

# Can Reporting Heterogeneity Explain Differences in Depressive Symptoms Across Europe?

Renske Kok · Mauricio Avendano · Teresa Bago d’Uva ·  
Johan Mackenbach

Accepted: 26 May 2011 / Published online: 15 June 2011  
© The Author(s) 2011. This article is published with open access at Springerlink.com

**Abstract** Depression is one of the leading causes of disability in the developed world. Previous studies have shown varying depression prevalence rates between European countries, and also within countries, between socioeconomic groups. However, it is unclear whether these differences reflect true variations in prevalence or whether they are attributable to systematic differences in reporting styles (reporting heterogeneity) between countries and socioeconomic groups. In this study, we examine the prevalence of three depressive symptoms (mood, sleeping and concentration problems) and their association with educational level in 10 European countries, and examine whether these differences can be explained by differences in reporting styles. We use data from the first and second waves of the COMPARE study, comprising a sub-sample of 9,409 adults aged 50 and over in 10 European countries covered by the Survey of Health, Ageing and Retirement in Europe. We first use ordered probit models to estimate differences in the prevalence of self-reported depressive symptoms by country and education. We then use hierarchical ordered probit models to assess differences controlling for reporting heterogeneity. We find that depressive symptoms are most prevalent in Mediterranean and Eastern European countries, whereas Sweden and Denmark have the lowest prevalence. Lower educational level is associated with higher prevalence of depressive symptoms in all European regions, but this association is weaker in Northern European countries, and strong in Eastern European countries. Reporting heterogeneity does not explain these cross-national differences. Likewise, differences in depressive symptoms by educational level remain and in

---

R. Kok (✉) · M. Avendano · J. Mackenbach  
Department of Public Health, Erasmus University Medical Center Rotterdam, Rotterdam,  
The Netherlands  
e-mail: renskekok@gmail.com

M. Avendano  
Harvard Center for Population and Development Studies, Harvard University, Cambridge, MA, USA

T. Bago d’Uva  
Erasmus School of Economics, Erasmus University of Rotterdam, Rotterdam, The Netherlands

T. Bago d’Uva  
Tinbergen Institute, Rotterdam, The Netherlands

some regions increase after controlling for reporting heterogeneity. Our findings suggest that variations in depressive symptoms in Europe are not attributable to differences in reporting styles, but are instead likely to result from variations in the causes of depressive symptoms between countries and educational groups.

**Keywords** Socioeconomic · Depression · Reporting style · International · Vignettes

## 1 Introduction

According to projections by the World Health Organization (WHO) depression will be the leading cause of Disability Adjusted Life Years (DALYs) lost in high income countries in 2030 (Mathers and Loncar 2006). Depression influences the risk of other chronic conditions such as heart failure, heart disease and stroke (Blazer 2003; Larson et al. 2001; Liebetrau et al. 2008; Arbelaez et al. 2007; Barth et al. 2004). Previous studies indicate that depression prevalence varies between countries (Castro-Costa et al. 2007), and within countries (Lorant et al. 2003). Disentangling true variations in depression prevalence from variations attributable to differences in reporting is an essential part of understanding the causes of these differences.

Direct international comparisons of depression prevalence between European countries are scarce but the existing evidence suggests large differences across Europe. Castro-Costa et al. (2007) used a representative sample of the European population aged 50 or above, taken from the Survey of Health, Ageing and Retirement in Europe (SHARE), to evaluate the prevalence of depressive symptoms in Europe. Using of the EURO-D depression scale, which was specifically developed as a standardized measure of depression across European countries (Prince et al. 1999b), their study revealed variation in the prevalence of depressive symptoms across Western European countries. For all symptoms, the prevalence was higher in Mediterranean countries (France, Italy and Spain), lower in Northern European countries (Sweden and Denmark), and of average level in Western European countries (Castro-Costa et al. 2007). Copeland et al. (1999) compared random community samples of older people (65+) using the Geriatric Mental State–AGECAT (GMS-AGECAT) package in eight European cities. Adjusting for gender, they classified centers into two groups: the high depression prevalence group (17.3–23.6%), which comprised London, Berlin, Verona and Munich; and the low prevalence group (8.8–12.0%), which comprised Iceland, Liverpool, Zaragoza, Dublin and Amsterdam. (Copeland et al. 1999, 2004; Castro-Costa et al. 2007). Prince et al., using the same dataset and the adjusted EURO-D scale as outcome, found the lowest prevalence of depressive symptoms in the UK and Ireland, followed by Mediterranean countries (France, Spain, Italy), Benelux countries, Nordic countries and finally Germany (Prince et al. 1999a). Zunzunegui et al. performed a cross-national analysis of depressive symptoms among older adults between 75 and 84 years old, and found the highest prevalence in Italy, followed by Israel, Sweden, Leganes (Spain), and the Netherlands (Zunzunegui et al. 2007).

These cross-national studies do not show a clear geographical pattern of differences in prevalence of depression across Europe. Although Castro-Costa et al. and Prince et al. found a North–South divide, other studies show different patterns. A limitation of these studies is their focus on Western European countries only, while little is known about depressive symptoms in Eastern and Central Europe. In addition, studies were based on samples that were neither comparable nor nationally representative, and only a few studies relied on a standardized measure of depression across countries.

Studies on the association between depressive symptoms and socioeconomic status (SES) have shown relatively consistent results, suggesting that lower socioeconomic status is associated with higher depression rates. The results of a meta-analysis of socioeconomic inequalities in depression covering 51 prevalence studies, five incidence studies, and four persistence studies indicate that low-SES individuals have higher odds of: (1) being depressed (odds ratio = 1.81,  $p < 0.001$ ), (2); developing a new depressive episode (odds ratio = 1.24,  $p < 0.004$ ); and (3) suffering from persistent depression (odds ratio = 2.06,  $p < 0.001$ ) (Lorant et al. 2003). Miech and Shanahan (2000) analyzed the association between socioeconomic status and depression over the life-course, using a nationally representative sample of 2,031 adults between the ages of 18 and 90 in the US. Their study suggests that the strength of the association between depression and educational level increases with age. In addition, higher prevalence of physical health problems among lower-educated adults accounted for most of the diverging gap in depression (Miech and Shanahan 2000). Freyers et al. reviewed major European population studies published from 1980 to 2005 on the distribution of common mental disorders, and found a higher prevalence of anxiety and depression in social disadvantaged groups (Fryers et al. 2005). Overall, although these studies have been conducted in different setting and applied different methodologies, they consistently show an inverse association between socioeconomic status and the prevalence of depression.

Typically, population wide studies on depression are based on self-reports of symptoms, rather than on clinical diagnosis. It is therefore unknown to what extent observed differences reflect true variations in depressive symptoms. Cultural, socioeconomic and demographic factors may influence the reference scale that respondents use when rating the presence and severity of their own depressive symptoms (Bago d'Uva et al. 2008b). Differences in response scales, also known as reporting heterogeneity (King et al. 2004), can result in variations in ratings that are not attributable to true differences in depressive symptoms. If reporting heterogeneity is systematic across countries or socioeconomic groups, estimated differences in prevalence across these groups based on self-reports may be biased.

Over recent years, the anchoring vignette approach has been developed to quantify and correct for reporting heterogeneity in subjective categorical self-assessments. Applied to the domain of depression, anchoring vignettes are concrete descriptions of depressive symptoms of hypothetical individuals, which participants are asked to rate on the same scale that they use to rate their own level of depressive symptoms. As vignettes describe fixed levels of depressive symptoms, ratings of respondents are in theory be attributable to heterogeneity in the scale respondents use to assess depressive symptoms. Based on this information, the Hierarchical Ordered Probit (HOPIT) model estimates the magnitude of reporting heterogeneity and uses this to identify differences in depressive symptoms that are not attributable to heterogeneity.

The HOPIT model has recently been used to examine cross-national differences in several outcomes other than depression. Kapteyn et al. (2007) used this approach to analyze work disability differences between the Netherlands and the United States. Their results showed that Dutch respondents are more likely to report work disability than their US counterparts, but about half of this difference is explained by reporting heterogeneity. More detailed descriptions and applications of the HOPIT model are available elsewhere (King et al. 2004; Salomon et al. 2004; Bago d'Uva et al. 2008b; Kristensen and Johansson 2008).

In this paper, we compare prevalence rates of mood, sleeping and concentration problems, which are symptoms included in the EURO-Depression scale. We hypothesize that cross-national and socioeconomic differences in depressive symptoms are attributable

**Table 1** Descriptive variables of covariates by country

	N	Male	Female	Age		Years of schooling	
				Mean	SD	Mean	SD
Denmark	992	447	545	64	9.5	13	3.5
Germany	1,407	650	757	64	9.1	13	3.2
Netherlands	707	342	365	62	9.7	12	3.8
Belgium	1,037	483	554	65	9.7	12	3.7
France	936	404	532	65	10.2	11	5.0
Italy	844	388	456	64	8.9	8	4.4
Spain	723	323	400	64	10.3	8	4.8
Greece	655	328	327	62	9.8	10	4.4
Poland	552	244	308	62	9.6	9	3.0
Czech Republic	896	367	528	64	9.6	11	3.1
Total	9,409	4,282	5,127	64	9.7	11	4.3

to systematic differences in reporting styles. We first examine cross-national differences in the prevalence of self-reported depressive symptoms by country and educational level. In a second step, we use the HOPIT model to assess the extent to which cross-national and socioeconomic variations in depressive symptoms are attributable to reporting heterogeneity.

## 2 Methodology

### 2.1 Data

This study is based on data from The Survey of Health, Ageing and Retirement in Europe (SHARE), a longitudinal investigation of the health, social networks, economic situation, and well-being of Europeans aged 50 years and over (Börsch-Supan and Jürges 2005). A drop-off questionnaire of SHARE contains self-assessments and vignette evaluations in several health and work disability dimensions and four other domains of satisfaction and well-being, and was applied to a sub-sample of the main SHARE sample. We refer to these data as the COMPARE sample ([www.compare-project.org](http://www.compare-project.org)). A description of the development of the vignettes is given elsewhere (Van Soest 2008). We have used the two waves available in the COMPARE sample, 2004 (4,544 participants) and 2006/2007 (7,186 participants). Three vignettes per item were available in the first wave. In the second wave this was cut back to one vignette per item in order to shorten the questionnaire. The description of these repeated vignettes was equal in both waves. All vignettes are presented in “Appendix”. For respondents who participated in both waves (1991), we only use data from the second wave. Our findings are robust to using data from the first wave for respondents who participated in both waves. We dropped 330 individuals with missing values on at least one variable, resulting in a final sample of 9,409 adults from 11 European countries.<sup>1</sup> Table 1 summarizes the distribution of basic covariates of the sample.

<sup>1</sup> Note that not all SHARE countries are included in the COMPARE sample. The COMPARE questionnaire was not issued in: Austria, Czech Republic, Israel, Sweden, Switzerland and Poland.

### 2.1.1 Depression Measure

In this analysis, we used three items of the EURO-D (1999b), a standardized scale of depressive symptoms designed to enhance cross-national comparability. The EURO-D consists of 12 items: mood, pessimism, death wish, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness. The COMPARE dataset includes self-reports and anchoring vignettes in the dimensions of depression, sleep and concentration. Self-reports and anchoring vignettes are rated on the scale: 1—no problems; 2—mild problems; 3—moderate problems; 4—severe problems; 5—extreme problems. Since only a small number of respondents reported more than moderate problems (i.e. categories 4 and 5) on each of the depressive items, we collapsed the scale into: 1—no problems, 2—mild problems, and 3—at least moderate problems.<sup>2</sup>

### 2.1.2 Education

Two main education items were included in SHARE wave 2, namely the highest completed educational level and the number of years of schooling. To enhance comparability of national educational levels, we transformed national levels into the 1997 International Standard Classification of Education (ISCED-97) of the United Nations Educational, Scientific, and Cultural Organization (UNESCO). The distribution of education varies greatly across European countries. For example, if we consider the following categorization of ISCED 0–2, ISCED 3 and ISCED 4–6, the group with the lowest educational category comprises 17.01% in Germany, but 80.15% in Spain. We therefore operationalize education based on the number of years of schooling. This variable varies between 0 and 25, with an average of 11 years. Because years of schooling are entered as a continuous variable in all models, we used smooth non-parametric LOESS function curves (Cleveland 1979) to examine deviations from linearity. Results indicate that associations with depressive symptoms are approximately linear, which justifies the use of a non-quadratic form of years of schooling in our models.

## 2.2 Analysis

For each item, we estimate both ordered probit and HOPIT models to analyze cross-national and socioeconomic differences in depressive symptoms. This section gives a short description of both models.

### 2.2.1 The Basic Model: Ordered Probit

The self-reported ratings of depressive symptoms are ordinal categorical variables. The ordered probit model is the standard model for estimation of the probabilities of ordinal responses. This model assumes an underlying true latent variable  $Y_{ij}^*$  with  $i$  indicating the individual and  $j$  the depressive symptom. The latent variable is constructed as follows:

---

<sup>2</sup> Tentative estimation of the HOPIT model with five categories is not possible, because the thresholds between the worse categories are not well identified due to the small number of observations in the top categories.

$$Y_{ij}^* = \beta X_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \tag{1}$$

where  $X_i$  is a vector of covariates and  $\beta$  is a vector of coefficients to be estimated.<sup>3</sup> The latent variable  $Y_{ij}^*$  corresponds to the observed ordinal ratings  $y_{ij}$  in the following way:

$$\begin{aligned} y_{ij} &= 1 & \text{if } Y_{ij}^* < \tau^1 \\ y_{ij} &= 2 & \text{if } \tau^1 \leq Y_{ij}^* < \tau^2 \\ y_{ij} &= 3 & \text{if } Y_{ij}^* \geq \tau^2 \end{aligned} \tag{2}$$

where  $\tau^1 < \tau^2$  are the cut-points on the latent scale between the response categories. These thresholds are equal for all respondents, which corresponds to assuming no reporting heterogeneity. However, if reporting heterogeneity is present,  $\beta$  reflects a mixture of the true association between the covariates and  $y_{ij}$ , and systematic differences in reporting  $y_{ij}$  associated with the covariates.

### 2.2.2 The Extended Model: HOPIT

The Hierarchical Ordered Probit (HOPIT) model is an extension of the ordered probit model that allows for systematic variation in the cut-points, and thus incorporates adjustment for reporting heterogeneity.<sup>4</sup> In contrast with the ordered probit model, the HOPIT model estimates the cut-points between categories using the ratings of the anchoring vignettes given by respondents. As the level of depressive symptoms described in the vignettes is fixed across respondents, the variation in ratings can be attributed to reporting heterogeneity.

The HOPIT model consists of two main components: the first estimates the cut-points on the latent scale of depressive symptoms, and the second estimates the corrected respondent level of depressive symptoms on the same latent scale. The first component considers an unobserved latent variable  $Y_{ij}^{v*}$  that represents the underlying unobserved level of depressive symptom  $j$  represented by the vignettes. Formally, the latent variable  $Y_{ij}^{v*}$  is described by:

$$Y_{ij}^{v*} = \alpha + \varepsilon_{ij}^v, \quad \varepsilon_{ij}^v \sim N(0, 1) \tag{3}$$

where  $\alpha$  is a vector of dummy variables identifying the respective vignette. The observed vignette ratings  $y_{ij}^v$  results from the mapping of  $Y_{ij}^{v*}$  into three categories, using person-specific cut-points ( $k = 1, 2$ ):

$$\begin{aligned} y_{ij}^v &= 1 & \text{if } Y_{ij}^{v*} < \tau_{ij}^1 \\ y_{ij}^v &= 2 & \text{if } \tau_{ij}^1 \leq Y_{ij}^{v*} < \tau_{ij}^2 \\ y_{ij}^v &= 3 & \text{if } Y_{ij}^{v*} \geq \tau_{ij}^2 \end{aligned} \tag{4}$$

In order to allow for systematic variation in the cut-points, these are modeled as functions of covariates in the following way:

$$\tau_{ij}^k = \gamma^k X_i \tag{5}$$

<sup>3</sup> Estimation of the ordered probit model requires normalization of location and scale, usually done by including no constant term in  $x$  and setting  $\sigma^2 = 1$ .

<sup>4</sup> See Tandon et al. (2002) for a more detailed description of the HOPIT model.

where  $\gamma^k$  is a vector of coefficients to be estimated. Note that the covariates are solely included in the estimation of the cut-points (i.e., in Eq. 5 but not in Eq. 3), which corresponds to assuming that there is no systematic variation in the perceived level of depressive symptoms represented by the vignettes (assumption of vignette equivalence, King et al. 2004).

The second component of the HOPIT models the individual’s own level of depressive symptoms imposing the cut-points as given by the first component. As in the ordered probit model, the level of depressive symptoms  $Y_{ij}^{s*}$  and the observed categorical responses  $y_{ij}^s$  are defined as:

$$Y_{ij}^{s*} = \beta Z_i + \varepsilon_i^s, \quad \varepsilon_i^s \sim N(0, \sigma^2) \tag{6}$$

$$\begin{aligned} y_{ij}^s = 1 & \quad \text{if } Y_{ij}^{s*} < \tau_{ij}^1 \\ y_{ij}^s = 2 & \quad \text{if } \tau_{ij}^1 \leq Y_{ij}^{s*} < \tau_{ij}^2 \\ y_{ij}^s = 3 & \quad \text{if } Y_{ij}^{s*} \geq \tau_{ij}^2 \end{aligned} \tag{7}$$

where  $Z_i$  and  $\beta$  are defined as in 2.2.1 but  $\tau_{ij}^1$  and  $\tau_{ij}^2$  are allowed to be functions of covariates, defined by Eq. 5. This corresponds to assuming that individuals use the same response scale when rating the vignettes and their own situation (response consistency, King et al. 2004).

Using the HOPIT model, we compute the prevalence of depressive symptoms that would be observed under a counterfactual: that all respondents used an identical response scale. In the computation of the thresholds, all covariates in Eq. 5 will be set to counterfactual reference characteristics.

### 2.3 The Analysis

Our analysis aims to assessing the influence of reporting heterogeneity on (a) cross-national differences and (b) socioeconomic differences in the prevalence of depressive symptoms. Models for cross-national differences include sex, age and country dummies, taking Sweden as reference country. We estimate both a basic and the extended model and calculate predicted probabilities of depressive symptoms for 64-year-old males for each country using both models. Because the sample size is relatively small in each country, male and female samples are pooled together to sustain power in the estimations. We chose 64 as the reference age, because this is the mean age in the sample. For our second objective, we examine educational disparities in depressive symptoms in Europe by adding years of schooling to the model. We cannot assume similar education effects across Europe and so we also allow for variation in these through inclusion of interaction terms between European area indicators and years of schooling. In order to increase statistical power, we consider, instead of country-specific education effects, interactions with indicators of the following areas: Nordic countries (i.e. Sweden and Denmark); Central-West countries (i.e. Germany, The Netherlands and Belgium); Mediterranean countries (i.e. France, Italy and Spain); and finally Eastern countries (i.e. Czech Republic and Poland).<sup>5</sup>

<sup>5</sup> It was especially difficult to obtain precise estimates of country-specific education effects for Poland and Denmark, as these countries were solely included in Wave 2, where only one vignette per item is available.

### 3 Results

#### 3.1 Cross-National Differences in Depressive Symptoms

Table 2 shows the estimated coefficients of thresholds between none and mild depressive problems in the HOPIT models for each item. Because the thresholds are measured on a latent scale, the coefficients have no quantitative interpretation. However, their signs and the  $p$  values indicate the direction and the significance of effects of the variables, i.e., a negative value indicates a lower reporting threshold with respect to the reference category. Results show that females have lower thresholds for reporting mood and sleep symptoms, indicating that they report symptoms more easily than men. Increasing age is associated with higher reporting thresholds to report mood and sleep symptoms, but lower thresholds to report concentration problems.

Table 2 shows that there are large differences between countries. Compared to Sweden, respondents from the Netherlands, Belgium and Czech Republic use similar reference scales for mood and sleep problems, but they have higher thresholds when reporting concentration problems. When reporting sleeping problems, Danish respondents use a scale similar to the Swedish, but are more lenient when reporting other symptoms. Respondents from Greece, France, Italy and Spain have higher thresholds than the Swedish for reporting all symptoms. Polish respondents have higher thresholds for reporting mood and sleep symptoms, but report concentration symptoms more easily than Swedish respondents. Reporting thresholds for mood symptoms do not differ significantly between Swedish and German respondents, but the latter have higher thresholds for reporting sleeping and concentration problems.

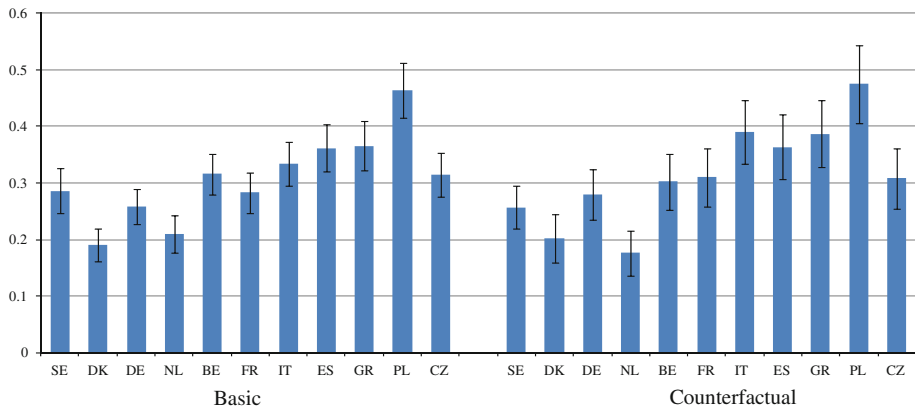
Figure 1 presents the predicted probability of mood problems for a 64-year-old male in each country. The left-hand panel shows the results using the Ordered Probit model, which assumes fixed thresholds for all respondents. The right-panel presents the predicted probabilities obtained from the HOPIT model, using the response scale of a 64-year-old

**Table 2** Estimation of the thresholds of depressive symptoms with the HOPIT model

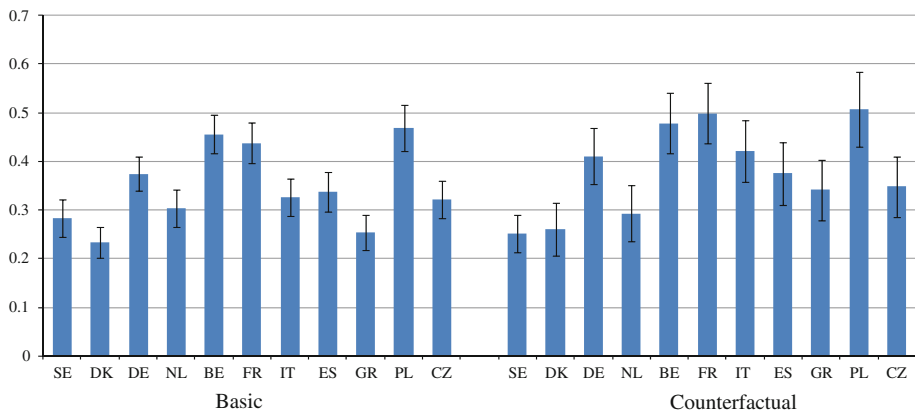
	Mood		Sleep		Concentration	
	Coefficient	$p$ value	Coefficient	$p$ value	Coefficient	$p$ value
Constant	<b>1.134</b>	<b>0.000</b>	<b>1.125</b>	<b>0.000</b>	<b>1.757</b>	<b>0.000</b>
Female	<b>-0.068</b>	<b>0.001</b>	-0.044	0.090	0.032	0.141
Age	<b>0.004</b>	<b>0.001</b>	<b>0.003</b>	<b>0.011</b>	<b>-0.004</b>	<b>0.001</b>
Denmark	<b>0.121</b>	<b>0.024</b>	0.105	0.116	<b>0.245</b>	<b>0.000</b>
Germany	0.085	0.080	<b>0.133</b>	<b>0.024</b>	<b>0.178</b>	<b>0.000</b>
Netherlands	-0.098	0.077	-0.013	0.847	<b>0.113</b>	<b>0.030</b>
Belgium	-0.056	0.276	0.065	0.293	<b>0.133</b>	<b>0.008</b>
France	<b>0.158</b>	<b>0.002</b>	<b>0.225</b>	<b>0.000</b>	<b>0.226</b>	<b>0.000</b>
Italy	<b>0.186</b>	<b>0.000</b>	<b>0.222</b>	<b>0.001</b>	<b>0.232</b>	<b>0.000</b>
Spain	<b>0.191</b>	<b>0.000</b>	<b>0.155</b>	<b>0.019</b>	<b>0.159</b>	<b>0.002</b>
Greece	<b>0.206</b>	<b>0.000</b>	<b>0.375</b>	<b>0.000</b>	<b>0.465</b>	<b>0.000</b>
Poland	<b>0.165</b>	<b>0.009</b>	<b>0.216</b>	<b>0.005</b>	<b>-0.500</b>	<b>0.000</b>
Czech Republic	-0.088	0.125	-0.019	0.786	<b>0.131</b>	<b>0.048</b>

Bold indicates significance at 5% level. Reference country is Sweden



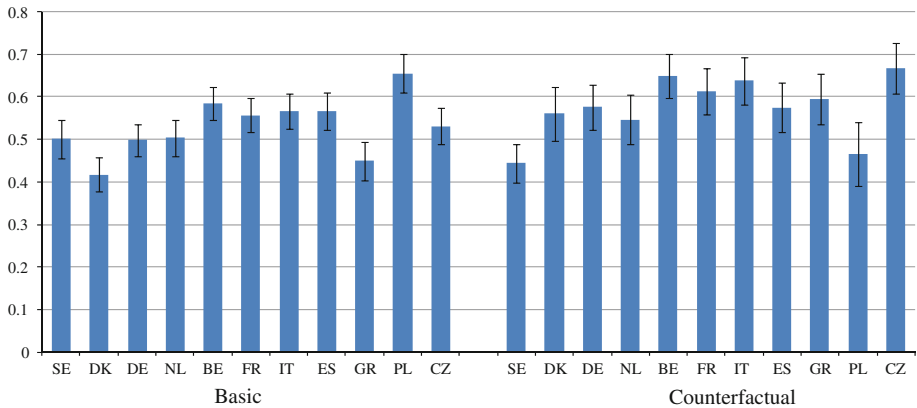


**Fig. 1** Predicted probability of having mood problems for a 64-year-old male



**Fig. 2** Predicted probability of having sleeping problems for a 64-year-old male

Swedish male as reference. Comparison of the outcomes of the basic with the counterfactual situation shows that countries have similar rankings with regards to the probabilities of mood problems. In other words, the cross-national differences in probabilities remain after the Swedish response scale has been applied to all countries. Figures 2 and 3 show the same results for sleeping and concentration problems. Again, the cross-national differences are not explained by differences in reporting scales. Cross-national differences remain, and the ranking of the countries are similar in both the HOPIT and Probit models. For sleeping problems, the contrast between Italy, France, Greece and Poland vis-à-vis the Nordic countries is somewhat more marked after controlling for reporting heterogeneity. This reflects the fact that respondents from these countries have higher thresholds for reporting sleeping problems than respondents in Nordic European countries, as shown in Table 2. For concentration problems, the probability of symptoms increases in all countries when applying the response scale of a 64-year-old Swedish male, except in Poland, for which the prevalence decreases. Based on the HOPIT model, respondents from Czech Republic have the highest probability of concentration problems.



**Fig. 3** Predicted probability of having concentration problems for a 64-year-old male

Table 3 shows the rate ratios comparing the probability of depressive symptoms in Sweden to the probability in other European countries. The rate ratios are computed using probabilities of 64-year-old males. A rate ratio below/above 1 indicates a lower/higher probability of depressive symptoms compared to Sweden. For Central-West European countries (Germany, the Netherlands, Belgium) and Denmark, rate ratios (RRs) of mood problems are similar for both the Probit and HOPIT models. For Mediterranean European countries (France, Italy, Greece and Spain) and Poland, rate ratios from the HOPIT model are somewhat larger than ratios from the Probit model. Overall, cross-national differences in the prevalence of mood problems remain and in some instances increase somewhat after controlling for reporting heterogeneity.

For sleeping, almost all rate ratios increase after adjustment for reporting heterogeneity, except for The Netherlands. RRs for Denmark and Greece are significantly lower than 1 in the basic model, while in the counterfactual situation the RR is not significant for Denmark, but for Greece it is significantly larger than 1. Rate ratios for concentration problems show a similar pattern to those of sleeping problems, with adjustment for reporting heterogeneity increasing significantly almost all RRs. In addition, the RRs of Germany, the Netherlands and Czech Republic are insignificant in the Probit model but significant in the HOPIT model. These results suggest that differences in sleeping and concentration problems are not explained by reporting heterogeneity.

### 3.2 Differences in Mental Health Problems by Education

In order to assess whether systematic reporting heterogeneity by level of education is present, we now specify a separate model that includes years of schooling. Table 4 shows the estimated coefficients (and their  $p$  values) in thresholds between having no problems and mild problems, for each of the three depressive symptoms. In the case of mood problems, number of years of schooling has a negative effect (significant at a 5% level) in the Central-West and Mediterranean European countries, indicating that higher educated respondents report problems more easily. Educational differences in reporting of sleeping problems are only significant (at a 5% level) for Central-Western countries, also with a negative effect. In the estimation of concentration problems, the estimated effect of years of schooling is positive (significant at 5%) in Nordic countries, suggesting that the higher the level of

**Table 3** Rate ratios comparing probabilities of depressive symptoms in European countries to those in Sweden

	Probit model		HOPIT model	
	RR	0.95CI	RR	0.95CI
<i>Mood</i>				
Sweden	1		1	
Denmark	<b>0.66</b>	<b>(0.56–0.77)</b>	<b>0.79</b>	<b>(0.62–0.95)</b>
Germany	0.90	(0.79–1.02)	1.09	(0.90–1.27)
Netherlands	<b>0.74</b>	<b>(0.62–0.85)</b>	<b>0.68</b>	<b>(0.53–0.84)</b>
Belgium	1.10	(0.96–1.25)	1.18	(0.97–1.38)
France	0.99	(0.85–1.13)	<b>1.21</b>	<b>(1.00–1.41)</b>
Italy	<b>1.17</b>	<b>(1.01–1.33)</b>	<b>1.52</b>	<b>(1.27–1.77)</b>
Spain	<b>1.26</b>	<b>(1.09–1.44)</b>	<b>1.41</b>	<b>(1.17–1.66)</b>
Greece	<b>1.28</b>	<b>(1.10–1.46)</b>	<b>1.51</b>	<b>(1.25–1.76)</b>
Poland	<b>1.62</b>	<b>(1.40–1.84)</b>	<b>1.85</b>	<b>(1.53–2.16)</b>
Czech Republic	1.10	(0.95–1.25)	1.20	(0.98–1.42)
<i>Sleep</i>				
Sweden	1		1	
Denmark	<b>0.82</b>	<b>(0.70–0.94)</b>	1.04	(0.82–1.26)
Germany	<b>1.32</b>	<b>(1.16–1.48)</b>	<b>1.63</b>	<b>(1.37–1.90)</b>
Netherlands	1.07	(0.92–1.22)	1.16	(0.92–1.40)
Belgium	<b>1.61</b>	<b>(1.41–1.81)</b>	<b>1.90</b>	<b>(1.60–2.21)</b>
France	<b>1.54</b>	<b>(1.35–1.74)</b>	<b>1.98</b>	<b>(1.67–2.30)</b>
Italy	<b>1.15</b>	<b>(1.00–1.31)</b>	<b>1.67</b>	<b>(1.38–1.96)</b>
Spain	<b>1.19</b>	<b>(1.02–1.35)</b>	<b>1.49</b>	<b>(1.21–1.77)</b>
Greece	0.89	(0.75–1.03)	<b>1.36</b>	<b>(1.09–1.62)</b>
Poland	<b>1.65</b>	<b>(1.43–1.88)</b>	<b>2.01</b>	<b>(1.65–2.38)</b>
Czech Republic	1.13	(0.98–1.28)	<b>1.38</b>	<b>(1.11–1.65)</b>
<i>Concentration</i>				
Sweden	1		1	
Denmark	<b>0.83</b>	<b>(0.75–0.91)</b>	<b>1.26</b>	<b>(1.11–1.42)</b>
Germany	1.00	(0.91–1.08)	<b>1.30</b>	<b>(1.16–1.44)</b>
Netherlands	1.00	(0.91–1.10)	<b>1.23</b>	<b>(1.08–1.38)</b>
Belgium	<b>1.17</b>	<b>(1.07–1.26)</b>	<b>1.46</b>	<b>(1.31–1.62)</b>
France	<b>1.11</b>	<b>(1.02–1.21)</b>	<b>1.38</b>	<b>(1.23–1.53)</b>
Italy	<b>1.13</b>	<b>(1.03–1.23)</b>	<b>1.44</b>	<b>(1.28–1.59)</b>
Spain	<b>1.13</b>	<b>(1.03–1.23)</b>	<b>1.30</b>	<b>(1.14–1.45)</b>
Greece	<b>0.90</b>	<b>(0.80–0.99)</b>	<b>1.34</b>	<b>(1.19–1.50)</b>
Poland	<b>1.31</b>	<b>(1.19–1.43)</b>	1.05	(0.87–1.22)
Czech Republic	1.06	(0.97–1.15)	<b>1.50</b>	<b>(1.34–1.67)</b>

Bold indicates significance at 5% level. Reference country is Sweden

education, the more lenient respondents are to report problems. Systematic differences in reporting depressive symptoms by age, gender and country remained similar to our earlier findings, indicating that these differences are robust to educational differences in reporting.

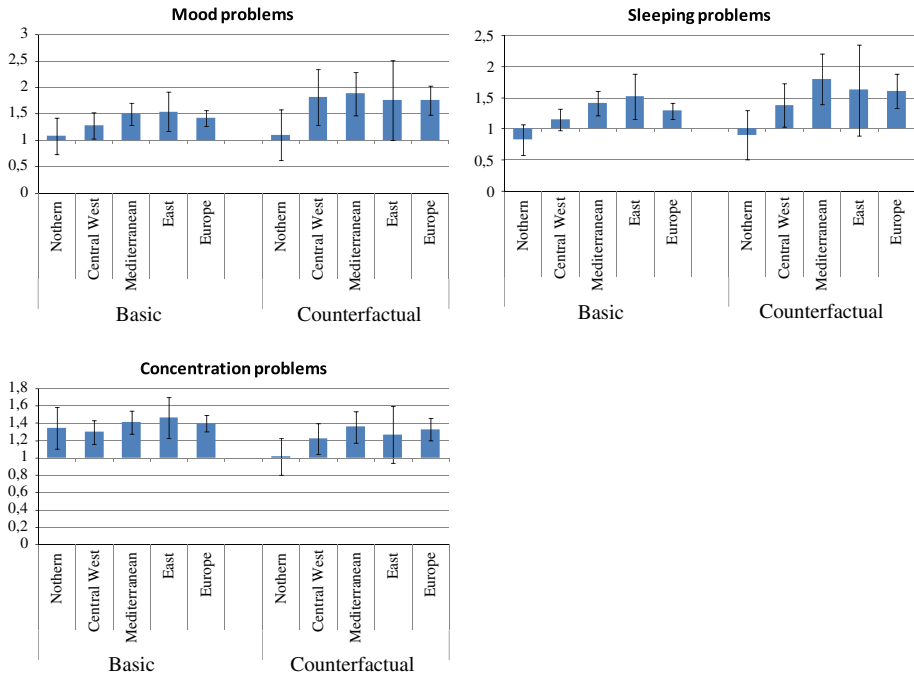
**Table 4** Estimation of the thresholds of depressive symptoms with the HOPIT model including years of schooling

	Mood		Sleeping		Concentration	
	Coefficient	<i>p</i> value	Coefficient	<i>p</i> value	Coefficient	<i>p</i> value
Constant	<b>0.893</b>	<b>0.000</b>	<b>1.102</b>	<b>0.000</b>	<b>1.172</b>	<b>0.000</b>
Female	<b>-0.084</b>	<b>0.000</b>	<b>-0.056</b>	<b>0.034</b>	0.036	0.100
Age	0.002	0.054	0.002	0.145	<b>-0.003</b>	<b>0.007</b>
Denmark	<b>0.120</b>	<b>0.033</b>	0.109	0.113	<b>0.202</b>	<b>0.002</b>
Germany	<b>0.350</b>	<b>0.004</b>	0.224	0.128	<b>0.382</b>	<b>0.003</b>
The Netherlands	0.136	0.257	0.061	0.673	<b>0.326</b>	<b>0.010</b>
Belgium	0.177	0.131	0.140	0.326	<b>0.346</b>	<b>0.006</b>
France	<b>0.299</b>	<b>0.004</b>	<b>0.327</b>	<b>0.012</b>	<b>0.511</b>	<b>0.000</b>
Italy	<b>0.290</b>	<b>0.005</b>	<b>0.285</b>	<b>0.025</b>	<b>0.515</b>	<b>0.000</b>
Spain	<b>0.296</b>	<b>0.004</b>	0.217	0.091	<b>0.443</b>	<b>0.000</b>
Greece	<b>0.329</b>	<b>0.002</b>	<b>0.458</b>	<b>0.000</b>	<b>0.751</b>	<b>0.000</b>
Poland	<b>0.283</b>	<b>0.045</b>	0.181	0.291	<b>-0.358</b>	<b>0.042</b>
Czech Republic	0.057	0.713	-0.047	0.802	0.248	0.193
Years edu Nordic	0.001	0.938	-0.004	0.648	<b>0.025</b>	<b>0.003</b>
Years edu Central-west	<b>-0.019</b>	<b>0.001</b>	-0.011	0.101	0.006	0.270
Years edu Mediterranean	<b>-0.012</b>	<b>0.001</b>	<b>-0.014</b>	<b>0.002</b>	0.000	0.995
Years edu East	-0.012	0.233	-0.002	0.867	0.014	0.285

Bold indicates significance at 5% level. Reference country is Sweden

Rate ratios in Fig. 4 compare the probability of having problems for a 64-year-old man with 3 years of schooling with the probability for a 64-year-old man with 17 years of schooling, i.e. for the bottom 5% of the educational distribution and the top 5%. A rate ratio above 1 indicates a higher probability of problems for respondents with 3 years of schooling than for respondents with 17 years of schooling. Again, we included interaction terms of years of schooling by European area. However, the last bar in the figures corresponds to the educational effect of the pooled European sample, excluding the interaction terms. The rate ratios incorporating reporting heterogeneity are estimated for the counterfactual situation that all respondents use the response scale of a 64-year-old male with 3 years of schooling in their respective country.

Correcting for reporting heterogeneity does not attenuate rate ratios comparing low and high education (Fig. 4). This suggests that different reporting scales do not explain educational differences in depressive symptoms. For all countries pooled, counterfactual rate ratios for mood and sleeping problems are greater than rate ratios from the Probit model. This is expected, since the estimated coefficients of years of schooling on the thresholds are mostly negative, which indicates that more highly educated people are more likely to report problems. Therefore, without including a correction for reporting heterogeneity in the analysis, the educational differences are somewhat underestimated. Although the confidence intervals in the separate areas are large, we see the impact of correcting for reporting heterogeneity is most profound in Central-Western and Mediterranean countries. For concentration problems, the rate ratio for the pooled European sample in the counterfactual analysis is somewhat lower than in the Probit model. However, educational differences in concentration problems are still present in the counterfactual analysis,



**Fig. 4** Rate ratios comparing low and high education

indicating that reporting heterogeneity does not completely explain these differences. The influence of differences in reporting by years of schooling is the largest in Central Western and Mediterranean countries, and as expected from the positive coefficient in Table 4, we find a reversed effect in Nordic countries after controlling for heterogeneity.

#### 4 Discussion

This study investigates the effect of reporting behavior on cross-national and socioeconomic differences in depression symptoms. Our results indicate that differences in the prevalence of depressive symptoms between countries are generally not explained by systematic cross-national differences in reporting thresholds. Similarly, within countries, differences in depressive symptoms by education are not explained by differences in reporting styles. Our findings suggest that variations in depressive symptoms in Europe are not attributable to differences in reporting styles, but might result from variations in the causes of depressive symptoms between countries and education groups.

Our study is the first to examine the role of reporting heterogeneity in explaining depression variations across and within European countries. Being the first some limitations should be considered. The HOPIT model is based upon two main assumptions. First, response consistency must be satisfied, which assumes that respondents use the same response scale for the vignettes as for their own self-assessment of symptoms. Van Soest et al. (2007) assessed this assumption by comparing a subjective measure of drinking problems with an objective measure. Albeit for a different outcome measure than the ones considered here, their results do support the assumption of response consistency. The use

of vignettes improved the fit of the model and raised the correlation between the subjective and the objective measures, which is in line with the main purpose of the vignette methodology (Van Soest et al. 2007). Bago d'Uva et al. (2010) also test for response consistency by comparing reporting heterogeneity inferred from conditioning on objective measures with that obtained from vignettes, using English data in the domain of concentration problems. Their results do not reject this assumption with respect to education (Bago d'Uva et al. 2010).

The second assumption in the HOPIT model implies that the level of depressive problems in vignettes is perceived in the same way by all respondents (vignette equivalence), irrespective of their age, sex, income, education, country of residence or other socio-demographic variables (Salomon et al. 2004). This assumption is difficult to test and has therefore been seldom tested. It is nevertheless supported by a high degree of consistency across individuals in the ranking of vignettes in six health domains (including concentration problems) and eight responsiveness domains, suggesting that they are understood similarly across age and education groups (Murray et al. 2003). On the other hand, more recent results of a formal, more demanding test applied to English data in the domain of concentration do not support this assumption with respect to education (Bago d'Uva et al. 2010). Further results of that paper, comparing vignette adjustment with adjustment by means of objective measures, suggest also that the vignette methodology may not be able to revise estimated disparities in concentration problems by education in the correct direction. A subset of similar objective measures is available in the dataset used here and has been exploited by Vonková and Hullege (2010). They assessed whether the vignette adjustment improves correlation between self-reports of concentration and objective measures. The results are mixed, showing weakening correlation when one of the vignettes is used (the one which was kept between waves one and two) but an improvement when either of the other two is used. Their analysis does not however permit assessment of the quality of the adjustment of education gradients and country rankings.

The subjective nature of the underlying constructs of the items makes the vignette equivalence a rather strong assumption. The subjective nature of the items not only makes it disputable that the interpretation of the level of problems described in the vignettes does not vary systematically across respondents but also raises the issue of whether there may be varying interpretations of the problem itself. It is unclear the extent to which these problems may affect the ability of vignettes to correct differential cut-point shift. In this light, our findings should be interpreted cautiously as a first approximation to the problem of reporting heterogeneity in depressive symptoms, rather than as a final proof of the absence of reporting heterogeneity between countries and educational groups.

On the other hand, although several depressive symptoms, including the items of mood and concentration, are indeed by definition subjective, some of the components of the Euro-Depression scale do have a clear objective counterpart. For these items vignette equivalence is probably less restrictive. Examples of this include: Troubles with sleep or recent changes in sleep patterns; diminution in appetite or desire for food; changes in eating patterns; fatigue; and events of tearfulness (crying). From these 'more objective' items, we had vignettes on sleep. The vignettes for this item asked respondents to make ratings of hypothetical individuals experiencing presumably objective behaviours such as: waking up two nights a week; taking 2 h every night to fall asleep; and waking up once every hour during the night and taking about 15 min to fall back asleep again. Differently from the mood and concentration items, these descriptions of sleep patterns are not by

definition subjective (they do not refer to individual's feelings), but instead refer to objective behaviour, which individuals are asked to judge using a given scale. The fact that for sleep patterns we also did not find that correcting for reporting heterogeneity diminished the cross-country differences strengthens our conclusion that at least part of the items to measure depressive symptoms are not strongly influenced by reporting heterogeneity. However, whether this finding holds for the more subjective components of the scale is indeed uncertain.

Previous studies have found that reporting heterogeneity explains cross-national differences in some health-related outcomes. Kapteyn et al. (2007) studied reporting differences in working disability between the US and The Netherlands. They found that more than half of the observed difference in reported work disability originates from the fact that residents of these two countries use different response scales in answering standard questions on whether they have a work disability. Essentially, for the same level of actual work disability, Dutch respondents have lower response thresholds in claiming disability than the American respondents (Kapteyn et al. 2007). Our study did not find strong support for the role of heterogeneity in explaining the large differences in depression observed across countries. This suggests that reporting heterogeneity across countries might play an important role for some physical health outcomes but not mental health outcomes.

Salomon et al. (2004), Bago d'Uva et al. (2008a, b) analyzed educational disparities in various health outcomes before and after correction for reporting heterogeneity. Outcomes of these studies are somewhat mixed. Bago d'Uva et al. (2008b) rejected reporting homogeneity by different educational groups in India, Indonesia and China, where correcting for reporting heterogeneity somewhat reduced disparities in health by education. However, using European data, it was found that higher educated older Europeans are more likely to rate a given health state negatively, resulting in underestimation or even undetected educational differences without correction (Bago d'Uva et al. 2008a). These results are similar to our own, where educational differences in health do not diminish but tend to increase after adjustment for reporting heterogeneity.

The COMPARE sample includes vignettes and self-reports on only three of the twelve items of the EURO-D scale. The lack of information on the other items does not allow us to draw conclusions concerning the influence of reporting heterogeneity on the EURO-D depression scale as a whole. In addition, the fact that differences in reporting heterogeneity differed somewhat across items and across countries makes it even harder to synthesize a single effect that can be generalized to the EURO-D measure. Thus, although we find no strong evidence of reporting heterogeneity for the three items assessed, it is still possible that heterogeneity exists in other items not assessed in our study. Whether this may lead to reporting differences in the overall Euro-D scale needs to be further examined.

The SHARE sample excludes the institutionalized population and therefore includes only adults living in the community. Observed cross-national variations in depression in our study may therefore be attributable to differential sample selection by country. In particular, institutionalization rates are generally higher in Northern and Central-Western European countries than in the Eastern and Southern parts of Europe. Depressive symptoms are likely to be associated with the risk of institutionalization, which may contribute to the low rates of depressive symptoms in Central-Western and Northern countries. To examine the extent of this bias, we did sensitivity analysis restricting the sample to ages 50–64, at which age institutionalization rates are relatively low. We found the same pattern

as for the entire sample, suggesting that institutionalization differences do not account for cross-national and educational variations in depressive symptoms.

When analysing educational differences across countries we compared the 5% highest and 5% lowest educational groups. We performed sensitivity analyses to assess whether our results are robust to this definition. We calculated the Relative Index of Inequality (RII) for categories of education based upon the International Standard Classification of Education (ISCED). On the basis of the RII we estimate rate ratios based on the total effect of the educational ranking on the three items for each European area as well as for all countries pooled together. These analyses confirm our original findings: we find differences in the prevalence of depressive symptom by education, which remain largely unchanged after controlling for reporting heterogeneity using the vignette approach.

Our results provide some support to the hypothesis that variations in depressive symptoms by country and education are the result of variations in the risk factors for depression, rather than an artifact of reporting heterogeneity. An important part of the differences may be explained by physical health problems. It is known that physical health is worse in lower educational groups (Kunst et al. 2005), which may underlie their depressive symptoms prevalence, as this was found in several studies (Geerlings et al. 2000; Lenze et al. 2001; Braam et al. 2005; Koster et al. 2006). In our study, we see that in countries where overall physical health is known to be worse, such as Spain and Italy (Jürges 2005), the prevalence of depressive symptoms is also higher. In addition, different effectiveness of depression treatment across European countries and socioeconomic levels may influence the duration of depression, leading to high point-prevalence of depression.

Economic factors may also be important in explaining differences in prevalence of depressive symptoms. Lorant et al. (2007) showed a lower material standard of living is associated with increased depressive symptoms and caseness of major depression. The higher prevalence of depression in Poland, Czech Republic and some of the Southern Mediterranean countries may be partly explained by higher rates of unemployment, less favorable employment conditions, and more economic hardship (Lyberaki and Tinios 2008; Siegrist and Wahrendorf 2008). Future studies should examine whether these economic and social factors explain the large variations in depressive symptoms observed in Europe.

In summary, using the vignette approach, we find no strong evidence that cross-national and educational differences in depressive symptoms across Europe are attributable to reporting heterogeneity. However, whether the HOPIT model assumptions hold on subjective outcomes such as depressive symptoms, or whether the cross-country differences in depression are attributable to variations in risk factors such as social and economic conditions, behaviour or physical health is a relevant question that deserves further investigation.

**Acknowledgments** Renske Kok was supported by a grant from the 6th Framework Programme from the European Commission (project SHARELIFE, CIT4-CT-2006-028812). And Mauricio Avendano was supported by a grant from the Netherlands Organisation for Scientific Research (NWO, grant no. 451-07-001), a Fellowship from the Erasmus University, and a David E. Bell fellowship from the Harvard Center for Population and Development studies. Teresa Bago d'Uva was funded by the NETSPAR theme "Income, health and care across the life cycle" and by a VENI grant from the Netherlands Organisation for Scientific Research (NWO).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.



## Appendix

### Self-Rated Items

#### *Mood*

Overall in the last 30 days, how much of a problem did you have with feeling sad, low, or depressed?

#### *Sleep*

In the last 30 days, how much difficulty did you have with sleeping such as falling asleep, waking up frequently during the night or waking up too early in the morning? (version wave 1)

In the last 30 days, how much difficulty did you have with sleeping? (version wave 2)

#### *Concentration*

Overall in the last 30 days how much difficulty did you have with concentrating or remembering things?

### Vignettes

#### *Mood Vignette 1*

Karen enjoys her work and social activities and is generally satisfied with her life. She gets depressed every 3 weeks for a day or two and loses interest in what she usually enjoys but is able to carry on with her day-to-day activities.

Overall in the last 30 days, how much of a problem did Karen have with feeling sad, low, or depressed? (version wave 1)

How much of a problem does Karen have with feeling sad, low, or depressed? (version wave 2)

#### *Mood Vignette 2 (only in wave 1)*

Anna feels depressed most of the time. She weeps frequently and feels hopeless about the future. She feels that she has become a burden on others and that she would be better dead.

Overall in the last 30 days, how much of a problem did Anna have with feeling sad, low, or depressed?

#### *Mood Vignette 3 (only in wave 1)*

Maria feels nervous and anxious. She worries and thinks negatively about the future, but feels better in the company of people or when doing something that really interests her. When she is alone she tends to feel useless and empty.

Overall in the last 30 days, how much of a problem did Maria have with feeling sad, low, or depressed?

### *Sleep Vignette 1*

Alice falls asleep easily at night, but two nights a week she wakes up in the middle of the night and cannot go back to sleep for the rest of the night.

In the last 30 days, how much difficulty did Alice have with sleeping, such as falling asleep, waking up frequently during the night or waking up too early in the morning? (version wave 1)

In your opinion, how much difficulty does Alice have with sleeping? (version wave 2)

### *Sleep Vignette 2 (only in wave 1)*

Maria takes about 2 h every night to fall asleep. She wakes up once or twice a night feeling panicked and takes more than 1 h to fall asleep again.

In the last 30 days, how much difficulty did Maria have with sleeping, such as falling asleep, waking up frequently during the night or waking up too early in the morning?

### *Sleep Vignette 3 (only in wave 1)*

Karen wakes up almost once every hour during the night. When he wakes up in the night, it takes around 15 min for her to go back to sleep. In the morning she does not feel well-rested.

In the last 30 days, how much difficulty did Karen have with sleeping such as falling asleep, waking up frequently during the night or waking up too early in the morning?

### *Concentration Vignette 1*

Lisa can concentrate while watching TV, reading a magazine or playing a game of cards or chess. Once a week she forgets where her keys or glasses are, but finds them within 5 min.

Overall in the last 30 days, how much difficulty did Lisa have with concentrating or remembering things? (version wave 1)

In your opinion, how much difficulty does Lisa have with concentrating or remembering things? (version wave 2)

### *Concentration Vignette 2 (only in wave 1)*

Sue is keen to learn new recipes but finds that she often makes mistakes and has to reread several times before she is able to do them properly.

Overall in the last 30 days, how much difficulty did Sue have with concentrating and remembering things?

### *Concentration Vignette 3 (only in wave 1)*

Eve cannot concentrate for more than 15 min and has difficulty paying attention to what is being said to her. Whenever she starts a task, she never manages to finish it and often forgets what she was doing. She is able to learn the names of people she meets.

Overall in the last 30 days, how much difficulty did Eve have with concentrating or remembering things?

## References

- Arbelaez, J. J., Ariyo, A. A., Crum, R. M., Fried, L. P., & Ford, D. E. (2007). Depressive symptoms, inflammation, and ischemic stroke in older adults: A prospective analysis in the cardiovascular health study. *Journal of the American Geriatrics Society*, *55*, 1825–1830.
- Bago d'Uva, T., Lindeboom, M., O'Donnell, O., Van Doorslaer, E. (2010). Slipping anchor? Testing the vignettes approach to identification and correction of reporting heterogeneity. mimeo. Erasmus School of Economics.
- Bago d'Uva, T., O'Donnell, O., & Van Doorslaer, E. (2008a). Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. *International Journal of Epidemiology*, *37*, 213–245.
- Bago d'Uva, T., Van Doorslaer, E., et al. (2008b). Does reporting heterogeneity bias the measurement of health disparities? *Health Economics*, *17*(3), 352–375.
- Barth, J., Schumacher, M., & Herrmann-Lingen, C. (2004). Depression as a risk factor for mortality in patients with coronary heart disease: A meta-analysis. *Psychosomatic Medicine*, *66*, 802–813.
- Blazer, D. G. (2003). Depression in late life: Review and commentary. *The Journal of Gerontology*, *58*, 249–265.
- Börsch-Supan, A., & Jürges, H. (Eds.). (2005). *The survey of health, aging, and retirement in Europe—methodology*. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).
- Braam, A. W., Prince, M. J., et al. (2005). Physical health and depressive symptoms in older Europeans. Results from EURODEP. *The British Journal of Psychiatry*, *187*, 35–42.
- Castro-Costa, E., Dewey, M., et al. (2007). Prevalence of depressive symptoms and syndromes in later life in ten European countries: The SHARE study. *The British Journal of Psychiatry*, *191*, 393–401.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*(386), 829–836.
- Copeland, J. R., Beekman, A. T., et al. (1999). Depression in Europe. Geographical distribution among older people. *The British Journal of Psychiatry*, *174*, 312–321.
- Copeland, J. R., Beekman, A. T., et al. (2004). Depression among older people in Europe: The EURODEP studies. *World Psychiatry*, *3*(1), 45–49.
- Fryers, T., Melzer, D., et al. (2005). The distribution of the common mental disorders: Social inequalities in Europe. *Clinical Practice and Epidemiology in Mental Health*, *1*, 14.
- Geerlings, S. W., Beekman, A. T., et al. (2000). Physical health and the onset and persistence of depression in older adults: An eight-wave prospective community-based study. *Psychological Medicine*, *30*(2), 369–380.
- Jürges, H. (2005). Cross-country differences in general health. In A. Börsch-Supan, et al. (Eds.), *Health, ageing and retirement in Europe—first results from the survey of health, ageing and retirement in Europe* (pp. 95–101). Mannheim Research Institute for the Economics of Ageing (MEA): Mannheim.
- Kapteyn, A., Smith, J., & Van Soest, A. (2007). Vignettes and self-reports of work disability in the United States and the Netherlands. *American Economic Review*, *97*(1), 461–473.
- King, G., Murray, C. J. L. M., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, *98*(1), 191–207.
- Koster, A., Bosma, H., et al. (2006). Socioeconomic differences in incident depression in older adults: The role of psychosocial factors, physical health status, and behavioral factors. *Journal of Psychosomatic Research*, *61*(5), 619–627.
- Kristensen, N., & Johansson, E. (2008). New evidence on cross-country differences in job satisfaction using anchoring vignettes. *Labour Economics*, *15*(1), 96–117.
- Kunst, A. E., Bos, V., et al. (2005). Trends in socioeconomic inequalities in self-assessed health in 10 European countries. *International Journal of Epidemiology*, *34*(2), 295–305.
- Larson, S. L., Owens, P. L., Ford, D., & Eaton, W. (2001). Depressive disorder, dysthymia, and risk of stroke: Thirteen-year follow-up from the baltimore epidemiologic catchment area study. *Stroke*, *32*, 1979–1983.
- Lenze, E. J., Rogers, J. C., et al. (2001). The association of late-life depression and anxiety with physical disability: A review of the literature and prospectus for future research. *The American Journal of Geriatric Psychiatry*, *9*(2), 113–135.
- Liebetrau, M., Steen, B., & Skoog, I. (2008). Depression as a risk factor for the incidence of first-ever stroke in 85-year-olds. *Stroke*, *39*, 1960–1965.
- Lorant, V., Croux, C., Weich, S., et al. (2007). Depression and socio-economic risk factors: 7-year longitudinal population study. *British Journal of Psychiatry*, *190*, 293–298.

- Lorant, V., Deliege, D., et al. (2003). Socioeconomic inequalities in depression: A meta-analysis. *American Journal of Epidemiology*, *157*(2), 98–112.
- Lyberaki, A., & Tinios, P. (2008). Poverty and persistent poverty: Adding dynamics to familiar findings. In A. Börsch-Supan, et al. (Eds.), *Health, ageing and retirement in Europe—first results from the survey of health, ageing and retirement in Europe*. Mannheim: Mannheim Research Institute for the Economics of Ageing (MEA).
- Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine*, *3*(11), e442.
- Miech, R. A., & Shanahan, M. J. (2000). Socioeconomic status and depression over the life course. *Journal of Health and Social Behavior*, *41*(2), 162–176.
- Murray, C. J. L., Ozaltin, E., Tandon, A., et al. (2003). Empirical evaluation of the anchoring vignettes approach in health surveys. In C. J. L. Murray & D. B. Evans (Eds.), *Health systems performance assessment: Debates, methods and empiricism* (pp. 369–400). Geneva: World Health Organization.
- Prince, M. J., Beekman, A. T., et al. (1999a). Depression symptoms in late life assessed using the EURO-D scale. Effect of age, gender and marital status in 14 European centres. *British Journal of Psychiatry*, *174*, 339–345.
- Prince, M. J., Reischies, F., et al. (1999b). Development of the EURO-D scale—a European Union initiative to compare symptoms of depression in 14 European centres. *British Journal of Psychiatry*, *174*, 330–338.
- Salomon, J. A., Tandon, A., et al. (2004). Comparability of self-rated health: Cross-sectional multi-country survey using anchoring vignettes. *BMJ*, *328*(7434), 258.
- Siegrist, J., & Wahrendorf, M. (2008). Quality of work and well-being: The European dimension. In A. Börsch-Supan, et al. (Eds.), *Health, ageing and retirement in Europe—first results from the survey of health, ageing and retirement in Europe*. Mannheim: Mannheim Research Institute for the Economics of Ageing (MEA).
- Tandon, A., Murray, C. J. L., Salomon, J. A., & King, G. (2002). *Statistical models for enhancing cross-population comparability, global programme on evidence for health policy discussion paper no. 42*. Geneva: WHO.
- Van Soest, A. (2008). Enhancing international comparability using anchoring vignettes. In A. Börsch-Supan, A. Brugiavini, H. Jürges, A. Kapteyn, J. Mackenbach, J. Siegrist, & G. Weber (Eds.), *Health, ageing and retirement in Europe (2004–2007)* (pp. 351–355). Mannheim: Mannheim Research Institute for the Economics of Ageing.
- Van Soest, A. H., Delaney, L. D., et al. (2007). Validating the use of vignettes for subjective threshold scales, UCD Geary Institute Working Paper.
- Vonková, H., Hullegie, P. (2010). Is the anchoring vignettes method sensitive to the domain and choice of the vignette? Netspar Discussion Paper 01/2010–004.
- Zunzunegui, M. V., Minicuci, N., et al. (2007). Gender differences in depressive symptoms among older adults: A cross-national comparison: The CLESA project. *Social Psychiatry and Psychiatric Epidemiology*, *42*(3), 198–203.