1
	-	4	1
	1	2	2
	2	1	2
	-	3	2

Table 1. Health groups defined after response level in EQ-5D

Variable	Obs	Mean	Std. Dev.	Min	Max
Age	892	63.17	12.37	19	97
Male	931	.66	.473	0	1
Old	931	.44	.497	0	1
Well educated	931	.34	.472	0	1
Overweight/Obese	931	.64	.479	0	1
Stroke	931	.31	.461	0	1
Hypertension	931	.39	.488	0	1
Heart	931	.30	.460	0	1

Table 2. Descriptive statistics

Health Group	1	2	3	4	5
Age	62	63	63	66	64
Male	73%	68%	63%	57%	59%
Old	37%	44%	45%	54%	42%
Well educated	43%	35%	25%	30%	17%
Overweight/Obese	70%	64%	63%	61%	53%
Stroke	24%	27%	34%	39%	40%
Hypertension	53%	44%	32%	23%	21%
Heart	23%	29%	34%	38%	38%

Table 3. Group composition

Mean VAS	Mean	Male	Female	Old	Young	Well Educ.	Less Educ.	Overw/ Obese	Normal weight	Hypert.	Heart	Stroke
Group 1	8.34	8.23	8.64	8.44	8.28	8.43	8.27	8.29	8.45	8.44	8.06	8.38
N =312	(1.19)	(1.21)	(1.08)	(1.16)	(1.20)	(1.09)	(1.26)	(1.13)	(1.32)	(1.08)	(1.40)	(1.17)
Group 2	7.80	7.69	8.06	7.88	7.75	8.01	7.69	7.70	8.00	7.98	7.69	7.65
N= 209	(1.27)	(1.29)	(1.18)	(1.29)	(1.25)	(0.99)	(1.39)	(1.26)	(1.27)	(1.14)	(1.35)	(1.33)
Group 3	6.94	6.77	7.23	6.80	7.03	7.17	6.86	6.90	6.99	7.03	6.82	6.99
N=142	(1.45)	(1.52)	(1.28)	(1.53)	(1.39)	(1.52)	(1.42)	(1.21)	(1.79)	(1.52)	(1.27)	(1.57)
Group 4	5.76	5.57	6.02	5.80	5.71	6.55	5.42	5.64	5.95	6.21	5.63	5.62
N=181	(1.73)	(1.67)	(1.80)	(1.72)	(1.75)	(1.81)	(1.58)	(1.72)	(1.74)	(1.58)	(1.82)	(1.70)
Group 5	4.80	4.80	4.8	4.41	5.08	4.22	4.92	5.12	4.44	5.45	4.95	4.31
N=87	(1.96)	(2.07)	(1.82)	(1.99)	(1.91)	(1.70)	(2.00)	(2.08)	(1.78)	(2.06)	(1.83)	(1.97)

Table 4. VAS score by health group and sub-group. Standard deviation in brackets

Dependent variable:	(1)	(2)	(3)
Chronic diseases			
Female	.095***	.032	.034
	(.031)	(.029)	(.029)
Old	-.011	-.030	-.039
	(.030)	(.027)	(.027)
High educated	-.087***	-.041	-.042
	(.031)	(.029)	(.028)
Overweight	-.067*	-.038	-.044
	(.031)	(.028)	(.028)
Diagnoses		x	x
5 health groups		x	x
Extra health controls			x
N	931	931	931

Table 5: Mean regressions, chronic diseases. Standard errors in brackets.
*10 % significance level, **5 % significance level, ***1% significance level.

Dependent variable:	(1)	(2)	(3)
Diseases			
Female	.057*	-.030	-.029
	(.036)	(.031)	(.031)
Old	.036	.014	.007
	(.033)	(.029)	(.029)
High educated	-.035	.023	.023
	(.034)	(.030)	(.030)
Overweight	-.021	.001	-.002
	(.033)	(.030)	(.030)
Diagnoses		x	x
5 health groups		x	x
Extra health controls			x
N	931	931	931

Table 6: Mean regressions, **Diseases**. Standard errors in brackets.
*10 % significance level, **5 % significance level, ***1% significance level.

Dependent variable: VAS	(1)	(2)	(3)
Female	.120	.331***	.340***
	(.130)	(.104)	(.101)
Old	-.071	.027	.069
	(.126)	(.098)	(.096)
High educated	.725***	.359***	.362***
	(.131)	(.102)	(.010)
Overweight	.145	-.106	-.079
	(.129)	(101)	(.098)
Diagnoses		x	x
5 health groups		x	x
Extra health controls			x
N	931	931	931

Table 7: Mean regression, VAS. Standard errors in brackets.
*10 % significance level, **5 % significance level, ***1% significance level.

Dependent variable: VAS	(1)	(2)	(3)
Female	0	.375**	.300**
25th quantile	(.185)	(.155)	(.149)
	.400*	.367***	.400***
50th quantile	(.232)	(.111)	(.115)
	.200	.250**	.271**
75th quantile	(.180)	(.119)	(.112)
Old	-.100	0	0
25th quantile	(.232)	(.141)	(.139)
	0	-.100	0
50th quantile	(.185)	(.097)	(.108)
	0	.175*	.172*
75th quantile	(.150)	(.100)	(.101)
High educated	1***	.250	.200
25th quantile	(.222)	(.172)	(.165)
	.900***	.300**	.267**
50th quantile	(.226)	(.119)	(.120)
	.500***	.250**	.229**
75th quantile	(.150)	(.107)	(.102)
Overweight	.900***	-.125	-.100
25th quantile	(.319)	(.179)	(.166)
	0	-.333***	-.267**
50th quantile	(.206)	(.126)	(.125)
	-.200	-.175*	-.114
75th quantile	(.178)	(.100)	(.104)
Diagnoses		x	x
5 health groups		x	x
Extra health controls			x
N	931	931	931

Table 8: Quantile regressions. Standard errors in brackets.
*10 % significance level, **5 % significance level, ***1% significance level.

Dependent variable: VAS	(1)	(2)	(3)
Female	0	.400**	.300*
20th quantile	(.217)	(.170)	(.177)
	.300	.450***	.400***
40th quantile	(.214)	(.136)	(.141)
	.033	.440***	.389***
60th quantile	(.156)	(.113)	(.112)
	.300*	.200*	.320***
80th quantile	(.159)	(.119)	(.116)
Old	0	0	0
20th quantile	(.235)	(.163)	(.159)
	0	-.050	0
40th quantile	(.251)	(.107)	(.107)
	-.033	.020	.056
60th quantile	(.167)	(.103)	(.095)
	0	.200*	.189*
80th quantile	(.153)	(.116)	(.113)
High educated	1.3***	.300	.300
20th quantile	(.317)	(.195)	(.183)
	.800***	.300**	.300**
40th quantile	(.228)	(.152)	(.130)
	.467***	.440***	.333***
60th quantile	(.171)	(.131)	(.117)
	.300*	.300***	.260**
80th quantile	(.171)	(.107)	(.108)
Overweight	.500*	-.100	-.200
20th quantile	(.228)	(.184)	(.186)
	0	-.250*	-.300*
40th quantile	(.290)	(.155)	(.158)
	-.033	-.080	-.111
60th quantile	(.132)	(.129)	(.112)
	-.100	-.100	-.080
80th quantile	(.146)	(.107)	(.108)
Diagnoses		x	x
5 health groups		x	x
Extra health controls			x
N	931	931	931

Table 9: Quantile regressions: other quantiles Standard errors in brackets.
*10 % significance level, **5 % significance level, ***1% significance level.

Appendix

A) EQ-5D, with predefined answers/levels

Mobility

(1) I have no problems in walking about

(2) I have some problems in walking about

(3) I am confined to bed

Self-Care

(1) I have no problems with self-care

(2) I have some problems washing or dressing my self

(3) I am unable to wash or dress my self

Usual activities (e.g. work, study, housework, family or leisure activities)

(1) I have no problems with performing my usual activities

(2) I have some problems with performing my usual activities

(3) I am unable to perform my usual activities

Anxiety/Depression

(1) I am not anxious or depressed

(2) I am moderately anxious or depressed

(3) I am extremely anxious or depressed

Department of Economics
University of Bergen
Fosswinckels gate 14
N-5007 Bergen, Norway
Phone: +47 55 58 92 00
Telefax: +47 55 58 92 10
http://www.svf.uib.no/econ