Is the story about worrying women and stoical men true? Gender differences in health after adjustment for reporting behavior

Anna Oksuzyan[1], Maciej J. Dańko[1], Jennifer Caputo[1], Domantas Jasilionis[1] and Vladimir M. Shkolnikov[1,2]

1 – Max Planck Institute for Demographic Research, Konrad-Zuse Str.1, 18057, Rostock, Germany

2 – National Research University Higher School of Economics, Street, Postal code, Moscow, Russian Federation

Corresponding author: Anna Oksuzyan

Tel.: + 49 3812081 178

Email: oksuzyan@demogr.mpg.de

Acknowledgments

This work was supported by Max Planck Society within the framework of the project ”On the edge of societies: New vulnerable populations, emerging challenges for social policies and future demands for social innovation. The experience of the Baltic Sea States (2016-2021)”. Vladimir M. Shkolnikov was partly supported by the Russian program of the state support of leading universities of the Russian Federation “5-100”.

This paper uses data from SHARE Waves 1 (DOIs: 10.6103/SHARE.w1.600), see Börsch-Supan et al for methodological details (Börsch-Supan et al. 2013). The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

Abstract

Research indicates that women have higher levels of physical disability and depression and lower scores on physical performance tests compared with men, while the evidence on gender differences in self-reported health is equivocal. Scholars note that these patterns may be contributed to by women over-reporting and men under-reporting health problems, but gender differences in reporting behaviors have not been rigorously tested. Using Wave 1 of the Survey of Health, Ageing and Retirement in Europe (SHARE), the present study investigates the extent to which adjusting for differences in reporting behavior modifies gender differences in global health. We also examine whether men and women's reporting behaviors are consistent across different levels of education. After adjusting for heterogeneity in reporting behavior, gender differences in both poor and good health widened. However, we found no clear gender-specific patterns in reporting either poor or good health. Our findings also do not provide convincing evidence that education is an important determinant of global health reporting, although the female disadvantage in poor health and the male advantage in good health were more apparent in lower than higher education groups at all ages. The results challenge prevailing stereotypes that women over-report and men under-report health problems and highlight the importance of attending to health problems reported by women and men with equal care.

Key words: gender, self-reported health, reporting behaviors, gender stereotypes

Background

A substantial body of research has shown that women are less healthy than men. Women fare worse on physical tests (Bohannon et al. 2006) and have higher levels of disability, functional limitations (Palacios-Ceña et al. 2012), and depression than their male counterparts (Oksuzyan et al. 2010; Salk et al. 2017). While gender differences in morbidity and self- reported health are less consistent and appear to vary across chronic conditions and national contexts (Oksuzyan et al. 2018). While women suffer more from non-acute disabling conditions, e.g. arthritis, men are more likely to have acute life-threatening conditions, e.g. myocardial infraction (Case and Paxson 2005; Crimmins et al. 2011). Women in most European countries (EU) and in the U.S. also report poorer health than men (Case and Paxson 2005; Bambra et al. 2009; Crimmins et al. 2011; Dahlin and Härkönen 2013). Scholars have assessed a variety of explanations for these gendered patterns in self-rated (SRH). Studies show that accounting for differences in men and women's socio-demographic characteristics, chronic conditions, and lifestyle behaviors substantially reduces gender gaps in SRH and may even reverse men's disadvantage in a few nations (Crimmins et al. 2011; Dahlin and Härkönen 2013). Because at the same level of morbidity women and men had similar ratings of health, scholars suggested that gender inequalities in SRH can be explained by the distribution of non-lethal disabling vs. acute life-threatening conditions among men and women (Case and Paxson 2005).

Other researchers suggested that the female disadvantage in SRH could be a result of gender inequalities in social roles with the expectation that gender differences in health would be smallest in more egalitarian countries. Although across EU countries the risk of reporting poor health among women was highest in Southern countries (Italy and Portugal) and no gender differences in SRH were found in Corporatist countries (Belgium, France, and Germany), which have traditional gender and family roles. In Social Democratic countries (Denmark, Sweden, Norway, and the Netherlands), which are widely considered as the most progressive countries in terms of gender equality, women had moderately high risk of reporting poor health (Bambra et al. 2009). A possible double burden on women who combine work and family in more egalitarian countries and fewer women who choose to combine these roles was suggested to explain these findings.

Another explanation for why women report poorer health than men is that they have greater somatic awareness and are more willing to admit health problems and seek medical advice (Benyamini et al. 2000; Courtenay 2000). Consistent with this possibility, Macintyre and colleagues (1996) found that the female disadvantage in health was limited to malaise symptoms, such as sleep problems, concentration difficulties, and worrying, while the female excess in morbidity was less apparent or even reversed for physical symptoms. An earlier review study suggested that women were more likely than men to report somatic symptoms whether they were related to medical conditions or medically unexplained (Barsky et al. 2001). Findings of strong associations between SRH and serious diseases in both genders, and between SRH and mild diseases among women only (Benyamini et al. 2000), further support the perspective that women consider a wider range of health dimensions than men when processing information for the assessment of global health. Also, findings that men with a recent history of transient ischemic attack, but not community-dwelling men, have more accurate self-reports of their global health than their female counterparts suggest that previous experiences, encountering a health problem, may have differential impact on men's and women's reporting behaviors (Dave et al. 2013; Dey et al. 2015).

The view that women's poorer performance on most health indicators is driven at least in part by their somatic awareness and willingness to report problems is further reinforced by studies which document consistent higher use of primary healthcare services and medications by women than men (Juel and Christensen 2008; Oksuzyan et al. 2011). A study covering the whole Danish population in 2005 showed that women were more frequent users of primary healthcare services than men at all ages, independent of reproductive healthcare visits (Juel and Christensen 2008). In contrast, the number of hospital admissions after age 50 was far higher in men than in women, and it was similar in the two genders before age 50 when sex-specific admissions were removed. As the authors noted, these patterns point toward delayed treatment-seeking behavior among men, and perhaps, greater female disadvantage for less serious health conditions and male disadvantage for more serious health problems requiring hospital admissions.

Other studies challenge gender stereotypes in the context of seeking medical advice and reporting health problems. Analysis of primary healthcare data in the UK revealed very small gender differences in the number of doctor consultations within 24 months prior to diagnosis of three cancers (Wang et al. 2014). Similarly, there was no strong support for the perspective that women were more likely to consult professionals for back pain than men with similar morbidity, although there was some evidence for more active help-seeking behavior for headache in women than in men (Hunt et al. 2011). A recent qualitative study also found that men and women who were interviewed after receiving a lung cancer diagnosis had similar interpretations and reactions of symptoms of the disease (MacLean et al. 2017).

In contrast to the prevailing assumption of over-reporting among women, MacIntyre and colleagues (1999) found no evidence of gender differences in the reporting of health problems, irrespective of their seriousness and type. Additionally, research showing no gender differences in subjective health at the same level of morbidity and in its predictive ability for hospitalizations runs counter to the hypothesis that women and men use different standards for assessing self- reported health (Case and Paxson 2005). To directly tackle the question of whether women over-report and men under-report health conditions, researchers have compared self-reports with clinical examinations and found mixed results. Some suggest higher accuracy of self-reports among women than among their male counterparts (Vargas et al. 1997; Dave et al. 2013), while others indicate the opposite (Short et al. 2009; Dey et al. 2015). However, we know relatively little about gender differences in the reporting of global health.

SRH is the most frequently used indicator of health in social, economic, and epidemiological research, as it is a strong predictor of mortality (Mossey and Shapiro 1982; Idler and Benyamini 1997) and easy to include in surveys. Jylhä (2009) developed a conceptual model for SRH to enhance researchers' understanding of the different factors that may influence an individual's perception of his or her health. According to this model, the evaluation of own health includes the review of information about biological health—e.g., medical diagnoses, functional status, symptoms and signs of illness—as well as lifestyle behaviors, and this evaluation influenced by contextual social and cultural factors. Among these factors are the use of various reference groups (e.g., peers vs. younger/older persons), earlier experiences (e.g., experiencing pain or being diagnosed with a condition), and cross-cultural differences in using scales and reporting patterns (e.g., linguistic differences in response options or gender differences) (Jylhä 2009). If contextual differences in the evaluations of SRH across cultures or socio-demographic groups are large, the comparison of SRH may lead to misleading results as the observed differences will reflect not only the variations of actual/true health across these groups but also the differences in reporting styles.

Various approaches have been used to account for reporting heterogeneity and to improve the comparability of self-reported health measures across socio-demographic and cultural groups (Salomon 2004; Jürges 2007; Layes et al. 2012; Schneider et al. 2012). Using anchoring vignettes Grol-Prokopczyk and colleagues (Grol-Prokopczyk et al. 2011) showed that the female participants of the Wisconsin Longitudinal Study had more optimistic reports of global health, after accounting for which the female advantage in SRH disappeared. However, in the Health and Retirement Study (HRS) men had more optimistic reporting in the domains of sleep, mobility, shortness of breath, and depression and more pessimistic assessments for pain and memory compared with women (Dowd and Todd 2011). To assess the reporting differences between Canadian men and women, Layes et al. (2012) considered a preference-standardized health-related quality of life to be a measure of latent true health, and its systematic deviation from SRH as a measure of reporting behavior. The authors found that men assessed their health to be significantly lower and women reported their health to be significantly higher relative to the "average Canadian", but the magnitude of these gender differences was very small (Layes et al. 2012).

Anchoring vignettes were also used to examine the differences in reporting behavior across educational groups (Bago d'Uva et al. 2008; Dowd and Todd 2011). In the Survey of Health, Ageing and Retirement in Europe (SHARE) the direction and significance of educational inequalities before and after accounting for reporting heterogeneity varied across six selected health domains -pain, sleep, mobility, emotional health, cognition, and breathing – and across countries (Bago d'Uva et al. 2008). For example, although Dutch persons with high education tended to assess their health more critically compared with their lower educated peers for most health domains except cognition, this pattern was not evident among Swedes. The educational differences in pain and memory domains remained almost unchanged before and after adjustment for reporting heterogeneity in the HRS sample as well, but the differences by education increased substantially for shortness of breath, depression, and mobility after the adjustment (Dowd and Todd 2011).

Two important requirements have to be fulfilled in order to use of anchoring vignettes: response consistency and vignette equivalence (Salomon 2004). Response consistency implies that an individual evaluates both specific health questions and related hypothetical scenario in the same way, while vignette equivalence requires that the underlying health level depicted in each vignette to be understood in the same way by all respondents independent of socio-demographic or other characteristics. Although earlier research studies found no major violations of response consistency and vignette equivalence (Bago d'Uva et al. 2008; Grol-Prokopczyk et al. 2011), recent studies, which used stricter statistical methods to test these two assumptions, provided a clear evidence that the respondents from different cultures[1] and socio-demographic groups in the HRS and the English Longitudinal Study of Aging appear to perceive vignette text at fundamentally different levels of health (Bago d'Uva et al. 2011; Grol-Prokopczyk et al. 2015). These findings suggest that health vignettes asked in these studies cannot be used to correct for reporting heterogeneity.

An alternative method is to use (semi-)objective information to adjust self-reported data (Jürges 2007; Schneider et al. 2012). For example, Jürges (2007) and Rebelo and Pereira (2014) showed that

[1] World Health Organization Study on Global AGEing and Adult Health includes six low- and middle-income countries – China, Ghana, India, Mexico, Russia, and South Africa, while the World Health Survey included four high-income countries - Brazil, France, Netherlands, and the United Kingdom.

based on self-reports the healthiest individuals in Europe reside in Denmark and Sweden and the least healthy live in Italy and Spain. When cross-cultural differences in reporting behavior were accounted for, cross-national variations in global health were reduced and the order of the countries from most to least healthy changed substantially.

The present study investigates the extent to which adjusting for differences in reporting behaviors modifies gender differences in health and whether these changes are due to men and women over- and/or under-reporting their health. Since most previous studies have focused on domain-specific measures of health rather than on the global health or only positive evaluations of health, we also examined whether gender differences in reporting patterns are similar for the two opposite evaluations of health: poor and good. Following Jylhä's conceptual model, we separate SRH into two components: true health and influences of contextual characteristics on the evaluation of health. Although the second component is likely to reflect differences across socio-demographic groups in both the assessment and reporting of SRH, for simplicity hereafter we will refer to this component as reporting behavior. Assuming that women and men differ both in their reporting styles and true health, we expect to find that an initial male advantage in SRH lessens or even disappears after adjusting for reporting behaviors. If our analyses show either that women over-report and men under-report poor health, or that women under-report while men over-report good health, stereotypical expectations about gender differences in reporting behaviors will be supported. We may thus observe a reduction of the gender differences in both poor and good self-reported health when accounting for these gendered reporting behaviors. On the other hand, if women under-report and men over-report poor health, or if women over-report and men under-report good health, this would challenge stereotypes about gendered reporting behaviors. Finally, it is possible that women and men do not differ much in their reporting styles, and that after adjusting for them the gender gaps in health will remain unchanged.

Prior work in the U.S. showed that gender differences in health were particularly pronounced in socially disadvantaged groups (Cooper 2002). However, in some EU countries the female disadvantage in SRH was highest among most educated groups (Bambra et al. 2009). A greater toll on health due to combining work and family among highly educated women was suggested to explain these seemingly contradictory findings. It is also possible that reporting behavior by education is different among men and women. Although the analysis of the HRS data revealed quite comparable reporting patterns by education in two genders, the German data suggest that reporting heterogeneity between men and women is driven in part by socioeconomic factors (Dowd and Todd 2011; Schneider et al. 2012).

Therefore, in this study we also explored whether the observed reporting behaviors of men and women are consistent among individuals with different levels of education.

Materials and methods

We used data collected during Wave 1 (2004) of the Survey of Health, Ageing and Retirement in Europe (SHARE), which consists of 11 EU countries – Austria, Belgium, Denmark, France, Germany, Greece, , Italy, Netherlands, Spain, Sweden, and Switzerland – and Israel (Börsch-Supan et al. 2013; Stuck et al. 2018). The SHARE is a multidisciplinary and cross-national panel survey of community-dwelling individuals aged 50 and older. The initial total sample of 29,373 individuals was reduced to by 6.9 percent to 27,345 individuals (54% women) as a result of excluding persons with missing data on SRH (n = 2028) and/or independent variables. The major source of missing data on independent variables was due to missing values on grip strength (n = 1707) and/or depression (n = 468). However, the sex-specific distributions of SRH were similar between those who had complete and missing data.

The global health question asked interviewees to evaluate their health in general ("Would you say your health is ...?") with 5 possible responses: very good, good, fair, bad, very bad[2]. The SHARE data collection instrument also included a wide range of other health measures, such as physician-diagnosed reported chronic conditions, Basic Activities of Daily Living (BADL), mobility limitations, Instrumental Activities of Daily Living (IADL), physical performance tests (handgrip strength and walking speed), mental health, anthropometric measurements, and lifestyle behaviors. Using these characteristics we estimated a continuous latent health measure.

As stated above, we apply the method that Jürges (2007) and Rebelo and Pereira (2014) used to explore cross-national differences in reporting general health in order to assess gender differences in health reporting. Their approach assumes that when responding to a survey question about their general health, participants assess their true health, which is measured on a continuous scale and is unobserved, and project it to a provided discrete scale. The thresholds that each individual uses to categorize their true health into a specific response option may be affected by the choice of a reference group, earlier health experiences, and cross-cultural differences in using scales and reporting patterns, and thus, may differ across individuals depending on their gender, age, cultural background, education, and personality traits, among other factors. Following this method we estimate a latent, underlying health for each individual and further reassign a particular response option of global health question. The latent,

[2] This is the EU version of response options. The US version has the following five possible responses: excellent, very good, good, fair, and poor. We replicated our analysis using also the US version (see Sensitivity analysis).

underlying health is estimated using a list of health related variables described in details below, and we allow the thresholds/cutpoints between adjacent responses of SRH to vary across individuals according to a combination of individual-level characteristics. As a result, we obtain a health measure, which is adjusted for heterogeneity in reporting behaviors and is more comparable across individuals.

To build the latent, underlying health variable, we first fit a generalized ordered probit model where the original SRH as a dependent variable is regressed against two sets of independent variables (Jürges 2007; Rebelo and Pereira 2014). A formal description and calculations is given in Appendix. The model estimates the latent health as a function of health characteristics, which constitute the first set of independent variables. These variables such as chronic conditions, participants' mobility level and difficulties with a range of daily activities, performance on grip strength test, anthropometric measures, and lifestyle behaviors (Table 1) that describe individuals' health in a more specific and objective way. The model includes also additional relationships to handle differences in how the continuous latent health is projected onto the five-category self-rated health by different individuals. These differences in reporting behavior are assumed to be determined by a number of socio-demographic characteristics (the second set of independent variables) such as gender, age group, education, and country. The predicted latent health values are further used to calculate disability weights for each health variable. The disability weight of the individual health variable is equal to the ratio of the difference between the corresponding health coefficient and the lowest value of the predicted latent health to the difference between the lowest and highest values of predicted latent health (eq. 5 in the Appendix). Disability weights provide information about the impact of a particular health condition on the latent health. They are used to create an individual health index (H_i). The H_i is a proxy for real underlying health, and varies from 0 representing the (model-based) worst health state to 1 representing the (model-based) worst health in the sample. The health index for individual i (H_i) is calculated by subtracting the sum of individual specific disability weights from 1, i.e. perfect health (eq. 6 in Appendix). While the disability weights have been computed for an average individual in the study population (i.e. perfect health is reduced to the same extent by the presence of heart attack or other heart problems for all individuals), the H_i index is considered to reflect underlying health and its variation across individuals.

Finally, health index is reclassified into a new five-category health measure that is adjusted for the inter-individual differences in reporting behavior. To do so, we follow the original distribution of SRH categories for the entire study population irrespective of other characteristics (Rebelo and Pereira 2014). That is, if 10% of the study participants assessed their health to be very bad, the value of health index which corresponded to the lowest 10% of the HI distribution, was considered as a cutoff level between the lowest two adjacent categories, i.e. “very bad” and “bad” (Supplementary Table 1 and Supplementary Figure 1). Although the original response frequencies are maintained, because of the adjustment for heterogeneity in reporting behavior between men and women, across age groups and countries, a study participant who initially perceived her/his health as very bad may actually be classified as having fair health.

All health determinants which comprise the first set of variables used to predict the latent health measure were coded as binary. Some variables identify the presence of the following physician-diagnosed reported chronic conditions (“Has a doctor ever told you that you had any of the conditions on this card?”): 1) heart attack or other heart problems, 2) hypertension or use of anti-hypertensive medications, 3) high blood cholesterol or use of statins, 4) stroke or other cerebrovascular diseases, 5) diabetes or use of antidiabetic medications, 6) chronic respiratory diseases including asthma, 7) musculoskeletal diseases, 8) cancer (including leukemia or lymphoma, but excluding minor skin cancers), 9) stomach, duodenal, or peptic ulcer, 10) Parkinson disease, 11) hip/femoral fracture, 12) other chronic conditions, and 13) the presence of two or more chronic diseases as a quasi-interaction. Other determinants reflected participants’ mobility level and difficulties with a range of activities: 14) poor mobility (three or more reported mobility limitations, i.e. the ability to engage in activities broadly ranging from sitting for about two hours to climbing several flights of stairs without resting), 15) Activities of Daily Living (ADL) disability (one or more reported ADL limitations, i.e. the ability to perform such simple tasks as dressing, including putting on shoes and socks or walking across a room), and 16) Instrumental Activities of Daily Living (IADL) disability (one or more reported IADL limitations, i.e. the ability to manage tasks necessary for fully independent life at own home). We also included 17) low grip strength (lowest sex-specific tertile (< 23 kg for women and < 39 kg for men) and 18) unable to perform grip strength test as objective measures of health. The three anthropometric measures based on self-reports in the analysis are 19) underweight (BMI≤20), 20) overweight (BMI>20 and BMI<25), 21) obese (BMI≥30). Finally, we accounted for mental health in identifying those with 22) a score of 4 and above on the EURO-D depressive symptoms scale, and a lifestyle behavior with 23) being a current smoker. Threshold variables include gender, education level (primary, i.e. ISCED level 2 or less, vs. secondary+, i.e. ISCED levels 3 and above), country (12 countries), and age groups (from 50 to 80+ in 10-year intervals).

Although in the initial model we selected all health variables suggested by (Jürges 2007) and (Rebelo and Pereira 2014), in the final model we included only those significantly related to the original SRH. Although walking speed was significant in the initial model, we excluded it from the final model due to large number of individuals with missing values. We also merged some of the health variables.

Specifically, we combined arthritis, including osteoarthritis and rheumatism, and osteoporosis into musculoskeletal problems, while asthma and chronic lung disease were combined into respiratory problems. We also repeated all analyses with a 3-level measure of education, i.e. low (ISCED level 2 or less), medium (ISCED levels 3 and 4), and high (ISCED level 5 or higher). Since the results were very similar in medium and high education groups, we have opted for more parsimonious models with 2-level education.

We developed an R code that realizes the SRH adjustment method introduced by Jurges (2007) and further developed by Rebelo and Pereira (2014). A full description of the method and calculations applied in this study is given in Appendix 1, and the R code is available upon request. We extend the original method by (Jürges 2007), and (Rebelo and Pereira 2014) by estimating confidence intervals (CIs) for adjusted health and the differences between the original and adjusted health based on bootstrap percentiles (Bradley and Tibshirani 1993; Lumley 2004). All analyses were performed in R (R Core team, 2018).

Results

Table 1 shows the distribution of health variables (health-related characteristics) and the threshold variables (socio-demographic characteristics) of the study population by gender, as well as the estimated disability weights for each health variable. The prevalence of mobility limitations, musculoskeletal problems, depression, IADL disabilities, and underweight were substantially greater among women, while having had a heart attack/another heart problem, stroke/other cerebrovascular diseases, ulcer, and being a current smoker or overweight was more likely for men. In terms of disability weights, Parkinson's disease, poor mobility, being unable to perform the grip strength test, and having a history of heart attack and stroke had the largest contribution to the reduction of health. The lowest contributions were from high total cholesterol, other than normal BMI (underweight, overweight, and obese), and being a current smoker. Women were likely to be in the lower education group than men, while the age distribution was similar across the two genders.

Although all models were fitted with 5-level SRH as a dependent variable, to facilitate the interpretation of the results we combined the response options very bad and bad (hereafter poor health) and the response option very good and good (hereafter good health). To determine whether the original and adjusted prevalences of poor/good health are different, we examined whether the original prevalences are within 95% confidence intervals (CIs) of the adjusted ones.

Self-reported vs. adjusted poor health by age and gender

The prevalence of poor SRH on the original scale increased with increasing age from 5.8% among 50-59 year old women to 21% among women aged 80+, and from 5.7% among 50-59 year old men to 15.7% among 80+ year old men (Figure 1, upper panels). Gender differences (the absolute difference between the prevalence of poor health among men and women) were small and not significant among persons in their 50s and 60s, although they increased in the older age groups (Figure 1, lower left panel).

When SRH was adjusted for differences in reporting styles, the proportion of the youngest group of women (age 50-59) with poor health declined slightly (Figure 1, upper left panel). However, the adjusted prevalences of poor health were actually higher than the original estimates among older women, and the magnitude of this change increased with advancing age (Figure 1, upper left panel). Similarly, among men the prevalence of poor health was lower for the two younger age groups (age 50-59 and 60-69) after adjusting for reporting styles, remained almost unchanged among men 70-79 year olds, and increased substantially among men age 80+ (Figure 1, upper right panel). After adjusting for differences in the reporting styles, gender differences in poor health widened across all age groups(Figure 1, lower left panel). At all ages the proportion of those with poor health was substantially higher among women than among men, and, as with the original scale, gender differences in the adjusted fractions increased with advancing age: from 2.15% among 50-59 year olds to 8.0% in the oldest age group.

To facilitate the interpretation of these results, we plotted the differences between the adjusted and original proportions of men and women with poor health by age. Figure 2 (lower left panel) shows that the adjusted prevalences of poor health are smaller than the original prevalences (i.e., the differences are negative) among 50-59 year-old women and among 50-59 and 60-69 year-old men, suggesting that they reported worse health than they actually had. The difference between adjusted and original prevalences of poor health among 70-79 year-old men was negligible, suggesting that their reporting of poor health is fairly accurate. By contrast, the adjusted prevalence of poor health is larger than on the original scale (i.e., the differences are positive) among women age 60+ and men age 80+, suggesting that both older women and the oldest men under-reported their poor health (Figure 2, lower left panel). These results also imply that, although the male advantage in general health increases after adjusting for reporting behaviors, there are no clear gender-specific patterns in reporting behaviors. Rather, the reporting styles appear to be age-dependent, with both younger men and women over-reporting poor health and older individuals under-reporting poor health.

Self-reported vs. adjusted good health by age and gender

The percentage of people reporting good health on the original scale was higher among men than in women at all ages (Figure 1). Gender differences increased from 7.1% among persons aged 50-59 to 14.4% in the oldest age group and were statistically significant at all ages. The percentage of persons with good SRH declined with increasing age from 72.2% among 50-59 year-old women to 38.8% among women aged 80+, and from 75.6% among 50-59 year-old men to 44.4% among the oldest men.

After controlling for reporting behaviors, the prevalence of good health increased slightly among youngest women (50-59 years) (Figure 1, upper left panel). In all other age groups the adjusted percentage of women with good health declined, and the difference between the original and adjusted proportions increased with advancing age (Figure 2, right panel and Figure 1, upper left panel). In comparison with the original proportions, the adjusted prevalence of good health among men increased among men age 50 to 69, remained almost unchanged among men aged 70-79, and declined in the oldest age group (Figure 2, right panel and Figure 1, upper left panel). The gender gaps in the adjusted proportions of good health favoring men widened at all ages, and the magnitude of gender differences increased with increasing age.

On Figure 2 (right panel) a positive difference between the adjusted and original prevalences of good health among youngest women suggests that they under-reported good health, while a negative difference in all other age groups indicates that more older women reported good health than were actually in good health. Figure 2 also indicates that fewer younger men (50-59 and 60-69 years) reported than were in good health, while men age 70-79 reported being in good health fairly accurately. However, as among women, more men at the oldest ages report being healthy than actually were. In short, these findings suggest that the youngest men and women in our sample under-report good health, while at the oldest ages both genders tend to over-report good health. Among men and women in their 60s and 70s, the patterns are more complex. Women in both age groups tend to over-report good health. Men at age 60-69 tend to under-report good health, while the reporting is rather accurate among 70-79 year-old men.

Self-reported vs. adjusted poor health by education

We also examined whether changes in the initial vs. adjusted proportions of poor/good health vary by education, as well as whether accounting for educational differences affects gender gaps in SRH. Prevalences of poor SRH on both the original and adjusted scales were consistently higher among low educated men than among their counterparts with secondary+ education (Figure 3, upper panels). The original proportions of women and men with low education who were in poor health were similar at ages 50-59, while at older ages low educated women tended to have higher prevalence of poorer health than their male counterparts (Figure 3, lower left panel). The proportion of poor health among women and men with secondary+ education was also similar in the 50-59 age group, but the prevalences of poor health were slightly lower among women than among men in the 60-69 and 70-79 age groups. In the oldest age group, women with secondary+ education had a higher prevalence of poor health than their male counterparts. These findings suggest that on the original scale, the female disadvantage in poor health is consistent across all ages in low educated groups, but the direction of gender differences in poor health varies across age groups among individuals with secondary+ education.

When differences in reporting behaviors were adjusted for, the female disadvantage in poor health became consistent across all ages and in both education groups, and the magnitude of gender gaps in the prevalence of poor health was greater in low education groups than among better educated groups (Figure 3, lower left panel).

The differences between the adjusted and original proportions of men and women with poor health by education are plotted on Figure 4 (left panels). After the adjustment for reporting behaviors, the patterns of changes in the prevalences of poor health were alike in the two education groups and similar to the general patterns indicated earlier (see section *Self-reported vs. adjusted poor health by age and gender*). That is, negative differences between the adjusted and original prevalences of poor health among women aged 50-59 and men aged 50-69 with low education suggest that they reported worse health than they have. In contrast, positive differences between the adjusted and original prevalences of poor health among women aged 60+ and among men aged 80+ with low education suggest that older women and the oldest men reported better health than they have. All of these patterns were also found for women and men with secondary+ education, and there was no clear pattern in the magnitude of the changes in the prevalence of poor health after the adjustment for reporting behaviors between the two education groups. Lastly, the differences between the adjusted and original prevalences of poor health among men 70-79 years old were very small (and statistically not significant in the low educated group), again suggesting that the reporting of poor health in this group is fairly accurate.

To summarize, our findings show that irrespective of education level, younger women (50-59 years) and men (50-69 years) over-report poor health, while women aged 70+ and men aged 80+ under-report poor health. Men at ages 70-79 report having poor health relatively accurately. Although our results suggested that there were no clear education-specific and gender-specific patterns in reporting behaviors, adjusting for the differences in health reporting widened gender differences in the prevalence of poor health in both education groups. The female disadvantage was more pronounced in the low education group than among better educated groups, and the adjustment for reporting behaviors revealed that the female disadvantage in poor health also appears among better educated persons across all ages.

Self-reported vs. adjusted good health by education

The proportions of individuals reporting good health on the original scale were consistently higher among men and women with secondary+ education than among their lower educated counterparts (Figure 3 upper panels). Men tended to have higher prevalences of good health than women in both education groups and across all ages, while the gender differences were higher among individuals with lower education than in those with secondary+ education (Figure 3, lower right panel).

After adjusting for reporting behaviors, gender differences in the proportions of those with good health widened. The male advantage in the prevalence of good health was larger in low education groups than in better educated groups (Figure 3, lower right panel).

To elucidate changes in the prevalences of good health after adjusting for reporting styles, we plotted the difference between the adjusted and original proportions of men and women with good health separately by education group. The patterns were similar in both education groups, and to those found for the analysis by age and gender only. Figure 4 (upper right panel) shows that irrespective of the education level the adjusted prevalences of good health are higher than the original prevalences among the youngest men and women, while the opposite is observed in the oldest age groups. The percentage of 60-69 and 70-79 year-old women with good health in both education groups declined after adjusting for reporting behavior. Among men, however, the adjusted prevalences were higher than the original estimates in both education groups. The only exception are the 70-79 year-old men with secondary+ education, for whom the adjustment for reporting styles resulted in a decline in the prevalence of good health.

To summarize, our findings suggest that regardless of education level younger men and women tend to under-report good health, while older individuals of both genders tend to over-report good health. We also found that the magnitude of changes in the prevalences of good health after the adjustment for reporting behaviors were not consistent in low vs. secondary+ education groups suggesting that education had little influence on women's and men's reporting behaviors.

Sensitivity analysis

Although it is well established that chronic conditions are important determinants of SRH (Jylhä et al. 1986; Singh-Manoux et al. 2006), less research has been done to investigate whether the effects of chronic conditions on SRH differ when considered as single vs as multimorbid conditions (Mavaddat et al. 2014). Mavaddat et al. (2014) revealed that the odds or reporting poor health increased with number of chronic conditions, and this association was stronger for men than for women. To test whether including physician-diagnosed reported chronic conditions as multimorbid conditions rather than single conditions has an impact on the estimation of disability weights and adjusted proportions of poor and good health, we performed additional analyses by fitting a model where specific chronic conditions were replaced with a categorical variable indicating the number of chronic diseases from 0 to 9+. All gender- and age-specific patterns were similar to those found for the original models (results available on request).

Another potential concern is related to differences in the assessment of general health created by variations in the response options. Since the SHARE questionnaire included another version of general health question with response options commonly used in the US – excellent, very good, good, fair, poor – we replicated our analysis with the second version of response options on general health question. Although some small differences were observed, the overall results by gender and education levels were very similar and our conclusions remained the same (results available on request).

Discussion

Although previous works have suggested that the female disadvantage in SRH can be partially explained by women's greater attention to bodily symptoms and their willingness to report health problems, research rigorously testing this assumption is limited. This study set out to examine whether adjusting for differences in reporting behaviors modifies gender differences in health, and whether these changes are due to men and women over- and/or under-reporting their health.

We found small gender differences in the prevalence of poor health on the original scale among persons at aged 50-69, and a consistent male advantage in the prevalence of both poor and good health among the older age groups. Additionally, after adjusting for differences in men and women's reporting behaviors, gender differences in both poor and good health widened. Contrary to widespread assumptions about gender-stereotypical reporting behaviors which suggest that women over-report poor health and men over-report good health, we found no clear evidence for gender-specific patterns in reporting of either poor or good health. Rather, reporting styles varied greatly by age: younger (50-59 and 60-69 year old) women and men in our study population tended to over-report poor health and under-report good health, while the oldest women and men tended to under-report poor health and over-report good health. Men in their seventies had fairly accurate reporting of both poor and good health. Although gender differences in poor and good health became more apparent in both education groups after adjusting for reporting behaviors, there was no consistent pattern in over- or under-reporting of health across education groups and in the magnitude of the female disadvantage in health. These findings suggest that reporting behaviors were similar among men and women with lower and secondary+ education.

Our results are consistent with previous research which showed that being 80 years of age and older was the strongest determinant of an optimistic health evaluation, i.e., above the average Canadian (Layes et al. 2012). Although men evaluated their health more pessimistically in this study (lower relative to the average Canadian) and women assessed their health more optimistically, these gender differences were not clinically meaningful. Our findings of no clear gender patterns in reporting behaviors partially agree with previous research using the Wisconsin Longitudinal Study (Grol-Prokopczyk et al. 2011) and the Health and Retirement Survey (Dowd and Todd 2011). Like in these studies, the reporting patterns we found are not in line with dominating assumptions that women tend to over-report and men tend to "hide" their health problems, although adjusting for heterogeneity in reporting behaviors results in larger gender differences in health. In the Wisconsin Longitudinal Study, women provided more optimistic evaluations for some vignettes, whereas no gender differences in health ratings or more negative evaluations among women were found for other vignettes (Grol-Prokopczyk et al. 2011). Using the HRS, Dowd and Todd revealed mixed patterns for gender differences in health depending on the selected health domain (Dowd and Todd 2011). Although men tended to provide relatively optimistic evaluations of their health, adjusting for reporting heterogeneity had no substantive impact on gender differences in mobility, diminished female disadvantage in depression and sleep, and increased male disadvantage in shortness of breath. The use of domain-specific rather than global measures of health in the analysis by Dowd and Todd (Dowd and Todd 2011) may the inconsistencies between these findings and our own.

Research comparing self-reported information with administrative or biomedical data provides inconsistent evidence regarding gender differences in the accuracy of reporting of health related information. Some studies found that men were more likely to accurately report hypertension, hypercholesterolemia, and diabetes (Dey et al. 2015), as well as the number of doctor visits, absenteeism, and cardiovascular risk factors than women (Short et al. 2009). Others revealed higher odds of over-reporting of chronic diseases or even greater validity of self-reported hypertension among women than among men (Vargas et al. 1997; Dave et al. 2013; Galenkamp et al. 2014). Since under-reporting of healthcare use worsens with its intensity (Bhandari and Wagner 2006; Oksuzyan et al. 2009), and women are more frequent users than men (Juel and Christensen 2008; Oksuzyan et al. 2011), the use of different health measures, such as the number of doctor visits and medications use vs. specific health conditions, may partially explain these contradictory findings surrounding gender differences in the accuracy of self-reports. Some of the discrepant findings can also be attributable to differences in the study populations, i.e., patient vs. community-based samples, and/or previous experiences of health problem.

In line with prior work we find that education inequalities in health widen after adjusting for heterogeneity in reporting behaviors. However, our findings indicate that reporting patterns do not systematically differ among persons with lower and secondary+ education and that gender gaps in health are similar across the two education groups. Although a previous study based on the SHARE data showed that correcting for these differences generally widened educational inequalities in health, this enlargement was attributable to more negative ratings of a given health state by better educated individuals (Bago d'Uva et al. 2008). Possible explanations for these inconsistent findings may underlie in the methodological differences between this study and our own our, i.e. the use of anchoring vignettes vs. other (semi-)objective health measures to adjust for reporting heterogeneity and domain-specific measures of health vs global health.

Previous work regarding educational differences in reporting behaviors is contradictory. A study based on a Canadian community sample showed that higher socioeconomic groups assessed their health more pessimistically than their peers from low socioeconomic groups in, but the differences by education were not clinically meaningful (Layes et al. 2012). In other North American-based studies, the accuracy of reporting of cardiovascular risk factors, such as hypertension or hypercholesterolemia, was similar among people with different levels of educational attainment (Vargas et al. 1997; Dey et al. 2015). There is, however, some evidence that the concordance between self-reports and administrative data regarding health care utilization and absenteeism was higher among persons with an advanced educational degree than those with less education (Short et al. 2009).

A strength of the present study is that the measure of health considered in this study takes into account the multi-dimensionality of global health by using a wide range of health-related characteristics to better get at the level of "true" health. However, a substantial part of the health characteristics used in the present study are self-reports, and research evidence regarding the accuracy of self-reported chronic conditions across socio-demographic groups is conflicting (Kriegsman et al. 1996; Galenkamp et al. 2014), height, weight, BMI (Villanueva 2001; Gorber et al. 2007; Stommel and Schoenborn 2009), and smoking (Newell et al. 1999). Further, some individuals may be unaware of a serious health problem they suffer from, e.g. silent hypertension, myocardial infraction, and ischemic stroke, which could bias our estimates of disability weights. Although the prevalence of these silent chronic conditions varies across study populations depending on the patients' ages and the method used to detect the condition, in a general population the prevalence appear to be rather small, up to 5%, with no clear sex-specific pattern (Das et al. 2008; Valensi et al. 2011). Another study limitation may underlie in the methodology we applied to adjust for reporting heterogeneity. Despite being comprehensive, this method may not be able to eliminate entirely the heterogeneity in reporting behavior. In addition, the set of threshold variables may be incomplete. It does not include, for example, profession/occupation, partnership status, religion, personality characteristics and other features that may potentially influence reporting of health. Considering above-mentioned limitations, our future studies will focus on investigating how adjusting for reporting styles modifies gender differences in general health using other survey datasets where biomarker data are available. Since gender-related social norms may vary across European countries, it is also possible that gender differences in reporting behaviors differ across the SHARE countries, which is a topic for future investigation.

The present study adds to the existing literature on gender differences in reporting styles by assessing whether adjusting for reporting styles modifies gender differences in SRH. Our findings challenge prevailing gender stereotypes that women over-report and men under-report health problems. They also highlight the importance of attending to the health problems reported by women and men equally carefully, which is particularly salient for clinical settings and may help avoid delayed diagnosis and treatment of health problems more commonly seen in the opposite gender. Consolidating and extending similar observations in the UK (Hunt et al. 2011; MacLean et al. 2017), we argue that both research and medical communities should abandon traditional views that women report more readily poor health and men more likely to over-report good health and that we should regard a malaise symptom, such as general weakness, as it can be a sign of a slowly progressing potentially lethal disease equally likely in men and women. These steps are needed to advance our understanding of gender differences in health and gender-specific barriers, if any, to use healthcare services timely and appropriately.

References

Bago d'Uva T, Lindeboom M, O'Donnell O, van Doorslaer E (2011) Slipping Anchor? Testing the Vignettes Approach to Identification and Correction of Reporting Heterogeneity. J Hum Resour 46:875–906. doi: 10.2139/ssrn.1572319

Bago d'Uva T, O'Donnell O, Van doorslaer E (2008) Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. Int J Epidemiol 37:1375–1383. doi: 10.1093/ije/dyn146

Bambra C, Pope D, Swami V, et al (2009) Gender, health inequalities and welfare state regimes: A cross-national study of 13 European countries. J Epidemiol Community Health 63:38–44. doi: 10.1136/jech.2007.070292

Barsky AJ, Peekna HM, Borus JF (2001) Somatic symptom reporting in women and men. J Gen Intern Med 16:266–275. doi: http://dx.doi.org/10.1046/j.1525-1497.2001.016004266.x

Benyamini Y, Leventhal EA, Leventhal H (2000) Gender Differences in Processing Information for Making Self-Assessments of Health. Psychosom Med 62:354–364

Bhandari A, Wagner T (2006) Self-reported utilization of health care services: Improving measurement and accuracy. Med Care Res Rev 63:217–235. doi: 10.1177/1077558705285298

Bohannon RW, Peolsson A, Massy-Westropp N, et al (2006) Reference values for adult grip strength measured with a Jamar dynamometer: a descriptive meta-analysis. Physiotherapy 92:11–15. doi: 10.1016/j.physio.2005.05.003

Börsch-Supan A, Brandt M, Hunkler C, et al (2013) Data resource profile: The survey of health, ageing and retirement in europe (share). Int J Epidemiol 42:992–1001. doi: 10.1093/ije/dyt088

Bradley E, Tibshirani RJ (1993) An Introduction to the Bootstrap, 1st edn. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Washington D.C.

Case A, Paxson C (2005) Sex Differences in Morbidity and Mortality. Demography 42:189–214. doi: 10.2307/4147343

Cooper H (2002) Investigating socio-economic explanations for gender and ethnic inequalities in health. Soc Sci Med 54:693–706. doi: 10.1016/S0277-9536(01)00118-6

Courtenay WH (2000) Constructions of masculinity and their influence on men's well-being: A theory of gender and health. Soc Sci Med 50:1385–1401. doi: 10.1016/S0277-9536(99)00390-1

Crimmins EM, Kim JK, Solé-Auró A (2011) Gender differences in health: Results from SHARE, ELSA and HRS. Eur J Public Health 21:81–91. doi: 10.1093/eurpub/ckq022

Dahlin J, Härkönen J (2013) Cross-national differences in the gender gap in subjective health in Europe:

Does country-level gender equality matter? Soc Sci Med 98:24–28. doi: 10.1016/j.socscimed.2013.08.028

Das RR, Seshadri S, Beiser AS, et al (2008) Prevalence and correlates of silent cerebral infarcts in the framingham offspring study. Stroke 39:2929–2935. doi: 10.1161/STROKEAHA.108.516575

Dave GJ, Bibeau DL, Schulz MR, et al (2013) Predictors of congruency between self-reported hypertension status and measured blood pressure in the stroke belt. J Am Soc Hypertens 7:370–378. doi: 10.1016/j.jash.2013.04.007

Dey AK, Alyass A, Muir RT, et al (2015) Validity of Self-Report of Cardiovascular Risk Factors in a Population at High Risk for Stroke. J Stroke Cerebrovasc Dis 24:2860–2865. doi: 10.1016/j.jstrokecerebrovasdis.2015.08.022

Dowd JB, Todd M (2011) Does Self-reported Health Bias the Measurement of Health Inequalities in U . S . Adults ? Evidence Using Anchoring Vignettes From the Health and Retirement Study. J Gerontol 66:478–489. doi: 10.1093/geronb/gbr050.

Galenkamp H, Huisman M, Braam AW, et al (2014) Disease prevalence based on older people's self-reports increased, but patient-general practitioner agreement remained stable, 1992-2009. J Clin Epidemiol 67:773–780. doi: 10.1016/j.jclinepi.2014.02.002

Gorber SC, Tremblay M, Moher D, Gorber B (2007) A comparison of direct vs. self-report measures for assessing height, weight and body mass index: A systematic review. Obes Rev 8:307–326. doi: 10.1111/j.1467-789X.2007.00347.x

Grol-Prokopczyk H, Freese J, Hauser RM (2011) Using anchoring vignettes to assess group differences in general self-rated health. J Health Soc Behav 52:246–261. doi: 10.1177/0022146510396713

Grol-Prokopczyk H, Verdes-Tennant E, McEniry M, Ispány M (2015) Promises and Pitfalls of Anchoring Vignettes in Health Survey Research. Demography 52:1703–1728. doi: 10.1007/s13524-015-0422-1

Hunt K, Adamson J, Hewitt C, Nazareth I (2011) Do women consult more than men? A review of gender and consultation for back pain and headache. J Heal Serv Res Policy 16:108–117. doi: 10.1258/jhsrp.2010.009131

Idler EL, Benyamini Y (1997) Self-Rated Health and Mortality : A Review of Twenty-Seven Community Studies *. Heal (San Fr 38:21–37

Juel K, Christensen K (2008) Are men seeking medical advice too late? Contacts to general practitioners and hospital admissions in Denmark 2005. J Public Health (Bangkok) 30:111–113. doi: 10.1093/pubmed/fdm072

Jürges H (2007) True health vs response styles: exploring cross-country differences in self-reported health: TRUE HEALTH VS RESPONSE STYLES. Health Econ 16:163–178. doi: 10.1002/hec.1134

Jylhä M (2009) What is self-rated health and why does it predict mortality? Towards a unified conceptual model. Soc Sci Med 69:307–316. doi: 10.1016/j.socscimed.2009.05.013

Jylhä M, Leskinen E, Alanen E, et al (1986) Self-rated health and associated factors among men of different ages. J Gerontol 41:710–717

Kriegsman DMW, Penninx BWJH, Van Eijk JTM, et al (1996) Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly. A study on the accuracy of patients' self-reports and on determinants of inaccuracy. J Clin Epidemiol 49:1407–1417. doi: 10.1016/S0895-4356(96)00274-0

Layes A, Asada Y, Kepart G (2012) Whiners and deniers - What does self-rated health measure? Soc Sci Med 75:1–9. doi: 10.1016/j.socscimed.2011.10.030

Lumley T (2004) Analysis of complex survey samples. 1–19

MacIntyre S, Ford G, Hunt K (1999) Do women "over-report" morbidity? Men's and women's responses to structured prompting on a standard question on long standing illness. Soc Sci Med 48:89–98. doi: 10.1016/S0277-9536(98)00292-5

Macintyre S, Hunt K, Sweeting H (1996) Gender differences in health: Are things really as simple as they seem? Soc Sci Med 42:617–624. doi: 10.1016/0277-9536(95)00335-5

MacLean A, Hunt K, Smith S, Wyke S (2017) Does gender matter? An analysis of men's and women's accounts of responding to symptoms of lung cancer. Soc Sci Med 191:134–142. doi: 10.1016/j.socscimed.2017.09.015

Mavaddat N, Valderas JM, Van Der Linde R, et al (2014) Association of self-rated health with multimorbidity, chronic disease and psychosocial factors in a large middle-aged and older cohort from general practice: A cross-sectional study. BMC Fam Pract 15:1–11. doi: 10.1186/s12875-014-0185-6

Mossey JM, Shapiro E (1982) Self-rated health: a predictor of mortality among the elderly. Am J Public Heal 72:800–808

Newell SA, Girgis A, Sanson-Fisher RW, Savolainen NJ (1999) The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population11The full text of this article is available via AJPM Online at http://www.elsevier.com/locate/ajpmonline. Am J Prev Med 17:211–229. doi: 10.1016/S0749-3797(99)00069-0

Oksuzyan A, Crimmins E, Saito Y, et al (2010) Cross-national comparison of sex differences in health and mortality in Denmark, Japan and the US. Eur J Epidemiol 25:471–480. doi: 10.1007/s10654-010-9460-6

Oksuzyan A, Gumà J, Doblhammer G (2018) Sex Differences in Health and Survival. In: Doblhammer, Gabriele; Gumà J (ed) A Demographic Perspective on Gender, Family and Health in Europe. Springer International Publishing, pp 65–100

Oksuzyan A, Jacobsen R, Glaser K, et al (2011) Sex differences in medication and primary healthcare use before and after spousal bereavement at older ages in Denmark: Nationwide register study of over 6000 bereavements. J Aging Res 2011:. doi: 10.4061/2011/678289

Oksuzyan A, Petersen I, Stovring H, et al (2009) The Male-Female Health-Survival Paradox: A Survey and Register Study of the Impact of Sex-Specific Selection and Information Bias. Ann Epidemiol 19:. doi: 10.1016/j.annepidem.2009.03.014

Palacios-Ceña D, Jiménez-García R, Hernández-Barrera V, et al (2012) Has the prevalence of disability increased over the past decade (2000-2007) in elderly people? A spanish population-based survey. J Am Med Dir Assoc 13:136–142. doi: 10.1016/j.jamda.2010.05.007

Rebelo LP, Pereira NS (2014) Assessing Health Endowment, Access and Choice Determinants: Impact on Retired Europeans' (In)activity and Quality of Life. Soc Indic Res 119:1411–1446. doi: 10.1007/s11205-013-0542-1

Salk RH, Hyde JS, Abramson LY (2017) Psychological Bulletin Gender Differences in Depression in Representative National Samples: Meta-Analyses of Diagnoses and Symptoms Gender Differences in Depression in Representative National Samples: Meta-Analyses of Diagnoses and Symptoms. 143:783–822. doi: 10.1037/bul0000102

Salomon JA (2004) Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes. Bmj 328:258–0. doi: 10.1136/bmj.37963.691632.44

Schneider U, Pfarr C, Schneider BS, Ulrich V (2012) I feel good! Gender differences and reporting heterogeneity in self-assessed health. Eur J Heal Econ 13:251–265. doi: 10.1007/s10198-011-0301-7

Short ME, Goetzel RZ, Pei X, et al (2009) How accurate are self-reports? Analysis of self-reported health care utilization and absence when compared with administrative data. J Occup Environ Med 51:786–796. doi: 10.1097/JOM.0b013e3181a86671

Singh-Manoux A, Martikainen P, Ferrie J, et al (2006) What does self rated health measure? Results from the British Whitehall II and French Gazel cohort studies. J Epidemiol Community Health 60:364–372. doi: 10.1136/jech.2005.039883

Stommel M, Schoenborn CA (2009) Accuracy and usefulness of BMI measures based on self-reported weight and height: Findings from the NHANES & NHIS 2001-2006. BMC Public Health 9:1–10. doi: 10.1186/1471-2458-9-421

Stuck S, Zuber S, Kotte M, et al (2018) Survey of health, Ageing and Retirement in Europe. Release Guide 6.1.0. 1–83

Valensi P, Lorgis L, Cottin Y (2011) Prevalence, incidence, predictive factors and prognosis of silent myocardial infarction: A review of the literature. Arch Cardiovasc Dis 104:178–188. doi: 10.1016/j.acvd.2010.11.013

Vargas CM, Burt VL, Gillum RF, Pamuk ER (1997) Validity of Self-Reported Hypertension in the National Health and Nutrition Examination Survey III , 1988 – 1991. Prev Med (Baltim) 26:678–685. doi: 10.1006/pmed.1997.0190

Villanueva E V (2001) The validity of self-reported weight in US adults: a population based cross-sectional study. BMC Public Health 1:11. doi: 10.1186/1471-2458-1-11

Wang Y, Freemantle N, Nazareth I, Hunt K (2014) Gender differences in survival and the use of primary care prior to diagnosis of three cancers: An analysis of routinely collected UK general practice data. PLoS One 9:. doi: 10.1371/journal.pone.0101562

Table 1. Distribution of sociodemographic and health characteristics and estimated disability weights in the study population by gender

	Women (n=14,615)	Men (n=12,566)	Total (n=27,181)	Disability weight
Health variables (%)				
Parkinson's disease	0.5	0.6	0.5	0.165
Poor mobility	27.6	14.5	21.5	0.118
Grip test unable	2.6	1.7	2.2	0.095
Respiratory diseases	8.1	8.9	8.4	0.094
Heart attack/other hearts problems	9.5	15.9	12.4	0.093
Stroke/other cerebral vascular disease	3.0	4.0	3.5	0.089
Cancer	5.9	4.8	5.4	0.085
ADL	10.0	7.9	9.1	0.084
Other diseases	17.9	15.8	16.9	0.080
Diabetes	10.0	11.5	10.7	0.068
Musculoskeletal diseases	31.4	14.2	23.4	0.067
Depression	46.3	27.4	37.6	0.063
Underweight	1.7	0.4	1.1	0.055
Hypertension	37.4	34.6	36.1	0.053
IADL	19.2	11.1	15.5	0.049
Stomach, duodenal or peptic ulcer	5.4	6.8	6.0	0.038
Low grip strength	35.2	34.9	35.0	0.028
Current smoker	15.4	23.5	19.1	0.026
Hip or femoral fracture	2.1	1.7	1.93	0.024
Chronic diseases 2+	44.5	38.9	41.9	0.024
Obese	18.3	16.2	17.3	0.020
High blood cholesterol	24.9	25.6	25.2	0.019
Overweight	36.6	50.0	42.8	0.017
Threshold variables (%)				
Education				
Low	56.3	45.5	51.3	
Secondary+	43.7	54.5	48.7	
Age groups				
50-59	37.5	36.5	37.0	
60-69	31.3	33.4	32.3	
70-79	21.5	22.3	21.9	
80+	9.7	7.9	8.9	
Country				

Figure 1. Original (O) and adjusted (A) prevalences of poor and good health among women and men by age

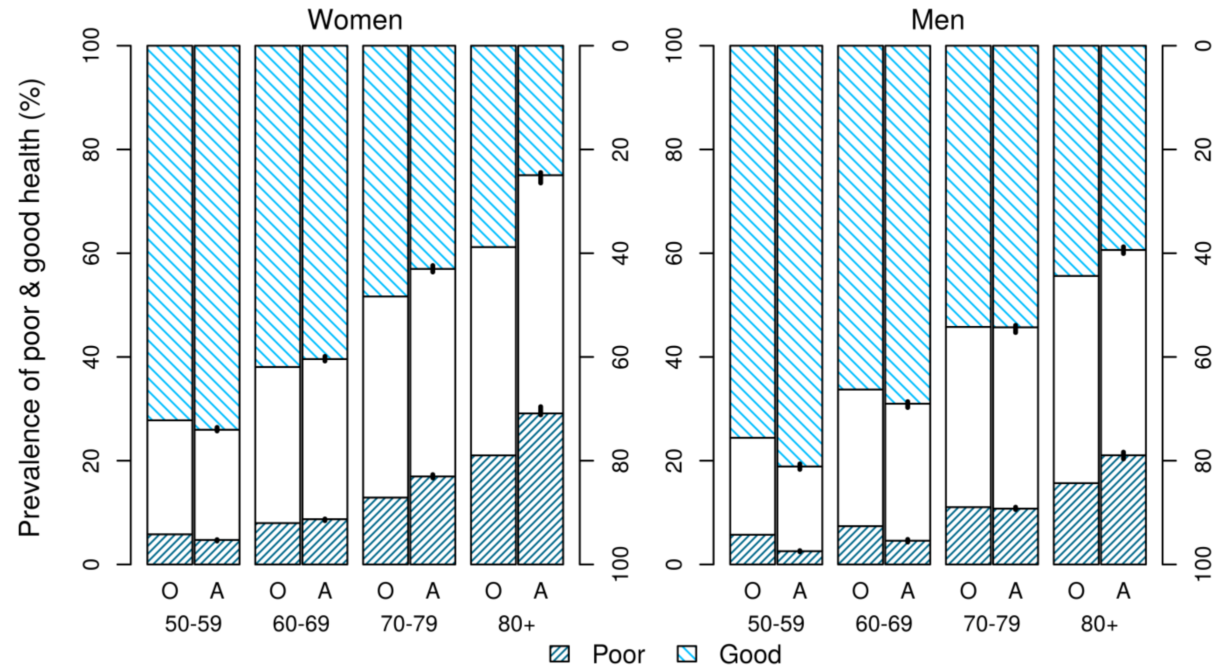


Figure 2. Gender (male – female) differences in the prevalences of poor and good health by age

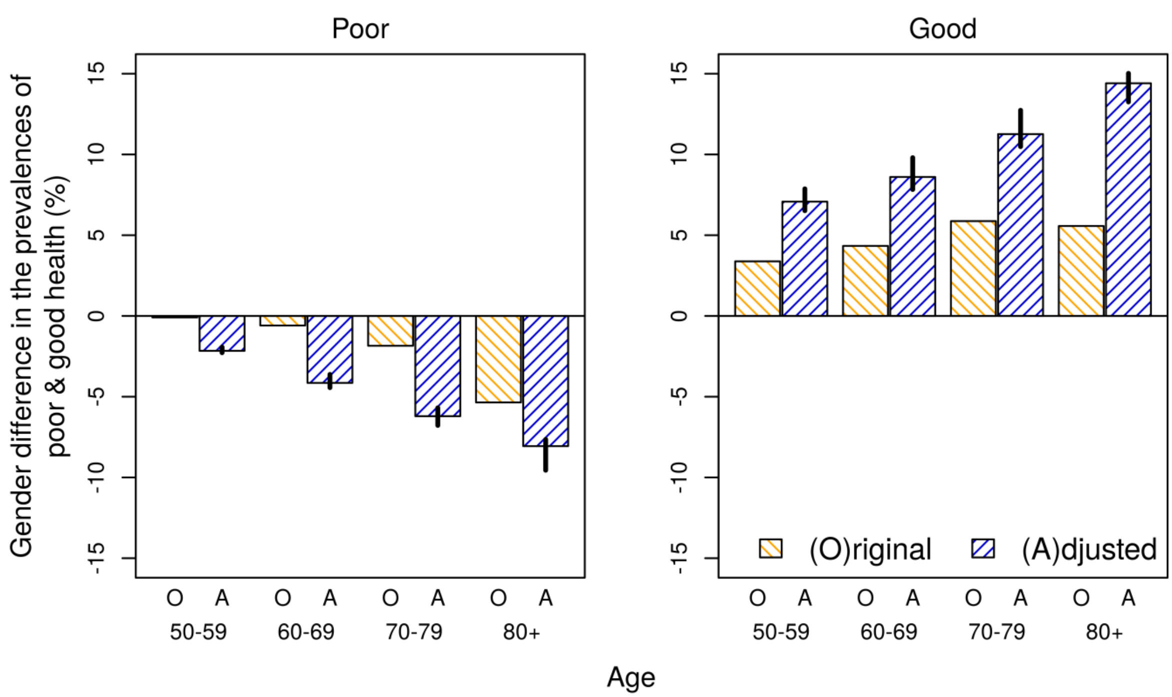


Figure 3. Differences between the adjusted and original prevalences of poor and good health by gender and age

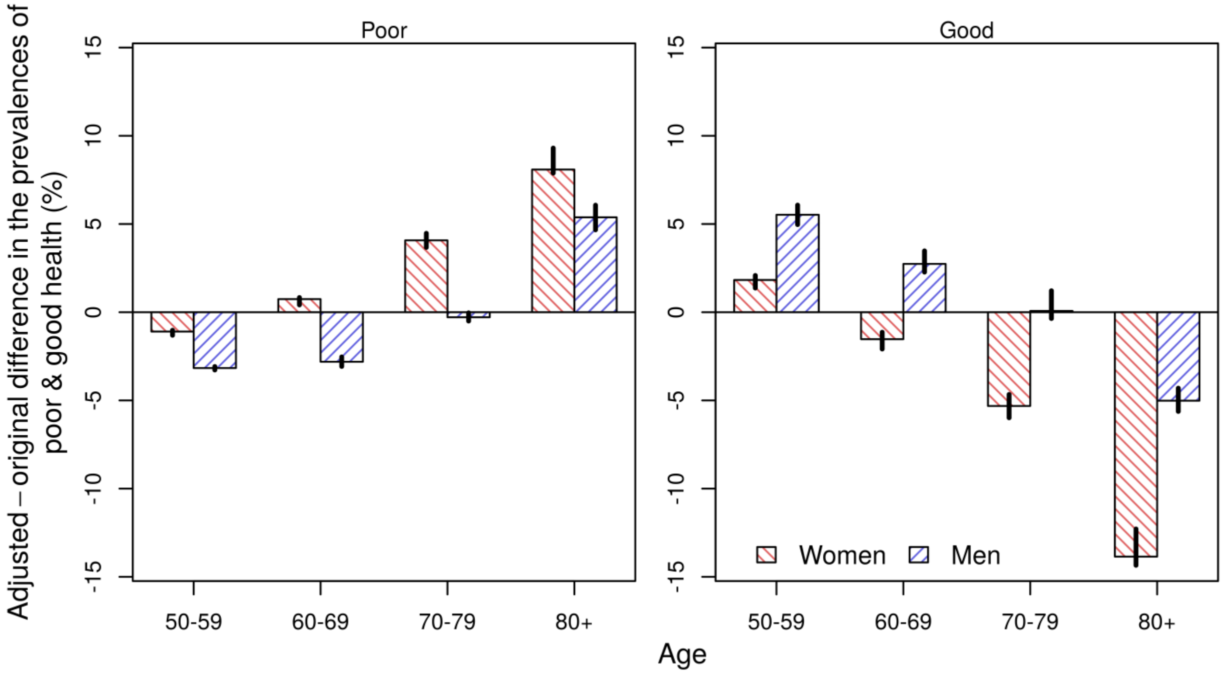


Figure 4. Original (O) and adjusted (A) prevalences of poor and good health among women and men by age and education

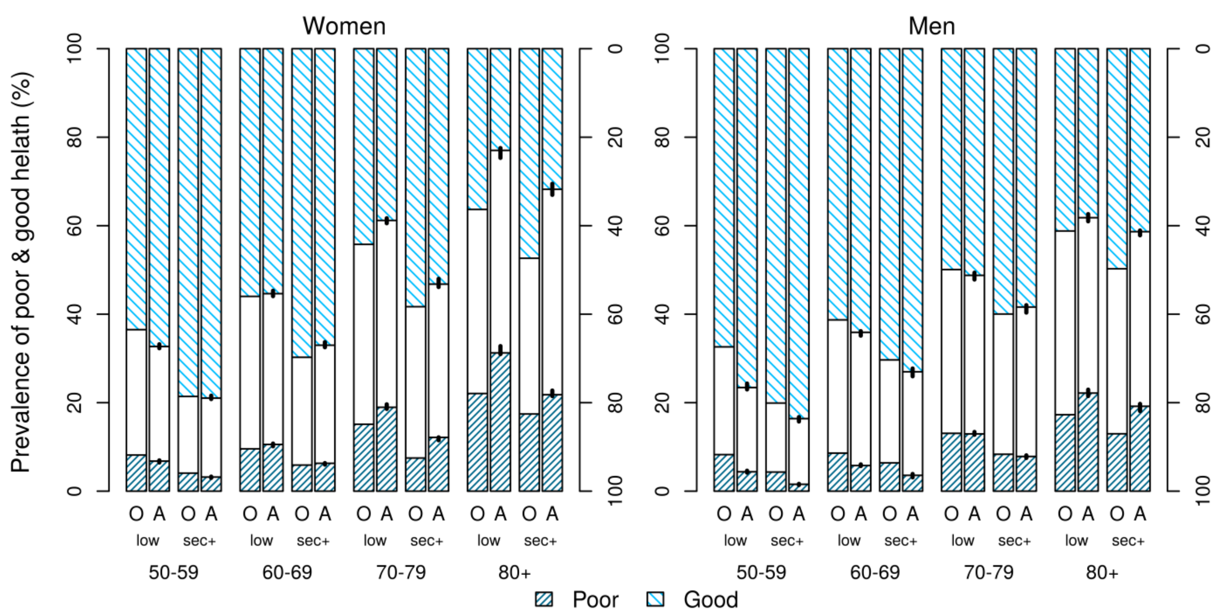


Figure 5. Gender (male – female) differences in the prevalences of poor and good health by age and education

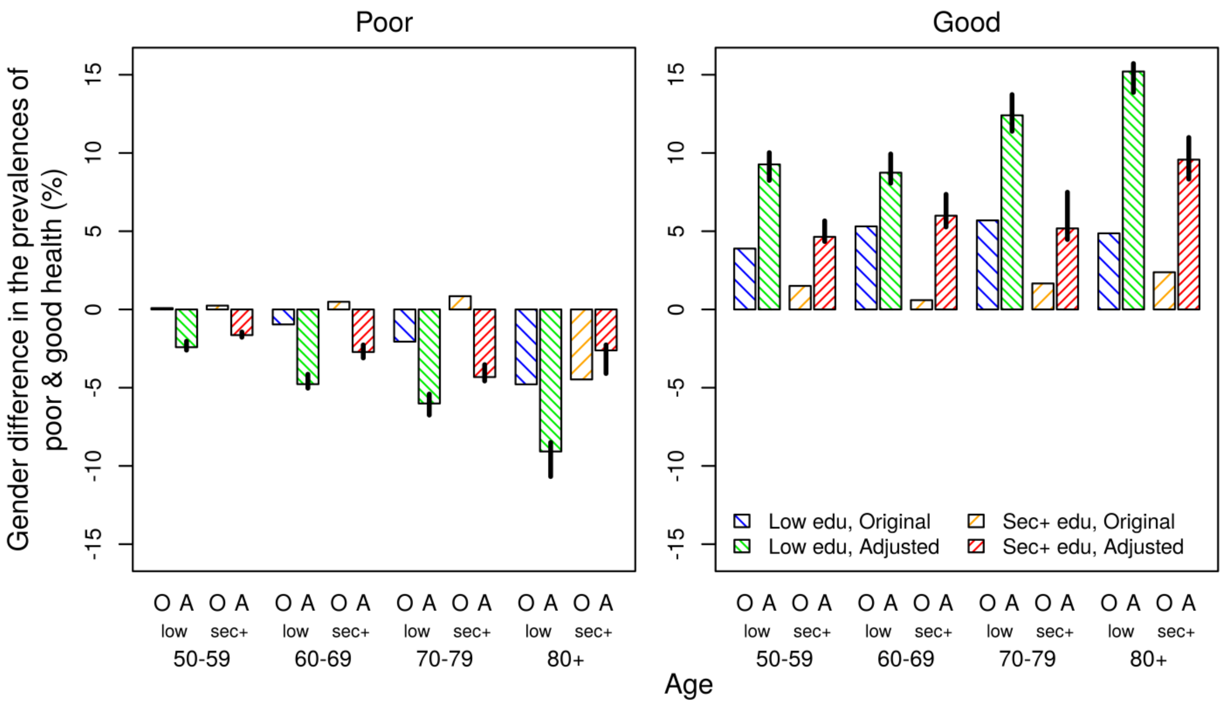
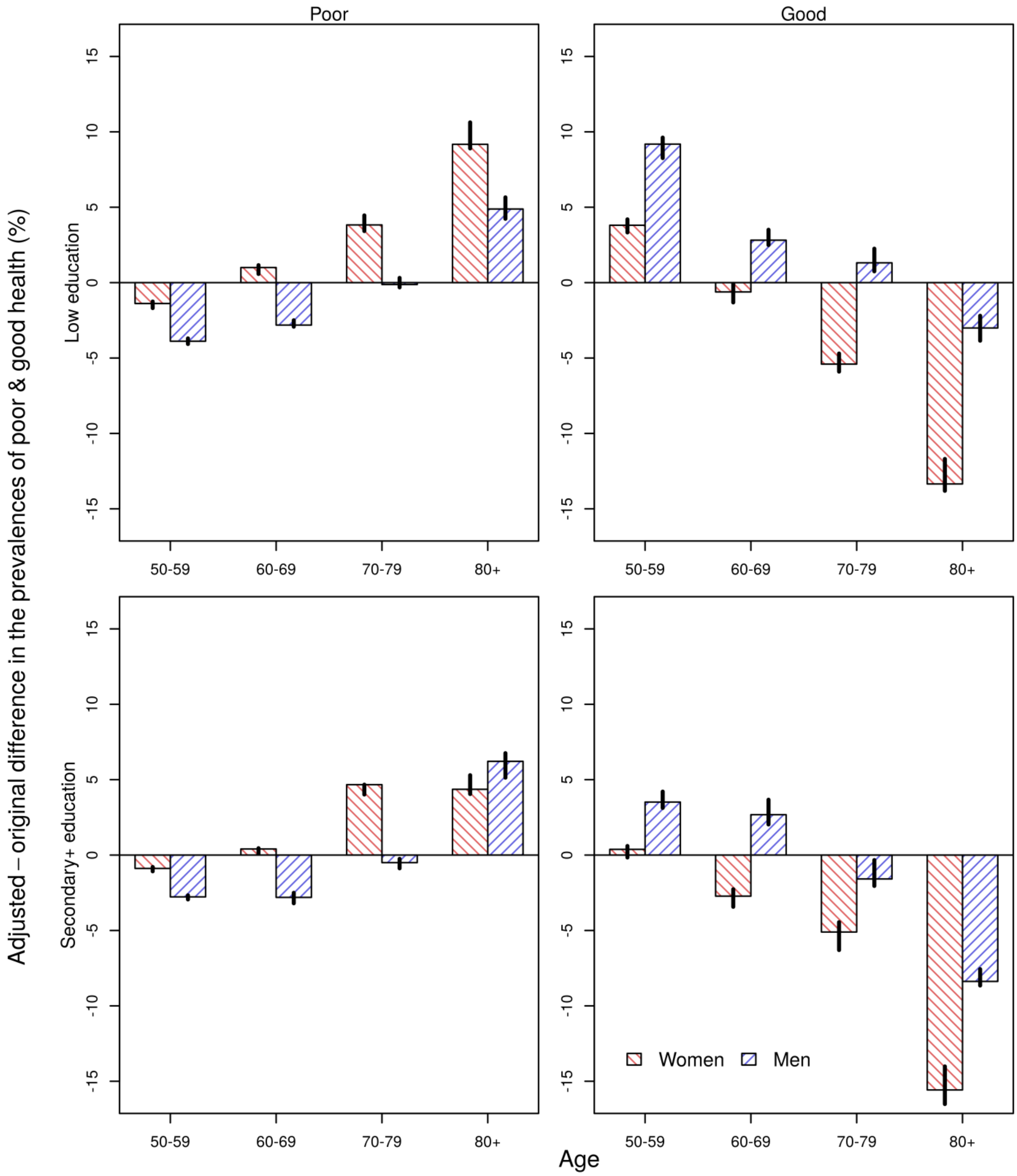


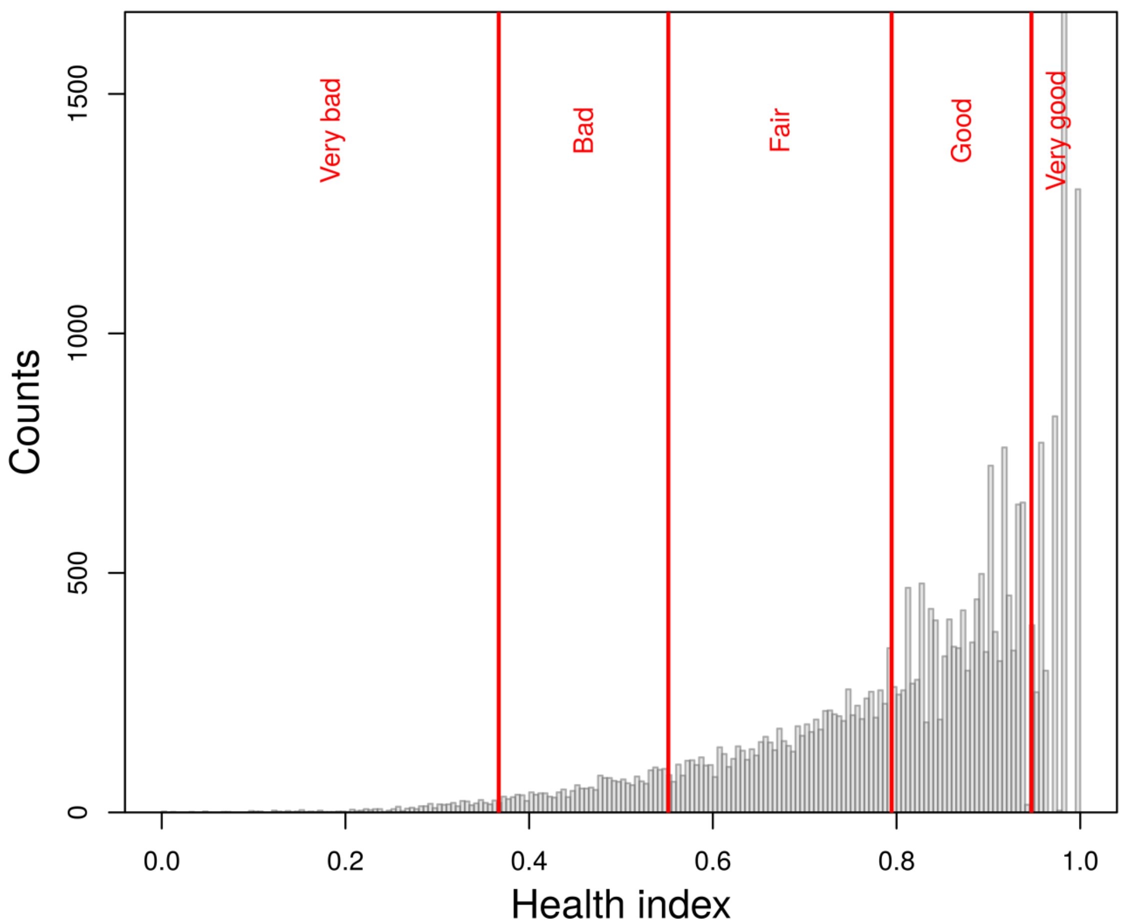
Figure 6. Differences between the adjusted and original prevalences of poor and good health among women and men by age and education


Poor
Good
Low education
Secondary+ education
Adjusted – original difference in the prevalences of poor & good health (%)
Women
Men
50-59
60-69
70-79
80+
Age
15
10
5
0
-5
-10
-15

Supplementary Table 1. Distribution of self-rated health and corresponding centiles of the health index

SRH	Frequency [%]	Cumulative Freq. [%]	Centile
Very bad	1.66	1.66	0.37
Bad	7.26	8.91	0.55
Fair	28.32	37.24	0.79
Good	43.76	81.00	0.95
Very good	19.00	100.00	1.00

Supplementary Figure 1. Distribution of the health index and thresholds for the adjusted general health



Appendix. Statistical model and computations for adjustment of the original self-rated health

A1. Jürges method

This part describes a specific variant of the method by Jürges (2007; Rebelo and Pereira 2014), which we used to calculate a categorical health measure ($J = 5$ categories: very bad, bad, fair, good, very good) (Ψ_i^a) for an individual i. The original self-rated health (SRH) - Ψ_i - is considered to be biased due to the influence of contextual social and cultural factors on an individual's perception of his or her health. The goal of the method is to estimate a health measure that is adjusted for the differences in the influences of these contextual social and cultural factors, i.e. reporting heterogeneity.

In the first step, following Jürges, we fit generalized ordinal probit model to obtain an individual latent health variable. According to the model, the latent, underlying health of an individual i ($i \in 1, 2, \dots, N$) is defined by a continuous latent health variable h_i, which is a linear function of K dichotomous health variables $X_{i,k}$ (including the physician-diagnosed reported chronic conditions, mobility limitations, and depression, abilities to perform simple or more demanding tasks, low score on or inability to perform grip strength, and lifestyle behaviors) and their corresponding coefficients β_k without an intercept. So, for *i*-th individual:

(1)
$$h_i = \sum_{k=1}^{K} \beta_k X_{i,k} \,,$$

Fitting generalized ordinal probit model allows the latent health h_i to be adjusted for the inter-individual heterogeneity in reporting behavior, i.e. in how different people transform their "real" latent health into the ordinal self-rated health categories. We assume that the variations in health reporting can be predicted by threshold variables, namely, gender, age, education level, and country. The procedure for fitting generalized ordinal probit model is described in detail the next section (section A2 of the Appendix).

The most important parameters calculated from the predicted latent health values h_i are the health index H and the disability weights D. These values are then normalized to 0 for the worst observed health state and 1 for the best observed health state:

(2)
$$H_i = 1 - \frac{h_i - \min_i h_i}{\max_i h_i - \min_i h_i}$$

The disability weight D_k inform about the impact of a particular health variable k ($k \in 1, 2, \dots, K$) on health status and is calculated as the normalized coefficient β_k. The normalization is done similarly as for health index.

(3) $$D_k = \frac{\beta_k - \min_i h_i}{\max_i h_i - \min_i h_i}$$

The relation between H_i and D_k can be expressed as:

(4) $$H_i = 1 - \sum_{k=1}^{K} D_k \, X_{i,k}$$

The health index is used to calculate a categorical health measure, which is adjusted for differences in reporting behavior. In the first step we calculate the cumulative frequency of each response of SRH. These frequencies are then used as probabilities in the calculation of percentiles of health index for the total population (Supplementary Table 1). Finally, obtained percentiles are used to re-classify the calculated health index into new categories to produce the adjusted categorical health variable Ψ_i^a (Supplementary Figure 1, see also Jürges (2007) and Rebelo and Pereira (2014)).

A2. Fitting the generalized ordinal probit model to the SRH data.

In section A1 we defined the health variable h_i (eq. 1), which is a function of K dichotomous health variables $X_{i,k}$ and their coefficients β_k. The generalized ordinal probit model has also other variables – threshold variables, which sort continuous health into the adjacent categories. Each threshold has its own coefficient for a modeled threshold variable Y. Thresholds λ are defined according to a discrete exponential cumulative function of a linear combination of threshold variables Y and their coefficients γ, such that a threshold of a lower order is never greater than thresholds of higher orders (King et al. 2004; Jürges 2007):

(5) $$\begin{cases} \lambda_{i,1} = \sum_{m=1}^{M} \gamma_{1,m} Y_{i,m} \\ \lambda_{i,j} = \lambda_{i,j-1} + exp\left(\sum_{m=1}^{M} \gamma_{j,m} Y_{i,m}\right), for\, J-1 \ \geq \ j \geq 2 \end{cases},$$

where M is a the number of threshold variables and j is the index of a threshold. The boundary thresholds for $j = 0$ and $j = J$ are set to $\lambda_{i,0} = -\infty$ and $\lambda_{i,J} = \infty$ respectively. If $J - 1$ determines the number of thresholds (J is number of response categories) then ($J - 1) \times M$ is the number of threshold coefficients.

The threshold and health variables are used to define the log likelihood function, which is maximized in the model to obtain health variable coefficients β and threshold coefficients γ:

(6) $$lnL = \sum_{i=1}^{N} \sum_{j=1}^{J} z_{i,j} ln\left[\Phi\left(\lambda_{i,j} - h_i\right) - \Phi\left(\lambda_{i,j-1} - h_i\right)\right],$$

where, Φ is the standard normal cumulative function, $z_{i,j}$ is an indicator function for the adjusted categorical response variable Ψ_i defined as:

(7) $$z_{i,j} = \begin{cases} 1 \ for \ \Psi_i = j \\ 0 \ for \ \Psi_i \neq j \end{cases}$$

The maximization of the log-likelihood function is done by the Nelder and Mead method (*optim* R-function).

A3. Calculation of the confidence intervals with bootstrap method

The confidence intervals for adjusted categorical health measure and the differences between the original and adjusted health are based on bootstrap percentiles (Bradley and Tibshirani 1993; Lumley 2004). In each of the 1000 bootstrap repetitions a set of new coefficients (βs' and γs') are drawn from the multivariate normal distribution, assuming originally estimated $\hat{\beta}$s' and $\hat{\gamma}$s' as a mean and the sandwich estimator for likelihood functions at $\hat{\beta}$ and $\hat{\gamma}$ as variance-covariance matrix. The drawn βs' and γs' are then used to calculate the measure of interest. After 1000 repetitions the 95% confidence intervals of the measure are calculated using a percentile method

All computations were performed in R (R Core team, 2018). R code is available upon request.