Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans

Teresa Bago d’Uva,1,2* Owen O’Donnell3 and Eddy van Doorslaer1,2,4

Accepted 18 June 2008

Background This study aims to establish whether health reporting differs by education level and, if so, to determine the extent to which this biases the measurement of health inequalities among older Europeans.

Methods Data are from the Survey of Health, Ageing and Retirement in Europe (SHARE) covering eight countries. Differential reporting of health by education is identified from ratings of anchoring vignettes that describe fixed health states. Ratings of own health in six domains (mobility, pain, sleep, breathing, emotional health and cognition) are corrected for differences in reporting using an extended ordered probit model. For each country and health domain, we compare the corrected with the uncorrected age-sex standardized high-to-low education rate ratio for the absence of a health problem.

Results Before correction for reporting differences across the 48 combinations of country by health domain, there was no inequality in health by education ($P>0.05$) in 32 of 48 cases. However, there were reporting differences by education ($P<0.05$) in 29 out of 48 cases. In general, higher educated older Europeans are more likely to rate a given health state negatively (except for Spain and Sweden). Correcting for these differences generally increases health inequalities (except for Spain and Sweden) and results in the emergence of inequalities in 18 cases ($P<0.05$), which may be considered ‘statistically significant’. The greatest impact is in Belgium, Germany and The Netherlands, where inequalities ($P<0.05$) appear only after adjustment in four of the six health domains.

Conclusions These results emphasize the need to account for differences in the reporting of health. Measured health inequalities by education are often underestimated, and even go undetected, if no account is taken of these reporting differences.

Keywords Health measurement, self-rated health, socioeconomic status, reporting differences, vignettes

1 Department of Applied Economics, Erasmus School of Economics, Erasmus University, Rotterdam, The Netherlands.
2 Tinbergen Institute, Amsterdam, The Netherlands.
3 Department of Balkan, Slavic & Oriental Studies, University of Macedonia, Thessaloniki, Greece.
4 Department of Health Policy and Management, Erasmus University Medical Centre, Rotterdam, The Netherlands.
* Corresponding author. Department of Applied Economics (Room H13-09), Erasmus School of Economics, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.
E-mail: bagoduva@few.eur.nl
Introduction

The measurement of socioeconomic inequalities in health has made extensive use of self-rated measures of health.\textsuperscript{1–9} The popularity of self-rated health (SRH) is driven by its low cost and feasibility in large-scale surveys, but is also justified by extensive evidence showing it to be a powerful predictor of mortality\textsuperscript{10,11} and of medical care use.\textsuperscript{12,13} There are, however, concerns that, besides containing valuable information on health status, SRH may vary with conceptions of what constitutes good health and expectations for own health.\textsuperscript{14} If such beliefs and expectations vary with socioeconomic status (SES), then differences in SRH will not provide an unbiased measure of socioeconomic inequality in health.

The problem is illustrated in Figure 1, which shows the hypothetical mapping from latent true health into categorical responses of a SRH instrument for representative high and low educated individuals. For illustrative purposes, all response thresholds are assumed higher for the more educated person, although different scenarios are possible. In this example, for any given true level of health, the more educated person will report worse health on the categorical scale. For example, if $H^*_L$ and $H^*_H$ represent the true health levels, then both report their health as ‘moderate’ despite the fact that the more highly educated person has better true health. Relying on SRH would lead to the false conclusion of no socioeconomic inequality in health. With data on SRH only, differences in reporting behaviour cannot be disentangled from differences in true health and SRH measures may provide a distorted picture of socioeconomic inequalities in health.

Differences in reporting of health by SES can arise because individuals report their health relative to that of peers, with the result that SRH underestimates socioeconomic inequality in health. The same bias would arise if more highly educated individuals are better informed of treatment options and so are less tolerant of a given health condition. But there can also be bias in the opposite direction if, for example, health problems are reported as a justification for not working,\textsuperscript{15} or if higher income individuals, perhaps driven by a belief that they should be in good health, use more lenient standards in reporting their own health status.\textsuperscript{16,17} The direction and extent of the bias has been found to differ across countries.\textsuperscript{18,19}

Several studies have investigated reporting differences in SRH by demographic and SES. One approach has been to test whether the ability of SRH to predict mortality varies across socio-demographic groups. With respect to SES, the results from different countries are strikingly mixed.\textsuperscript{18} Some studies have found no differences in the predictive ability of SRH by SES,\textsuperscript{20,21} others have found SRH to be more strongly related to mortality at higher SES\textsuperscript{22,23} and still others have found the opposite.\textsuperscript{24,25} Another approach has been to examine whether, after controlling for a health indicator that is presumed to be more objective, there is any systematic variation in SRH, which is argued to be attributable to reporting differences. Lindeboom and Van Doorslaer\textsuperscript{26} using Canadian data found no evidence of reporting heterogeneity by education and income, while Etilé and Milcent\textsuperscript{27} found some evidence in French data of reporting differences by income. The problem with these approaches is that truly objective indicators, such as mortality, are rarely available and, in any case, do not capture non-fatal health conditions, while other indicators, like diseases or health limitations, are usually also self-reported and potentially contaminated by similar measurement errors as SRH.\textsuperscript{28,29}

As noted in a recent editorial in this journal,\textsuperscript{18} a potentially attractive alternative is to identify reporting behaviour directly through the rating of case vignettes that describe fixed levels of functioning within a given health domain.\textsuperscript{30,31} Survey respondents are asked to rate both these hypothetical cases and their own health on the same response scale. If respondents evaluate the same hypothetical case differently, there is evidence of reporting differences. This makes it possible to identify systematic differences in response thresholds in relation to socio-demographic characteristics. Assuming that individuals rate the vignettes in the same way as their own health (response consistency), the thresholds obtained from the vignette responses can be imposed on the model for reported own health, making it possible to identify differences in true health by SES and not merely a mixture of health and reporting differences. One can

![Figure 1](image-url)
then compute the health that each group would report if they all used the response thresholds of a reference group. That is, one can measure health on a comparable scale. For example, in terms of Figure 1, the health of the high education individual, $H^H$, could be re-labelled ‘good’, while that of the low education individual would remain ‘moderate’.

Bago d’Uva et al.\(^{19}\) have used the vignette approach to measure health inequalities in China, Indonesia and India. They found significant reporting differences by education and income, in addition to age, gender and rural/urban residence. This led to overestimated inequalities by education but underestimated by income.

The current study tests for reporting differences by education and determines the impact of correcting for such differences on the measurement of health inequalities by education, using data on self-reported own health and vignette ratings for older individuals in eight European countries. In ageing developed countries, most morbidity and mortality occurs at higher ages and health inequalities among the older population have increasingly become the focus of attention.\(^6,7,32–34\)

**Methods**

**Study population**

The Survey of Health, Ageing and Retirement in Europe (SHARE) randomly sampled from the population aged 50 years and over (plus spouses) in 12 countries.\(^{35,36}\) The first wave of SHARE data were collected in 2004–05 and released in June 2007 (Release version 2.0). Supplementary samples containing the vignettes data are available for all but four countries. Probability samples have been carefully drawn in each country.\(^{37}\) Questionnaires were administered face-to-face by a computer-assisted personal interview to collect core information like age, gender and education, plus a self-completion drop-off part that covered self-assessments of health and vignette ratings.\(^{37}\) The overall response rate in the vignette samples was 57.7% (highest in France—77%—and lowest in Belgium—42%).\(^{38}\) The data were analysed separately for eight countries: Belgium, France, Germany, Greece, Italy, The Netherlands, Spain and Sweden.

**Measures**

Respondents were asked to classify their own health in six domains, in response to the questions: ‘Overall in the last 30 days, how much…’: ‘of a problem did you have with moving around?’ (mobility); ‘difficulty did you have with concentrating or remembering things?’ (cognition); ‘bodily aches or pains did you have?’ (pain); ‘difficulty did you have with sleeping such as falling asleep, waking up frequently during the night or waking up too early in the morning?’ (sleep); ‘of a problem did you have because of shortness of breath?’ (breathing); ‘of a problem did you have with feeling sad, low, or depressed?’ (emotional health). Domain selection was based on the World Health Survey (WHS), and guided by validity in terms of intuitive, clinical and epidemiological concepts of health; correspondence to the conceptual framework of the International Classification of Functioning, Disability and Health and comprehensiveness.\(^{39}\) Selection of SHARE health domains further took account of their contribution to self-assessed general health, which led to the inclusion of the additional domain of breathing.\(^{40}\) The response categories were: ‘None’, ‘Mild’, ‘Moderate’, ‘Severe’ and ‘Extreme’. In addition, for each domain, respondents were asked to evaluate three vignettes, each describing a fixed level of difficulty in that domain, on the same response scale. Descriptions of vignettes for all domains are given in the Supplementary data published online with this article.

As an indicator of SES, we use educational attainment based on the International Standard Classification of Education (ISCED 97): (i) finished at most primary education or first stage of basic education (ISCED 0–1); (ii) lower secondary or second stage of basic education (ISCED 2), (iii) upper secondary education (ISCED 3–4) and (iv) recognized third level education, which includes higher vocational education and university degree (ISCED 5–6).\(^{41}\) For Germany, the SHARE data do not distinguish between levels (i) and (ii). The lowest education group is used as the reference in the analysis. We control for age by means of a continuous variable. Because of relatively limited sample sizes, we opted to include a gender dummy variable, rather than estimating separate models for males and females.

**Data analyses**

For each country and domain, we first use an ordered probit model to estimate the association between self-reported health and educational level, controlling for sex and age but not for reporting differences. From these estimates, we compute the highest to lowest education group rate ratio for reporting no problem or difficulty in a given domain. We present these ratios for males at the sample mean age (64). We do not present results for females because we estimate a common model (controlling for sex), and so the rate ratios for males and females derive from the same education coefficients, exhibiting similar patterns. As robustness checks, we also computed rate ratios for reporting none or a mild problem or difficulty in a given domain and relative indices of inequality in the reporting of no problem or difficulty (which, contrary
Rate ratios corrected for reporting differences between socio-demographic groups were computed from a hierarchical ordered probit (HOPIT) model, in which the thresholds to report an extreme/severe/moderate/mild/no health problem are allowed to vary by age, sex and education level. The information to identify these socio-demographic varying reporting thresholds comes from the rating of the case vignettes. Having allowed for differences in reporting scales, we compute the corrected rate ratio from the education–health association while setting the response thresholds of the highest education group equal to those of men aged 64 with the lowest level of education. We test for differences in response scales by educational level using a log-likelihood ratio test for joint significance of the three education dummy variables in the four response thresholds.

All statistical analyses were performed with Stata 9.2, using the command `oprobit` for the ordered probit models and the code provided in Jones et al. for the HOPIT models. Rate ratios and their confidence intervals were obtained using the command `nlcom`.

**Results**

To illustrate reporting differences by education we present in Table 2, for The Netherlands and Sweden, the proportions classifying a specific vignette for each domain as no or mild difficulty. These particular vignettes are the ones that correspond to the middle level of difficulty, among the three levels represented by the vignettes for each domain. Except for cognition, the higher educated Dutch are much less likely than lower educated to designate the health problem described in the vignettes as one of no or mild difficulty. In general, the Swedes are less likely than the Dutch to report a given health state as representing no or mild difficulties and, more relevant for current purposes, they display less variability by education in the reporting of health. The tendency for the better educated to hold higher health standards is evident for only half of the domains.

There are differences in the response scales by education ($P < 0.05$) in 29 of the 48 domain/country combinations (Table 3). Differential reporting by education is most marked in the domain of mobility, for which differences exist in seven of the eight countries, and in France and The Netherlands, where differences emerge for all domains. On the other hand, Italy and Germany display differential reporting scales in only one and two domains, respectively. Additionally, we have tested whether the reporting of health by education differs between males and females by including gender-education interactions in all thresholds. Gender differences exist ($P < 0.05$) in only 8 out of 48 cases. Even in these cases, the impact of vignette adjustment on measured health inequalities is consistent with that obtained from the

---

**Table 1** Sample sizes by country, gender and education

<table>
<thead>
<tr>
<th></th>
<th>Belgium</th>
<th>France</th>
<th>Italy</th>
<th>Germany</th>
<th>Greece</th>
<th>Netherlands</th>
<th>Spain</th>
<th>Sweden</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$ (%)</td>
<td>$N$ (%)</td>
<td>$N$ (%)</td>
<td>$N$ (%)</td>
<td>$N$ (%)</td>
<td>$N$ (%)</td>
<td>$N$ (%)</td>
<td>$N$ (%)</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>55 (22.3)</td>
<td>145 (38.9)</td>
<td>98 (50.3)</td>
<td>– (–)</td>
<td>108 (32.7)</td>
<td>35 (13.8)</td>
<td>117 (60.6)</td>
<td>75 (38.3)</td>
</tr>
<tr>
<td>Lower secondary</td>
<td>63 (25.5)</td>
<td>20 (5.4)</td>
<td>44 (22.6)</td>
<td>16 (7.3)</td>
<td>34 (10.3)</td>
<td>94 (37.0)</td>
<td>36 (18.7)</td>
<td>34 (17.3)</td>
</tr>
<tr>
<td>Upper secondary</td>
<td>67 (27.1)</td>
<td>124 (33.2)</td>
<td>44 (22.6)</td>
<td>141 (61.4)</td>
<td>110 (33.3)</td>
<td>59 (23.2)</td>
<td>16 (8.3)</td>
<td>47 (24.0)</td>
</tr>
<tr>
<td>Tertiary</td>
<td>62 (25.1)</td>
<td>84 (22.5)</td>
<td>9 (4.6)</td>
<td>63 (28.6)</td>
<td>78 (23.6)</td>
<td>66 (26.0)</td>
<td>24 (12.4)</td>
<td>40 (20.4)</td>
</tr>
<tr>
<td>Total</td>
<td>247 (100.0)</td>
<td>373 (100.0)</td>
<td>195 (100.0)</td>
<td>220 (100.0)</td>
<td>330 (100.0)</td>
<td>254 (100.0)</td>
<td>193 (100.0)</td>
<td>196 (100.0)</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>81 (25.6)</td>
<td>241 (48.3)</td>
<td>156 (62.4)</td>
<td>– (–)</td>
<td>199 (51.3)</td>
<td>41 (14.7)</td>
<td>175 (64.8)</td>
<td>74 (33.9)</td>
</tr>
<tr>
<td>Lower secondary</td>
<td>83 (26.2)</td>
<td>45 (9.0)</td>
<td>41 (16.4)</td>
<td>73 (25.5)</td>
<td>35 (9.0)</td>
<td>129 (46.4)</td>
<td>57 (21.1)</td>
<td>41 (18.8)</td>
</tr>
<tr>
<td>Upper secondary</td>
<td>73 (23.0)</td>
<td>127 (25.5)</td>
<td>40 (16.0)</td>
<td>164 (57.3)</td>
<td>109 (28.1)</td>
<td>59 (21.2)</td>
<td>15 (5.6)</td>
<td>49 (22.5)</td>
</tr>
<tr>
<td>Tertiary</td>
<td>80 (25.2)</td>
<td>86 (17.2)</td>
<td>13 (5.2)</td>
<td>49 (17.1)</td>
<td>45 (11.6)</td>
<td>49 (17.6)</td>
<td>23 (8.5)</td>
<td>54 (24.8)</td>
</tr>
<tr>
<td>Total</td>
<td>317 (100.0)</td>
<td>499 (100.0)</td>
<td>250 (100.0)</td>
<td>286 (100.0)</td>
<td>388 (100.0)</td>
<td>278 (100.0)</td>
<td>270 (100.0)</td>
<td>218 (100.0)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>136 (24.1)</td>
<td>386 (44.3)</td>
<td>254 (57.1)</td>
<td>– (–)</td>
<td>307 (42.8)</td>
<td>76 (14.3)</td>
<td>292 (63.1)</td>
<td>149 (36.0)</td>
</tr>
<tr>
<td>Lower secondary</td>
<td>146 (25.9)</td>
<td>65 (7.5)</td>
<td>85 (19.1)</td>
<td>89 (17.6)</td>
<td>69 (9.6)</td>
<td>223 (41.9)</td>
<td>93 (20.1)</td>
<td>75 (18.1)</td>
</tr>
<tr>
<td>Upper secondary</td>
<td>140 (24.8)</td>
<td>251 (28.8)</td>
<td>84 (18.9)</td>
<td>305 (60.3)</td>
<td>219 (30.5)</td>
<td>118 (22.2)</td>
<td>31 (6.7)</td>
<td>96 (23.2)</td>
</tr>
<tr>
<td>Tertiary</td>
<td>142 (25.2)</td>
<td>170 (19.5)</td>
<td>22 (4.9)</td>
<td>112 (22.1)</td>
<td>123 (17.1)</td>
<td>115 (21.6)</td>
<td>47 (10.2)</td>
<td>94 (22.7)</td>
</tr>
<tr>
<td>Total</td>
<td>564 (100.0)</td>
<td>872 (100.0)</td>
<td>445 (100.0)</td>
<td>506 (100.0)</td>
<td>718 (100.0)</td>
<td>532 (100.0)</td>
<td>463 (100.0)</td>
<td>414 (100.0)</td>
</tr>
</tbody>
</table>
Before adjustment for reporting differences, high to low education rate ratios for reporting no problem or difficulty with own health are generally greater than 1, but their confidence intervals (CI) contain 1 in 32 of the 48 cases, including all domains in Sweden, The Netherlands and Belgium and all but pain in Germany (Table 4). Adjusting the ratios for reporting differences raises their magnitude in 39 of the 48 cases, and results in their CIs not containing 1 in 18 cases (Tables 4 and 5). The impact of adjustment is largest for Belgium, France, Germany and The Netherlands, where it increases the rate ratio in all domains and uncovers inequalities (lower bound of CI becomes larger than 1) in 3 (France) or 4 (Belgium, Germany and The Netherlands) domains. Spain and Sweden differ from the other countries in that the higher education group rates a given health state more positively in 3 and 4 domains, respectively, and consequently adjustment reduces the magnitude of the rate ratio. However, in the case of Spain, the adjustment is small and the CIs do not contain 1 neither before nor after adjustment and, in the case of Sweden, they contain 1 before and after the adjustment. The rate ratio for cognition falls slightly and loses significance (CI: 0.99–2.11) for Italy, but this is an artefact of the very small size of the highest education group. Rate ratios for the larger second and third education groups relative to the bottom do not display this pattern and, accordingly, the adjustment has a positive impact on the relative index of inequality. In all other cases that show negative impacts of the adjustment on the rate ratios, the same holds for relative indices of inequality. In terms of uncovering inequalities, the adjustment for differential reporting scales has the greatest impact in the domains of sleep and breathing.

**Discussion**

This study uses ratings of case vignettes to examine the extent to which the reporting of health in six domains differs by educational level and the degree to which any such differences bias the measurement of
health inequalities among the older population in eight European countries. We find clear evidence of reporting differences in most of the health domains and countries analysed. In general, more highly educated individuals are more critical of a given health state, although the opposite is true in Spain and Sweden for at least half of the health domains. In the remaining countries, failure to correct for differential reporting leads to underestimation of health inequalities by education. In particular for Belgium and The Netherlands, before correction, no inequalities by education in the probability of reporting no health problem or difficulty are present. But correction for differential reporting increases the ratios for all domains and reveals inequalities favouring the higher educated in four of the domains.

The study adds to the literature on SRH in several respects. First, ratings of vignettes provide a more direct test of differential reporting than is possible, for example, by comparing the degree to which SRH predicts mortality by SES. Second, the method not only provides a test for differential reporting, but also a means of correcting measured health inequalities for the phenomenon. Third, we are able to do this for a range of health domains, covering not only physical but also cognitive and emotional functioning. Differential reporting in the latter domains may not be reflected in the relationship between SRH and mortality. Fourth, comparable data from a cross-national survey allows examination of the extent to which the direction and the magnitude of differential reporting by SES varies across cultural settings. While we find that reporting differences generally lead to an

<table>
<thead>
<tr>
<th>Country</th>
<th>Pain</th>
<th>Sleep</th>
<th>Mobility</th>
<th>Emotional health</th>
<th>Cognition</th>
<th>Breathing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>Unadjusted</td>
<td>1.13 (0.78–1.48)</td>
<td>1.11 (0.83–1.39)</td>
<td>1.06 (0.86–1.25)</td>
<td>1.10 (0.90–1.31)</td>
<td>1.05 (0.75–1.35)</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>1.27 (0.89–1.66)</td>
<td>1.33 (1.04–1.63)</td>
<td>1.22 (1.02–1.42)</td>
<td>1.24 (1.03–1.45)</td>
<td>1.12 (0.76–1.48)</td>
</tr>
<tr>
<td>France</td>
<td>Unadjusted</td>
<td>1.39 (1.08–1.70)</td>
<td>1.05 (0.84–1.25)</td>
<td>1.21 (1.06–1.35)</td>
<td>1.05 (0.91–1.19)</td>
<td>1.40 (1.13–1.66)</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>1.54 (1.22–1.86)</td>
<td>1.32 (1.08–1.56)</td>
<td>1.33 (1.19–1.48)</td>
<td>1.23 (1.09–1.37)</td>
<td>1.61 (1.32–1.90)</td>
</tr>
<tr>
<td>Italy</td>
<td>Unadjusted</td>
<td>1.77 (0.98–2.57)</td>
<td>1.09 (0.66–1.52)</td>
<td>1.53 (1.24–1.82)</td>
<td>1.43 (1.06–1.81)</td>
<td>1.57 (1.03–2.11)</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>1.91 (1.06–2.75)</td>
<td>1.34 (0.88–1.80)</td>
<td>1.59 (1.32–1.85)</td>
<td>1.53 (1.19–1.88)</td>
<td>1.55 (0.99–2.11)</td>
</tr>
<tr>
<td>Belgium</td>
<td>Unadjusted</td>
<td>1.99 (1.14–2.84)</td>
<td>1.31 (0.94–1.68)</td>
<td>1.33 (0.95–1.71)</td>
<td>1.06 (0.81–1.32)</td>
<td>1.37 (0.92–1.81)</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>2.08 (1.25–2.91)</td>
<td>1.42 (1.03–1.80)</td>
<td>1.44 (1.04–1.84)</td>
<td>1.14 (0.85–1.42)</td>
<td>1.52 (1.01–2.04)</td>
</tr>
<tr>
<td>Italy</td>
<td>Unadjusted</td>
<td>1.23 (0.94–1.51)</td>
<td>1.29 (1.09–1.49)</td>
<td>1.03 (0.92–1.14)</td>
<td>1.34 (1.08–1.60)</td>
<td>1.37 (1.11–1.63)</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>1.45 (1.14–1.76)</td>
<td>1.34 (1.12–1.57)</td>
<td>1.04 (0.92–1.16)</td>
<td>1.33 (1.05–1.61)</td>
<td>1.54 (1.26–1.83)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>Unadjusted</td>
<td>1.11 (0.77–1.44)</td>
<td>0.99 (0.74–1.24)</td>
<td>0.99 (0.79–1.19)</td>
<td>1.10 (0.88–1.33)</td>
<td>1.32 (0.88–1.76)</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>1.35 (0.97–1.73)</td>
<td>1.31 (1.06–1.56)</td>
<td>1.18 (1.00–1.36)</td>
<td>1.38 (1.16–1.60)</td>
<td>1.39 (0.94–1.83)</td>
</tr>
<tr>
<td>Sweden</td>
<td>Unadjusted</td>
<td>0.95 (0.77–1.14)</td>
<td>1.09 (0.89–1.29)</td>
<td>1.32 (0.91–1.73)</td>
<td>1.38 (0.98–1.79)</td>
<td>1.19 (0.94–1.44)</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>1.02 (0.82–1.21)</td>
<td>1.07 (0.82–1.32)</td>
<td>1.08 (0.64–1.52)</td>
<td>1.27 (0.86–1.68)</td>
<td>1.03 (0.75–1.32)</td>
</tr>
<tr>
<td>Spain</td>
<td>Unadjusted</td>
<td>1.44 (1.07–1.80)</td>
<td>1.40 (1.11–1.69)</td>
<td>1.37 (1.12–1.62)</td>
<td>1.26 (1.07–1.45)</td>
<td>1.60 (1.29–1.90)</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>1.43 (1.02–1.84)</td>
<td>1.40 (1.06–1.75)</td>
<td>1.52 (1.25–1.79)</td>
<td>1.25 (1.02–1.47)</td>
<td>1.55 (1.21–1.89)</td>
</tr>
</tbody>
</table>

Probability of reporting no health problem or difficulty in each domain for the highest education group relative to the lowest. Unadjusted ratios estimated from the ordered probit model that imposes homogeneous reporting thresholds. Adjusted ratios estimated from the HOPIT model allowing reporting thresholds to vary by education, age and gender. All ratios computed for males at sample mean age (64).

This table provides a graphical summary of the results given in Table 3.  
▲—High to low education rate ratio increases with adjustment and CI contains 1 before and does not contain 1 after adjustment.  
▼—Rate ratio decreases with adjustment but CI contains 1 before and after adjustment.  
▼—Rate ratio increases with adjustment but CI contains 1 before and after adjustment.

health inequalities among the older population in eight European countries. We find clear evidence of reporting differences in most of the health domains and countries analysed. In general, more highly educated individuals are more critical of a given health state, although the opposite is true in Spain and Sweden for at least half of the health domains. In the remaining countries, failure to correct for differential reporting leads to underestimation of health inequalities by education. In particular for Belgium and The Netherlands, before correction, no inequalities by education in the probability of reporting no health problem or difficulty are present. But correction for differential reporting increases the ratios for all domains and reveals inequalities favouring the higher educated in four of the domains.

The study adds to the literature on SRH in several respects. First, ratings of vignettes provide a more direct test of differential reporting than is possible, for example, by comparing the degree to which SRH predicts mortality by SES. Second, the method not only provides a test for differential reporting, but also a means of correcting measured health inequalities for the phenomenon. Third, we are able to do this for a range of health domains, covering not only physical but also cognitive and emotional functioning. Differential reporting in the latter domains may not be reflected in the relationship between SRH and mortality. Fourth, comparable data from a cross-national survey allows examination of the extent to which the direction and the magnitude of differential reporting by SES varies across cultural settings. While we find that reporting differences generally lead to an
underestimation of socioeconomic inequalities in health, this is not always the case for Sweden and Spain and there is variation over the remaining countries in the degree of underestimation. Therefore, correction for systematic reporting biases may lead to different rankings of countries in international comparisons of health inequalities. For example, for the countries and domains analysed in this study, such correction moves Belgium and France mostly up and Spain and Sweden mostly down in the rankings.

One limitation of the study is that we do not allow for the possibility that differential reporting by education varies between males and females. With the relatively limited sample sizes available, separate estimation of the model by gender would have reduced the precision of the estimates. However, we control for sex throughout the analysis, allowing both the level of health and its reporting to differ by gender. We have also estimated models with interactions of gender and educational levels. Only in a few cases have we found differential health returns to education and/or differential reporting effects of education by gender \((P < 0.05)\), and even in these cases the gender-specific impacts of adjustment are still broadly consistent with the ones presented here.

A second limitation is that the analysis is restricted to the population aged 50 years and over. While health inequality among older individuals is of fundamental interest given the concentration of morbidity in this population, future studies should investigate whether correcting for reporting differences has the same impact on measured socioeconomic inequalities in health in younger populations. For older adults, education appears a more appropriate indicator of SES than income, but with a younger sample, differential reporting could be examined for alternative indicators of SES, like occupation and income.

Another potential limitation was the rather high non-response rate for most countries. It is not known whether the total non-response rates varied by education but we have examined item non-response to the health and vignette questions. Across domains, individuals with missing information on own health and vignettes tend to have lower levels of education \((P < 0.05)\) in Belgium, France, The Netherlands and Sweden. However, it is unlikely that this influences our results as the total number of observations with missing data on health or vignettes in all countries in the analysis does not exceed 147 (3%) in any health domain.

Using vignettes to standardize the reporting of health relies on the assumption that individuals rate the vignettes in the same way as they do their own health \((\text{response consistency})\). This is inherently difficult to test, but King \textit{et al.}\cite{King2015} and Van Soest \textit{et al.}\cite{VanSoest2016} provide some indirect evidence in favour of it by showing that the vignette adjustment brings self-reports of vision and drinking behaviour, respectively, closer to objective measures. Among other factors, the plausibility of the assumption depends upon the wording of the vignette descriptions. For example, the reference to obesity in one of the mobility vignettes may induce a bias that is not present in the reporting of own health. However, we found the results reported here to be robust to dropping this vignette, with the exception of there being less evidence of reporting differences by education in France and Spain although, even in these countries, the rate ratios after the vignette adjustment remained similar to those presented in Table 4. The vignette methodology further assumes that the level of health represented by each vignette is perceived by all respondents in the same way, apart from random measurement error \((\text{vignette equivalence})\). This is again difficult to test, but is supported by a high degree of consistency across individuals in the ranking of vignettes, suggesting that they are understood similarly across age and education groups.\cite{VanSoest2016}

A final, important limitation of the study is that it does not investigate reporting differences in the general SRH variable commonly used in the literature on socioeconomic inequalities in health, but in six more specific health domains. Reporting differences in general SRH may derive not only from the use of different scales to evaluate health within specific domains but also from differential weighting of the domains in the valuation of health overall.\cite{For example, one might expect individuals employed in manual occupations to place greater weight on mobility functioning, while those engaged in professional occupations may place relatively greater weight in cognitive functioning. An important topic for future research is the development of a methodology to aggregate vignette corrected reports of specific health domains into a measure of general health that takes account of differential weighting of the domains.

One plausible interpretation of our main finding is that less well-educated individuals have lower health standards and expectations because they observe more health problems in their surrounding reference groups and adjust their norms accordingly.\cite{But exposure to a higher prevalence of health problems need not reduce the tendency to report a given limitation. In a recent study on self-reported work disability, Van Soest \textit{et al.}\cite{VanSoest2016} find that higher prevalence of work disability in one’s reference group is associated with a greater likelihood of considering a given condition as work limiting. This is suggestive of peer effects reducing the threshold at which it is considered socially acceptable not to work because of a health problem. Besides having implications for the measurement of health inequality, differential reporting of health by SES may point to a potentially important cause of this inequality. If lower expectations for health result in the less well-educated being less likely to report a given health condition in a
survey, then it seems plausible that they will also be less inclined to take the preventive measures and the curative treatments that can protect and improve their health. Improved understanding of health reporting behaviour is, therefore, not only of interest from the rather technical perspective of measurement but also with respect to the substantive issue of identifying and rectifying the causes of socioeconomic inequality in health.

Acknowledgements
The Network for the Study of Pensions, Aging and Retirement (NETSPAR) theme ‘Health, income and work across the life cycle’ supports the research reported in this article. This article uses data from release 2 of SHARE 2004. The SHARE data collection has been primarily funded by the European Commission through the 5th framework programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life). Additional funding came from the US National Institute on Ageing (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OGHA 04-064). The Belgian Science Policy Office funded data collection in Belgium. Further support by the European Commission through the 6th framework programme (projects SHARE-I3, RII-CT-2006-062193, and COMPARE, CIT5-CT-2005-028857) is gratefully acknowledged. For methodological details see Börsch-Supan and Jürges. This article has benefitted from the comments received from three anonymous referees and the editor John Lynch.

Supplementary data
Supplementary data are available at IJE online.

Conflict of interest: None declared.

References

KEY MESSAGES
• Vignettes can be used to detect differences in the reporting of health by education.
• There is evidence of reporting differences by educational level for most health domains and countries analysed in this study.
• These reporting differences lead mostly to the underestimation of educational inequalities in health, except for Spain and Sweden.


