Later-Life Mental Health in Europe:
A Country-Level Comparison

George B. Ploubidis and Emily Grundy

Centre for Population Studies, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, University of London, UK.

Objectives. To investigate the influence of country of residence on depression and well-being among older Europeans, after establishing the between-country measurement invariance of both constructs.

Methods. We used data from a cross-sectional nationally representative population-based sample of older Europeans, the Survey of Health, Ageing, and Retirement in Europe (SHARE). The analysis sample comprised 13,498 older Europeans from nine countries. The EURO-D was used to measure depression, and a well-being outcome was derived from self-report items available in SHARE. The between-country measurement invariance of both mental health outcomes was established using modern psychometric modeling techniques.

Results. After adjustment for demographic characteristics and the presence of chronic illness, Spain was the country scoring highest on depression and Denmark highest on well-being. Optimal mental health was associated with higher educational attainment and being married.

Discussion. There is considerable between-country heterogeneity in later-life mental health in Europe. The Scandinavian countries, the Netherlands, and Austria, do best (low depression/high well-being), followed by Germany and France; whereas residents of Spain, Italy, and Greece report the worst mental health.

Key Words: Cross cultural comparison—Depression—Mental health—Structural equation modelling—Well-being

The 20th century witnessed significant improvements in somatic health in most countries, including substantial increases in survival to older ages and large reductions in late-age mortality. Although the initial driver of population aging in Europe has been falling fertility, these mortality changes have led to further population aging in many European countries, prompting an increased interest in the quality of life at older ages, including the important domain of mental health (Grundy, 2006). Aging often involves increased exposures to events and conditions associated with higher risks of depressive illness, including bereavement, somatic illness, and poverty, and rates of depressive illness are relatively high in older age groups, although it is unclear whether they continue to increase in later old age (Copeland, Beekman, Dewey, Hooijer, et al., 1999). It is also unclear whether older age is associated with reduced well-being, as studies have reported conflicting results (Lawton, Parmelee, Katz, & Nesselroade, 1996).

One of the key aims of government is to promote a society where citizens are happy, healthy, capable, and engaged (Marks & Hetan, 2005). Good mental health is a vital part of this aim, not only as an end in itself but also as an important contributor to other aims, such as economic productivity and social well-being. The concept of happiness, or satisfaction with life, is currently a major area of research in economics and psychology and has also recently become part of political discourse. Central to this area of research are country-level comparisons of levels of happiness or life satisfaction (Borooah, 2006), such as a recently published global map of happiness (White, 2007). National differences in depression and well-being may be valuable in providing insights into macrolevel influences on mental health in later life. Although each country is a unique case, the results of a comparative study may help to enhance our understanding of possible variations between European countries in the mental health status of older people and inform the debate about public health policy relevant to mental health in older age groups.

A common feature of previous studies is that they suffer from a major methodological limitation because none has formally addressed the issue of between-country measurement invariance. To engage in a meaningful country-level comparison of a mental health construct, the measurement invariance of the construct under study needs to be considered due to the possible influence of language, culture, different levels of expectations for the future, and other country-specific influences on mental health assessment. Furthermore, existing studies have either considered well-being (Borooah, 2006; Delhey, 2003) or depression (Ferring et al., 2004; Prince, Reischies, et al., 1999), but rarely both. Although it is clear that depressed individuals are not mentally healthy, it does not follow that nondepressed individuals have good mental health, and considering only one dimension may bias cross-country comparisons. If only depression is studied, relevant variation in positive aspects of mental health is not taken into account; similarly, if only
information from well-being indices is used, low scorers on well-being are grouped with those who are depressed. Presenting adjusted results from analyses that include indicators of both well-being and depression, as we do here, provides a more holistic view of country-level mental health than using either mental health construct alone.

In this paper, we use data from the Survey of Health, Ageing, and Retirement in Europe (SHARE) to derive invariant mental health measures for nine European countries and to analyze country differences in mental health using these derived measures. We also adjust for demographic characteristics because previous studies have reported associations between later-life depression and demographic variables (Prince, Beekman, et al., 1999), with women, the less educated, and the unmarried being at greater risk. Less is known about the effect of demographic characteristics on positive aspects of mental health, notably well-being, but, as might be expected, evidence suggests inverse associations to those reported for depression, with higher reported well-being among the married and those with higher levels of education (Isaacowitz & Smith, 2003). The primary aim of the paper is therefore to test for possible cross-national differences in depression and well-being when the demographic characteristics of the participants and their somatic health status are statistically controlled; a secondary aim is to assess the effect of demographic characteristics on the mental health of older Europeans. This study adds to the existing literature by considering both depression and well-being as outcomes and, to our knowledge, is the first country-level mental health comparison that takes into account the issue of between-country measurement invariance, using modern psychometric modeling techniques within the generalized structural equation modeling framework.

### Methods

#### Sample

SHARE is a multidisciplinary cross-national survey including data on the health, socioeconomic status, demographic characteristics, and social and family networks of individuals aged 50 years or over. Here we employ data from 9 of the 11 European countries included in the 2004 SHARE baseline study, excluding Switzerland due to its poor response rate and Belgium (data not available at the time of the analysis). The countries included are drawn from Northern (Denmark and Sweden), Western (Austria, France, Germany, and the Netherlands), and Mediterranean regions (Spain, Italy, and Greece). Full details of the SHARE sampling methodology have been reported elsewhere (Borsch-Supan & Jürges, 2005). Our initial sample comprised 6,782 male and 8,223 female respondents, of which 1,502 had to be excluded due to missing items of data leaving a sample of 13,498 respondents: 6,180 (45.8%) male and 7,318 (54.2%) female.

#### Measures

**Depression.**—We measured depression using the EURO-D scale that was developed and validated by the EURODEP Concerted Action Programme. The scale includes 12 binary (no/yes) items that enquire about depression, pessimism, wishing death, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment, and tearfulness (Prince, Reischies, et al., 1999). Positive responses indicate the presence of a symptom. The items originate from the Geriatric Mental State Scale (Gurland, Copeland, Sharpe, & Kelleher, 1976), the Center for Epidemiological Studies Depression Scale (Radloff, 1977), the Zung Self-Rating Depression Scale (Zung, 1965), and the Comprehensive Psychopathological Rating Scale (Asberg, Montgomery, Perris, Schalling, & Sedvall, 1978). The strong associations between the EURO-D and its parent instruments, measured as correlations for continuous measures and as areas under the receiver operating characteristic curves for the dichotomous measures, suggest that the EURO-D has captured the essence of the instruments from which it was derived (Prince, Beekman, et al., 1999).

**Well-being.**—We selected 10 items from the self-completion section of SHARE. Selection was made on the basis of the face validity of the items with respect to life satisfaction, presence of positive mood, and happiness—the three components of subjective well-being (Diener, Suh, Lucas, & Smith, 1999). Seven of the items were drawn from the Control–Autonomy–Self-realization–Pleasure questionnaire included in SHARE (Wiggins, Higgs, Hyde, & Blane, 2004). The response format was a 4-point Likert scale for 6 items (e.g., “How satisfied are you with your life in general?” Very satisfied, Satisfied, Dissatisfied, Very
dissatisfied; “I feel the future looks good for me.” Often, Sometimes, Rarely, Never) and a 5-point Likert scale for the remaining 4 (e.g., “I am always optimistic about my future.” Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly disagree).

Demographic characteristics.—It has been shown that depression and well-being are associated with socioeconomic status, age, and marital status (Copeland, Beeckman, Dewey, Jordan, et al., 1999; Simon, 2002; Stansfeld, Clark, Rodgers, Caldwell, & Power, 2008; Stordal, Mykletun, & Dahl, 2003). Somatic health and disability are also associated with both mental health outcomes as well as with demographic characteristics (Mykletun et al., 2007; Roysamb, Tambs, Reichborn-Kjennerud, Neale, & Harris, 2003), with this effect not necessarily being uniform between countries. For these reasons, it is clearly important to control for demographic characteristics as well as for mental health when investigating country-level differences in mental health.

Age ($M = 64.5$ years, $SD = 9.7$) was entered as a continuous variable in the model. Educational qualifications were classified according to the International Standard Classification of Educational Degrees (ISCED 1997) scheme (Hollmeyer-Zlotnik & Wolf, 2003) and recoded to three categories: grouping together ISCED codes 0, 1, and 2 (low: pre-primary level of education, primary level of education, and lower secondary level of education), ISCED code 3 (mid: upper secondary level of education), and ISCED codes 4, 5, and 6 (high: postsecondary nontertiary education and higher qualifications). Marital status was dichotomized into married or in a registered partnership and not married (widowed, divorced or separated, never married). Finally, a continuous summary variable representing the total reported number of chronic illnesses was derived based on questions to respondents. Participants were asked about the presence of the following chronic illnesses/conditions: heart attack or other heart problems; stroke or cerebral vascular disease; diabetes; chronic lung disease; asthma, arthritis or rheumatism; osteoporosis; cancer or malignant tumor; stomach, duodenal, or peptic ulcer; Parkinson’s disease; cataract; hip or femoral fracture, and other conditions. Positive responses were summed to derive a continuous summary variable representing the reported number of chronic illnesses.

Statistical Modeling

Psychometric modeling.—The derivation of unbiased mental health measures is usually approached using forms of factor analytic measurement models. At the individual level, these models, when appropriately used, capture the true variance of the construct that is being measured, although controlling for exogenous-to-mental-health sources of error such as social desirability and survey context. However, in international comparisons, this approach assumes a uniform distribution of error between countries. This assumption is unlikely to be valid, as there are additional sources of error arising from the translation of mental health items to several languages and cultural differences in the understanding of items and in response tendencies. These probably vary by country, so their distribution as sources of error cannot be assumed to be uniform. The establishment of between-country measurement invariance where both uniformly and not uniformly distributed sources of measurement error are controlled is needed in order to obtain an accurate country-level comparison of mental health.

Within the generalized latent variable modeling framework, measurement equivalence or invariance is analogous to factorial invariance. Factorial invariance is a set of hypotheses stating that measurement model parameters function without bias across groups or occasions. According to Drasgow (1984), measurement equivalence exists “when the relations between observed test scores and the latent attribute measured by a test are identical across subpopulations.” Factorial invariance can be tested with multigroup confirmatory factor analysis (CFA) (Joreskog, 1971). The multiple group (global) chi-square test statistic and the indices of fit derived from it for this multiple group model are used to assess invariance or, in other words, assess whether a confirmatory factor analytic model with fixed-model parameters among groups fits the data.

There are three types—levels—of factorial invariance (Meredith, 1993):

1. Weak or configural invariance, where the factorial structure of the model is the same across groups, but the model parameters are freed.
2. Metric invariance, where invariance is achieved when, in addition to the factorial structure, the scale metric is set to be equal among groups.
3. Strict factorial invariance, where the structure, metric, and all the measurement model parameters are fixed to be equal among groups.

In order to estimate an invariant between-country model, we first had to establish the factorial structure of the EURO-D and the well-being items in our sample data. Therefore, we first identified the appropriate factorial structure of the EURO-D and the well-being outcome using the pooled sample. We next estimated a between-country multigroup CFA of the best fitting model, constraining the item loadings and thresholds (common threshold structure) and their respective standard errors to be equal across groups (countries), therefore testing for strict factorial invariance. Latent trait scores derived from the multigroup CFA were calculated for both the EURO-D and the well-being outcome and were used in subsequent analyses. Latent trait scores can theoretically range from $-\infty$ (minus infinity) to $+\infty$ (infinity), but in practice, the range is usually from $-3$ to $3$. 
Our factor analytic models of well-being and depression are based on a multivariate probit analysis with latent variables (Muthén, 1983) through a two-parameter normal ogive item–response model and its extension to polytomous/ordinal data (Muthén, 1984). The factor loading reflects the strength of the association between the observed item and the latent construct. The threshold parameter reflects the point of the latent construct that needs to be reached for a particular response option to be endorsed. Factorial invariance for binary and polytomous/ordinal items is achieved when measurement parameters (thresholds, factor loadings, and their associated standard errors) function equivalently in each group of a multigroup CFA model (Joreskog & Moustaki, 2001). All latent variable models were estimated with the weighted least squares, mean- and variance-adjusted estimator using the Mplus 5.1 software (L. K. Muthén and B. O. Muthén, 1997–2008). Model fit was assessed with the comparative fit index (CFI), the Tucker–Lewis index (TLI), and the root mean square error of approximation (RMSEA) following the recommendations of Yu (2002) on their interpretation, as well as through difference tests between nested models.

**Factorial Structure and Establishment of the Between-Country Measurement Invariance of the Mental Health Measures**

Depression.—In the EURO-D literature, there is no clear guidance on its factorial structure, although the results of previously reported principal component analyses suggest a possible two-factor solution (Prince, Reischies, et al., 1999). We began to examine the structure of the EURO-D with an exploratory factor analysis (EFA) suitable for binary variables. There were two eigenvalues larger than 1, indicating that two latent factors were necessary to account for responses to the EURO-D items. With the EFA results as a guide, we proceeded by testing two restricted CFA models for the EURO-D. We first estimated a model with two first-order factors and second a general specific, bifactor model in which a global latent factor accounts for variation in all EURO-D items. Because both models were confirmed by the pooled sample data, we then performed and compared two multigroup CFAs, fixing all model parameters (loadings, thresholds, and errors) to be equal between countries. The general specific model was superior in both the pooled sample and multigroup CFA and is depicted in Figure 1. To further test the superiority of the general specific models against the first-order models in both the pooled sample and multigroup analyses, we performed formal comparisons of the overall $\chi^2$ (difference tests), with the appropriate procedures for the binary and ordinal nature of our data. In the pooled sample, the overall $\chi^2$ statistic of the general specific model, $\chi^2(39) = 403.3$, had a significantly lower value (indicating good fit) compared with the first-order model $\chi^2(49) = 1,762.8$; $\chi^2(\text{difference}) (12) = 1,167.8, p < .001$. Similarly in the multicountry analysis, the overall $\chi^2$ statistic of the general specific model, $\chi^2(448) = 1,611.4$, had a significantly lower value (indicating good fit) compared with the first-order model $\chi^2(459) = 2,668.3$; $\chi^2(\text{difference}) (3) = 1,016.1, p < .001$. The general specific structure was therefore superior compared with all other structures both in the pooled and multigroup analyses. The overall fit of the multicountry general specific model as evidenced by the global indices of fit (CFI, TLI, and RMSEA) provided evidence for strong between-country measurement invariance with respect to the EURO-D because the fit of such a restrictive model to the data (all models parameters fixed to be equal between countries) is considered to be the most conservative test for measurement invariance within the factor analytic framework (Meredith, 1993). A comparison of the selected model with less restrictive invariance models (factor loadings and/or thresholds free to vary between groups) was therefore not needed because despite the expected worse fit of the most

---

**Table 2. Goodness of Fit Criteria for Pooled Sample and Multigroup CFA**

<table>
<thead>
<tr>
<th></th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EURO-D</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pooled sample, two first-order factors</td>
<td>0.946</td>
<td>0.962</td>
<td>0.041</td>
</tr>
<tr>
<td>Multigroup CFA, two first-order factors</td>
<td>0.933</td>
<td>0.946</td>
<td>0.046</td>
</tr>
<tr>
<td>Pooled sample nested model</td>
<td>0.985</td>
<td>0.987</td>
<td>0.023</td>
</tr>
<tr>
<td>Multigroup (Country) CFA, nested general specific model</td>
<td>0.963</td>
<td>0.969</td>
<td>0.035</td>
</tr>
<tr>
<td>Multigroup (Gender) CFA, nested general specific model</td>
<td>0.982</td>
<td>0.985</td>
<td>0.024</td>
</tr>
<tr>
<td>Multigroup (Age) CFA, nested general specific model</td>
<td>0.980</td>
<td>0.983</td>
<td>0.026</td>
</tr>
<tr>
<td>Multigroup (Education) CFA, nested general specific model</td>
<td>0.989</td>
<td>0.991</td>
<td>0.019</td>
</tr>
<tr>
<td>Multigroup (Marital status) CFA, nested general specific model</td>
<td>0.982</td>
<td>0.985</td>
<td>0.024</td>
</tr>
<tr>
<td><strong>Well-being</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pooled sample, four first-order factors</td>
<td>0.987</td>
<td>0.994</td>
<td>0.045</td>
</tr>
<tr>
<td>Multigroup sample, four first-order factors, one second-order factor</td>
<td>0.935</td>
<td>0.973</td>
<td>0.100</td>
</tr>
<tr>
<td>Pooled sample, four first-order factors, one second-order factor</td>
<td>0.980</td>
<td>0.991</td>
<td>0.054</td>
</tr>
<tr>
<td>Pooled sample, nested general specific model</td>
<td>0.932</td>
<td>0.958</td>
<td>0.101</td>
</tr>
<tr>
<td>Multigroup (Country) CFA, nested general specific model</td>
<td>0.999</td>
<td>0.999</td>
<td>0.019</td>
</tr>
<tr>
<td>Multigroup (Gender) CFA, nested general specific model</td>
<td>0.963</td>
<td>0.979</td>
<td>0.077</td>
</tr>
<tr>
<td>Multigroup (Age) CFA, nested general specific model</td>
<td>0.997</td>
<td>0.998</td>
<td>0.026</td>
</tr>
<tr>
<td>Multigroup (Education) CFA, nested general specific model</td>
<td>0.997</td>
<td>0.998</td>
<td>0.024</td>
</tr>
<tr>
<td>Multigroup (Marital status) CFA, nested general specific model</td>
<td>0.996</td>
<td>0.997</td>
<td>0.030</td>
</tr>
</tbody>
</table>

Note: CFA = confirmatory factor analysis; CFI = comparative fit index, values >0.95 indicate good fit; TLI = Tucker–Lewis index, values >0.95 indicate good fit; RMSEA = root mean square error of approximation, values <0.06 indicate good fit.
restrictive model compared with either of the less restrictive models, the former is still well above (or below) the recommended cutoffs for model fit.

In Figure 1, we report the standardized factor loadings of all items to the latent factors. The factor loadings on the global EURO-D depression factor ranged from 0.40 to 0.88. Further multigroup CFAs, with the general specific model, revealed measurement invariance of the EURO-D with respect to gender, age (three category grouping, 50−64, 65−75, and 75+ years), educational attainment, and marital status (see Table 3 for all indices of fit), confirming the construct validity of the EURO-D in all subgroups in the SHARE sample. We calculated latent trait scores based on the multigroup general specific model and used the global EURO-D depression factor in all further analyses.

Well-being.—The selected well-being items were entered into an EFA suitable for ordinal data. There were four eigenvalues greater than 1, indicating the presence of four factors. We proceeded by testing three competing CFA models based on the EFA results. These included (a) a model with four first-order factors, (b) a model additionally including a second-order factor, and (c) a general specific model. The general specific model was the best fitting model in both pooled sample and multigroup CFAs. The value of the RMSEA (0.077) was within the range defined as indicative of an acceptable (values <0.08) rather than a good fit (values <0.06) (Yu, 2002). We considered this adequate evidence for model fit for the purposes of this paper, especially for such a heavily restricted model, taking into account the fact that both the CFI and TLI were well above the recommended criteria.

Table 3. Adjusted Marginal means of Depression and Well-being Scores from Main Effects Only MANCOVA Model

<table>
<thead>
<tr>
<th>Country</th>
<th>Pooled sample</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Depression</td>
<td>Well-being</td>
<td>Depression</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.040</td>
<td>0.433</td>
<td>−0.043</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.153</td>
<td>0.139</td>
<td>0.037</td>
</tr>
<tr>
<td>Austria</td>
<td>0.030</td>
<td>0.045</td>
<td>−0.059</td>
</tr>
<tr>
<td>Germany</td>
<td>0.076</td>
<td>−0.230</td>
<td>−0.038</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>0.095</td>
<td>0.227</td>
<td>0.021</td>
</tr>
<tr>
<td>France</td>
<td>0.321</td>
<td>−0.250</td>
<td>0.185</td>
</tr>
<tr>
<td>Spain</td>
<td>0.392</td>
<td>−0.083</td>
<td>0.179</td>
</tr>
<tr>
<td>Italy</td>
<td>0.241</td>
<td>−0.404</td>
<td>0.115</td>
</tr>
<tr>
<td>Greece</td>
<td>0.135</td>
<td>−0.365</td>
<td>−0.004</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISCED-97 codes 0, 1, and 2 (Low)</td>
<td>0.222</td>
<td>−0.173</td>
<td>0.086</td>
</tr>
<tr>
<td>ISCED-97 code 3 (Mld)</td>
<td>0.154</td>
<td>−0.060</td>
<td>0.046</td>
</tr>
<tr>
<td>ISCED-97 codes 4, 5, and 6 (High)</td>
<td>0.119</td>
<td>0.071</td>
<td>−0.001</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not married⁴</td>
<td>0.234</td>
<td>−0.170</td>
<td>0.107</td>
</tr>
<tr>
<td>Married⁵</td>
<td>0.096</td>
<td>0.062</td>
<td>−0.020</td>
</tr>
</tbody>
</table>

Notes: MANCOVA = multivariate analysis of variance; ISCED = International Standard Classification of Educational Degrees.

⁴Widowed, divorced or separated, or never married.
⁵Or in a registered partnership.
To further test the superiority of the general specific model against the first-order model in both the pooled sample and multigroup analyses, we performed formal comparisons of the overall \( \chi^2 \) statistic (difference tests), using appropriate procedures to take account of the ordinal nature of the well-being items. In the pooled sample, the overall \( \chi^2 \) statistic of the general specific model, \( \chi^2(11) = 66.9 \), had a significantly lower value (indicating good fit) compared with the first-order model \( \chi^2(25) = 915.6; \) \( \chi^2(\text{difference}) (15) = 814.5, p < .001 \). Similarly in the multicountry analysis, the overall \( \chi^2 \) statistic of the general specific model, \( \chi^2(272) = 2,938.7 \), had a significantly lower value (indicating good fit) compared with the first-order model \( \chi^2(307) = 5,147.1; \) \( \chi^2(\text{difference}) (40) = 1,071.8, p < .001 \). As with the EU-RO-D, the fit of the multigroup general specific model, where all model parameters were fixed to be equal between countries, provided evidence for strong between-country measurement invariance of the well-being measurement model.

The factor loadings on the general well-being factor ranged from 0.48 to 0.83 (see Figure 2). Further multigroup CFAs using our best fitting model showed the factorial and measurement invariance of our well-being outcome with respect to gender, age group (50–64, 65–74, and 75+ years), educational level, and marital status (see Table 3 for all indices of fit), confirming the construct validity of the derived well-being outcome in all subgroups in the SHARE sample. Because we were interested in a population-based continuum of well-being, we calculated latent trait scores based on the multigroup general specific model and used the global well-being factor in all further analyses.

Multivariate analysis of covariance.—The depression and well-being latent trait scores were entered as dependent variables in a multivariate analysis of covariance (MANCOVA). Country of residence, gender (pooled sample analysis only), educational level, and marital status were entered as independent categorical variables and age and number of chronic illnesses as continuous covariates. The effects of the independent variables and the covariates are adjusted for by each other as well as by the negative moderate correlation between depression and well-being (\( r = -0.440 \)). All independent variables, the two covariates, plus the two-way interactions tabulated in Table 4 were included in the models. In Table 4, main effects are derived from a main effects model only, whereas the two-way interactions come from a model with added interaction terms. Although a similar analysis could be performed within the standard error of the mean (SEM) framework, the MANCOVA approach was preferred because it allows for straightforward testing of possible interactions among the independent variables, as well as allowing tests of pairwise comparisons between levels of the independent variables, whereas within the SEM framework all comparisons would have been against a selected reference category.

**Results**

**The MANCOVA**

**Men.**—The multivariate Wilk’s lambda criterion indicated that the linear combination of the dependent variables was significantly associated with country of residence, age, education, number of reported chronic illnesses, and marital status. No interactions reached statistical significance.

**Depression.**—Overall the contribution of the combined independent variables to depression was 8.3%. Country of residence accounted for 1.5% of the overall variance. The estimated marginal means of depression by country are presented in Table 3 (for women as well as for men) and shown graphically in Figure 3. Spain and France were the countries with the highest depression scores; Austria,
Germany, Sweden, and Denmark had the lowest scores. Age and the reported number of chronic illnesses were positively associated with depression in men. Educational level was inversely associated with depression, with the least educated respondents having the highest depression score. Finally, married respondents had significantly lower depression scores than the not married. We did not observe any significant interactions.

Well-being.—The overall effect of the combined independent variables on well-being was 13%. Country of residence, age, number of chronic illnesses, education, and marital status were all significantly associated with well-being, with country of residence being the strongest predictor, accounting for 8.2% of the overall variance of well-being. The effect of country (for both men and women) is shown in Figure 4. Denmark, the Netherlands, and Sweden had the highest scores on well-being; Greece and Italy the lowest. Age and the number of reported chronic illnesses were inversely associated with well-being, whereas respondents with the highest educational qualifications (ISCED codes 4, 5, and 6) had the highest well-being. Married respondents had a significantly higher well-being score than the non married. We did not observe any significant interactions among the independent variables.

Women.—The multivariate Wilk’s lambda criterion indicated that the linear combination of the dependent variables was significantly associated with country of residence, age, education, number of reported chronic illnesses, and marital status. As for men, no interaction reached statistical significance.

Depression.—Overall the contribution of the combined independent variables to depression was 12.8%. Country of residence accounted for 3.6% of the overall variance. The

![Figure 3. Country-level depression comparison (adjusted means).](image-url)
estimated marginal means of depression by country are shown in Table 3 and Figure 3. Spain and France were the countries with the highest depression scores, and Denmark had the lowest. Number of reported chronic illnesses, but not age, was positively associated with depression. As for men, educational level was inversely associated with depression, and married respondents had significantly lower depression scores than the not married.

Well-being.—The overall effect of the combined independent variables on well-being was 20.6%. Country of residence, age, the number of chronic illnesses, education, and marital status were all significantly associated with well-being, with country of residence being the strongest predictor, accounting for 11.5% of the overall variance. As shown in Figure 4, women in Denmark, Sweden, and the Netherlands had the highest scores on well-being, and those in Greece and Italy the lowest. Age had a significant inverse association with well-being, as did number of reported chronic illnesses. Additionally, respondents with the highest educational qualifications had higher well-being scores than the less well educated. We observed a significant interaction between country of residence and marital status, with Italian women not receiving the benefit of being married.

Pooled sample.—The multivariate Wilk’s lambda criterion indicated that the linear combination of the dependent variables was significantly associated with the country of residence, age, number of reported chronic illnesses, education, and marital status. We observed significant two-way interactions (with either or both depression and well-being) of country with gender, country with marital status, as well as country with age.

Depression.—Gender had a significant effect on depression with women scoring higher than men in all countries. Country of residence, age, education, number of chronic illnesses, and marital status were also associated with depression. We found a significant interaction between country of residence and gender. This was due to the fact that in Denmark, the Netherlands, Austria, and Sweden, there was no significant difference between the depression scores of men and women, whereas in all other countries women had significantly higher depression scores than men. Results also showed a significant interaction between age and gender, with age being associated with higher risks of depression only in men, as well as between country of residence and gender and country of residence and marital status.

Well-being.—We did not find a significant effect of gender on well-being, but results indicated an interaction between gender and country of residence. In Austria, France, and Greece, men reported significantly higher well-being than women; in the other countries considered, women reported similar well-being to men or even higher well-being in the case of Denmark and Sweden. Country of residence, age, education, number of chronic illnesses, and marital status were also associated with well-being. Results also showed a significant interaction between country of residence and gender and country of residence and marital status.

Discussion

Our results indicate that, after adjustment for demographic characteristics, there is considerable between-country heterogeneity in both later-life depression and well-being. This heterogeneity is unlikely to be an artifact of between-country measurement bias because the measurement invariance of measures of depression and well-being was established using modern psychometric modeling techniques. The absolute value of country means for depression and well-being may to some extent be influenced by response tendencies, with the residents of the Scandinavian countries being susceptible to over reporting of positive health (Jurges, 2007). However, the strict factorial invariance established here and especially the between-country equation of the threshold parameters implies that the influence of this exogenous source of error has been equalized between countries. Furthermore, our results are unlikely to be influenced by other sources of country-level bias, such as cultural differences and differences in expectations and in language because these potential biases were also explicitly controlled. Therefore, the differences in the depression and well-being country means and the resulting ranking of countries cannot be attributed to differences in response tendencies or other country-level sources of bias. Similarly, differences in the reported means by gender, educational level, and marital status cannot be attributed to variation in response styles or other influences exogenous to depression and well-being because measurement invariance was established for each of these separately, confirming the construct validity of the EURO-D and the well-being outcome across all subpopulations present in our sample.
For Europeans aged over 50 years, the Scandinavian countries, Austria, and the Netherlands seem to do best in terms of mental health (high well-being and low depression), followed by France, Austria, and Germany (medium or low depression, medium or high well-being), whereas older people in Italy, Greece, and Spain have the worst mental health (high depression, low well-being). Spain is the country with the highest average depression score and Denmark the country with the highest score on well-being. Our findings are generally in agreement with previous studies on depression (Copeland, Beekman, Dewey, Jordan, et al., 1999) and well-being (Delhey, 2003), despite the methodological differences with our study.

One possible factor underlying the observed country-level variation might be the availability of state-provided supports and services, as these may be particularly important for mental health because the availability of state-provided safety nets may enhance feelings of security and reduce anxiety. With the notable exception of France, it appears that in countries with generous social security schemes, such as the Scandinavian countries, the Netherlands, and Austria, where per capita public expenditure on health and welfare services is among the highest in Europe, people are less depressed and have higher well-being than in other countries where the state provides less. These findings are consistent with previous research which has examined associations between welfare state regime typologies and population health and concluded that population health is enhanced by the relatively generous and universal welfare provision of the Scandinavian countries (Navarro et al., 2006). In a similar vein, it has been suggested that the introduction of universal old pensions contributed to historical declines in suicide rates at older ages observed in England and Wales (Murphy, Lindesay, & Grundy, 1986).

Our results also indicate that country of residence is more strongly associated with well-being than with depression and that the association is stronger for women than men. Further research is required to identify factors that might account for these differences. A related implication is that when mental health outcomes are investigated using pooled data from several European countries, a multilevel approach should be adopted because not accounting for country-level heterogeneity may lead to biased standard errors and overrejection of the null hypothesis (Type A error).

With regard to associations between demographic characteristics and mental health, in models in which the number of reported chronic illnesses was controlled, older age was associated with higher depression and lower well-being in both men and women. Associations between physical and mental health in later life have been widely reported (Yang, 2007), but it seems that for older Europeans increases in somatic health conditions cannot account for all the effect of age on depression and well-being. However, we lacked other indicators of physical health or disability. With respect to the other demographic characteristics, higher educational attainment was associated with better mental health (low depression, high well-being), and being married was equally beneficial (Inaba et al., 2005), except among Italian women, for whom being married was associated with higher levels of depression.

We confirmed one of the most robust findings in psychiatric epidemiology, with levels of depression being higher in women than men (Minicuci, Maggi, Pavan, Enzi, & Crepaldi, 2002). The female excess in depression seems to be more prominent in Southern European countries as indicated by the observed country-by-gender interaction, a finding in accordance with the EURODEP report (Copeland, Beekman, Dewey, Hooijer, et al., 1999) and other studies (Zunzunegui et al., 2007), but it was also evident in Austria. Gender differences in depression may reflect genetic influences, with Kendler (Kendler, Gardner, Neale, & Prescott, 2001), for example, showing that the heritability of depression is greater in women than in men, but it has been also attributed to environmental influences, such as socioeconomic conditions (Jeon, Jang, Rhee, Kawachi, & Cho, 2007; Zunzunegui et al., 2007). With respect to the environmental component of depression, two nonexclusive explanations have been offered to account for how social factors relate to gender differences in health: the differential vulnerability hypothesis and the differential exposure hypothesis. According to the differential vulnerability hypothesis, the higher prevalence of depression in women reflects the stronger impact of socioeconomic position, social relations, and chronic stressors on the physical and mental health of women compared with men (Rieker & Bird, 2005). The differential exposure hypothesis states that women experience more depression-related adverse circumstances, such as social and economic disadvantage and lifetime stressors (Zunzunegui et al., 2007). The relative contribution of the two processes cannot be inferred from our results and by the information given at present in the SHARE dataset. Further research especially using life course data is needed to investigate whether the gender gap in depression is due to different vulnerability between women and men or to differences in the accumulation of adverse exposures or a possible interaction between the two processes.

Women scored higher than men in well-being as the pooled sample MANCOVA including gender as an independent variable revealed a finding in disagreement with previous research (Pinquart & Sorensen, 2001). In contrast with our findings on depression, a significant interaction between gender and country was observed with the pattern of country-level mean scores on well-being not being the same for both genders. In some countries, women report lower well-being than men, whereas in others they report higher. Men in Austria, France, and Greece report higher well-being than women, and these were also the countries where the female excess in depression was most prominent. In the other countries, women reported similar well-being to men or even higher well-being in the case of Denmark, Sweden, and the Netherlands.
A possible explanation for the interactions between gender and country with respect to both depression and well-being may be differences in gender roles, which are thought to lead to differences in the experience of life events, which in turn may lead to depression (Nazroo, Edwards, & Brown, 1998), as well as access to resources such as social support that may buffer the effects of stress (Matud, Ibanez, Bethencourt, Marrero, & Carballeira, 2003). Furthermore, Southern European countries are characterized in many typologies as having a strong “familistic” culture (Reher, 1998). Previous comparisons of levels of family support for older people have found that these tend to be high both in Southern Europe (where formal service provision is low) and in Austria and Germany (Pacolet, Bouten, Lnaoye, & Versieck, 2000). Possibly associated differences in women’s actual and perceived responsibilities for caregiving may contribute to the country-level variations in mental health we identified, although further research would be needed to see if this is the case.

Strengths of this study include the use of a population-based sample, the establishment with the use of modern psychometric methods of the between-country measurement equivalence of depression and well-being, and the inclusion of both in a country-level comparison. A weakness is potential bias arising from sample nonresponse and missing data. We tried to address this issue by analyzing partial incomplete datasets in which missing data were imputed using the multiple imputation method under the missing at random assumption, and the results were similar to the ones presented here (results available from corresponding author). Future studies are needed to see if our findings are replicated using samples from other countries and to identify the factors that may account for the effect of country of residence on mental health—particularly because we found that this effect was stronger than the effect of demographic characteristics on well-being and depression, therefore enhancing our understanding of variations in mental health at the country, as well as at the individual, level.

Assuming that the observed between-country heterogeneity in late-life mental health in Europe is mostly due to environmental influences, this would suggest some potential for policies and population-based interventions to alleviate depression and enhance well-being in the older population. As the older population is large and growing and rates of depression increase with age, any success in these areas would potentially have a large public health impact, especially as the children, grandchildren, and other relatives of older people might also benefit from improvements in the mental health of their relatives.

**Funding**

This research was partly funded by the Major Aging and Gender Issues in Europe project 28571, within the 6th European Commission Framework.

**Conflict of Interest**

None declared.

**Acknowledgments**

Author contributions: George Ploubidis undertook the statistical modeling, manuscript preparation, and revision. Emily Grundy advised on the statistical modeling and contributed to manuscript preparation and revision as well as the identification of the research questions.

**Correspondence**

Address correspondence to George Ploubidis, MSc, PhD, Centre for Population Studies, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, University of London, 49/50 Bedford Square, London, WC1B 3DP, UK. Email: George.Ploubidis@LSHTM.ac.uk

**References**


