

Muszyńska-Spielauer, Magdalena; Luy, Marc

Working Paper

Well-being adjusted health expectancy - a new summary measure of population health

Vienna Institute of Demography Working Papers, No. 01/2021

Provided in Cooperation with:

Vienna Institute of Demography (VID), Austrian Academy of Sciences

Suggested Citation: Muszyńska-Spielauer, Magdalena; Luy, Marc (2021) : Well-being adjusted health expectancy - a new summary measure of population health, Vienna Institute of Demography Working Papers, No. 01/2021, Austrian Academy of Sciences (ÖAW), Vienna Institute of Demography (VID), Vienna, <https://doi.org/10.1553/0x003cb424>

This Version is available at:

<https://hdl.handle.net/10419/247164>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

ÖAW

AUSTRIAN
ACADEMY OF
SCIENCES

VIENNA INSTITUTE OF DEMOGRAPHY

WORKING PAPERS

01/2021

WELL-BEING ADJUSTED HEALTH EXPECTANCY— A NEW SUMMARY MEASURE OF POPULATION HEALTH

MAGDALENA MUSZYŃSKA-SPIELAUER AND MARC LUY

Vienna Institute of Demography
Austrian Academy of Sciences
Vordere Zollamtsstraße 31 | 1030 Vienna, Austria
vid@oeaw.ac.at | www.oeaw.ac.at/vid



ABSTRACT

We propose a new summary measure of population health (SMPH) called *well-being-adjusted healthy life expectancy* (WAHE). WAHE belongs to a subgroup of health-adjusted life expectancy indicators and measures life expectancy equivalent to full health. The measure combines health and mortality information into a single indicator with weights that quantify the reduction in well-being associated with decreased health. WAHE's advantage over other common SMPHs lies in its ability to differentiate between the consequences of health limitations at various levels of severity and its transparent, simple valuation function. We also demonstrate that WAHE accounts for how health impairment impacts overall well-being according to selected individual and contextual factors.

Our paper empirically evaluates WAHE's performance in terms of agreement and reliability, as well as how strongly it correlates with the other SMPHs when applied to 29 European countries. Health and well-being data are from the 2018 EU-SILC, while life tables are from Eurostat. WAHE's sensitivity to univariate and multivariate state specifications is tested using three Minimum European Health Module health dimensions: chronic morbidity, activity limitations, and self-rated health.

The findings show that WAHE has the highest and most significant correlation with all well-known SMPHs (health expectancy and disability-adjusted health expectancy from the Global Burden of Disease). Moreover, WAHE estimates are in agreement with all other SMPHs. Additionally, like the other SMPHs, all WAHE variants form a group of reliable indicators for studying population health in European countries. Finally, WAHE estimates are robust, regardless of whether health is defined across one or multiple simultaneous dimensions of health. We conclude that WAHE has several advantages over other standard SMPHs and propose adopting it as a universal health indicator for descriptive use in public health.

KEYWORDS

Health expectancy, health adjusted life expectancy, well-being, health, EU-SILC, well-being adjusted life expectancy, WAHE

AUTHORS

Magdalena Muszyńska-Spielauer, Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), E-Mail: magdalena.muszynska@oeaw.ac.at

Marc Luy, Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), E-Mail: marc.luy@oeaw.ac.at

ACKNOWLEDGEMENTS

The authors acknowledge the constructive comments from Isabella Buber-Ennsner on the first draft of the manuscript. We are also thankful to Paola di Giulio for recommending literature on the generalised ordered probit models for estimating the weights and Leora Courtney-Wolfman for language editing the manuscript.

The paper's empirical application is based on data from Eurostat, EU Statistics on Income and Living Conditions survey for the year 2018 (EU-SILC 2018). The responsibility for all conclusions drawn from the data lies entirely with the authors. This work was supported by the European Research Council within the EU Framework Programme for Research and Innovation Horizon 2020, ERC Grant Agreement No. 725187 (LETHE).

WELL-BEING ADJUSTED HEALTH EXPECTANCY— A NEW SUMMARY MEASURE OF POPULATION HEALTH

MAGDALENA MUSZYŃSKA-SPIELAUER AND MARC LUY

1 INTRODUCTION

Summary measures of population health (SMPHs) combine information on health and mortality into a single value. Depending on their intended application, they can be classified as having descriptive, causative, or evaluative use (Essink-Bot and Bonsel, 2002; Murray et al., 2000; Van der Mass, 2002). SMPHs most familiar to demographers are health expectancy (HE) and disability-adjusted years of life (DALY) from the Global Burden of Disease (GBD). HE belongs to the group of measures of descriptive use that describe differences in the distribution of health between populations, sub-groups of populations, or changes in population health over calendar time. Meanwhile, DALY belongs to the group of measures of causative use in health policy, which facilitate analysing the burden of diseases from specific causes. The third group of measures, which have evaluative use, tend to be less familiar to demographers, and are used in economic cost-effectiveness analyses to assess the benefits of health interventions in order to establish priorities for health care policies. Since its development in the 1960s, quality-adjusted life years (QALYs) is the most commonly used SMPH of evaluative use (Gold et al., 2002).

Despite their widespread use in population health, both HE and DALY have received frequent criticism from the scientific community. Because HE is based on a dichotomy of full and decreased health, a main concern is its inability to distinguish between specific levels of disease severity. Additionally, such distinctions between full and decreased health are based on their arbitrary threshold definitions (Murray et al., 2002). DALY overcomes these limitations of dichotomies and arbitrary thresholds by including information on different health states. Nevertheless, it has been criticised for its solutions used to construct weights for health states, differential age weighting, and discounting procedures (Anand and Hanson, 1997; Voigt and King, 2014). Since its development in the early 1990s, however, some issues regarding how weights for health states are constructed in the DALY—albeit not all—have been solved, and its age-weighting and discounting have been discontinued (Salomon et al., 2015; Solberg et al., 2020).

Both HE and DALY focus exclusively on the physiological consequences of decreased health and do not take into account significant and differential effects of health on people's overall well-being (Broome, 2002; Saito et al., 2014). Additionally, neither measure addresses the fact that decreased health conditions occur in a social context that shapes peoples' awareness of health problems and influences the consequences of decreased health (Anand and Hanson, 1997; Voigt and King, 2014). In 1997, Anand and Hanson raised this concern and proposed addressing the issue by using compensated disability weights, as "compensated disability weights would come closer to reflecting the true burden of disability as experienced by the individual. The DALY approach does not distinguish between the quantity of ill-health and the 'burden' associated with it" (p. 694). More recently, Hausman (2015) suggested that weighting health states by well-being would better accomplish the purpose of using SMPHs to compare health across populations. In these measures, well-being weights would quantify the "personal or private value of health" (p.151).

This article aims to propose a new SMPH that accounts for the consequences of decreased health on well-being. Drawing on SMPH classifications according to their intended use, the proposed indicator is of descriptive use for public policy. Its construction, however, is analogous to quality-adjusted life years indices (QALYs) derived from health economics cost-benefit analyses. We have named this new indicator *well-being adjusted health expectancy* (WAHE), which was chosen to reflect its technical solutions in constructing the new measure: The valuation function for different health states is derived according to the subjective well-being of those who experience a given health state. This valuation is also context-specific, which means that it and can be estimated separately depending on the selected individual- and contextual-level characteristics. In the empirical application, we apply the WAHE indicator to study population health from 2018 across 29 European countries. These results are then compared to the corresponding HE, disability-adjusted life expectancy (DALE) from the GBD, and life expectancy (LE). Next, we study the correlation between WAHE and the other measures by applying them to the 29 European populations. Additionally, we address quality issues concerning our new indicator using standard criteria for evaluating new measurement instruments in medical, epidemiological, and psychological studies: agreement and reliability. In the WAHE indicator, health status is specified using either single- or multiple dimensions across the Minimum European Health Module (MEHM), including chronic morbidity, limitations in activities of daily living measured with the Global Activity Limitation Indicator (GALI), and self-rated health.

2 METHODS

This section introduces the methods used to construct SMPH and addresses critical technical issues in developing WAHE, namely: Defining and measuring health dimensions and health states across these dimension; constructing well-being weights for health states; the Sullivan method for redistributing the number of life years lived amongst those spent in a specific health state.

2.1 HEALTH-ADJUSTED LIFE EXPECTANCY

Health expectancy (HE) and health-adjusted life expectancy (HALE) quantify the average number of years lived in a stationary population. Both measures are constructed based on the same formula, while they differ in the number of distinguished and valued health states: HE is based on a dichotomous health state definition that simply divides the number of years lived into those spent in full and decreased health states. Conversely, HALE defines health across a discrete number of states, with decreased health defined across more than one state. The years are then combined into a single indicator by weighting the number of years lived in a given decreased health state according to the degree of dysfunction in which these years are lived. While HE quantifies the average number of years lived in good health in a stationary population, HALE measures the average number of years equivalent to full health (Mathers, 2002). HALE's advantage over HE lies in its technical ability to account for the severity of health impairments and incorporating information about the specific value of decreased health states compared to full health. Analogous to LE, HALE at age x is calculated based on the survivorship function and the prevalence of discrete health states as:

$$HALE_x = \int_x^{\infty} \sum_k [\pi_{k,a} h_{k,a}] S(a) da \quad (1)$$

where $S(a)$ is the probability of survival until age a ; $h_{k,a}$ is the share of the population in health state k at age a , and $\pi_{k,a}$ is the weight assigned to the health state k at age a . The HALE weights indicate the relative value of a year spent in a given health state k compared to a year spent in full health at age a (Mathers, 2002). By contrast, HE only considers a single state k of full health with weight $\omega_{k,a} = 1$.

What distinguishes WAHE from other measures of descriptive use is that the latter are most commonly estimated as HE. Some exceptions include Manuel et al. (2002) and Wolfson (1996) who studied population health based on the Canadian-developed HUI Mark II multi-attribute health classification system. Furthermore, Cutler et al. (1997) and Cutler and Richardson (1998) examined changes in health and mortality within the United States population between 1970 and 1990 by valuating health states across selected chronic diseases according to self-rated health. Erickson (1998) and Erickson et al. (1995) also estimated population health in the United States, using the health and activity limitation index (HALEX), which was later applied by Asada (2007). This index was also applied by Asada and Ohkusa (2004) to describe the health of the Japanese population. The HALE formula is additionally used to estimate disability-adjusted health expectancies in the GBD study, which is an example of a SMPH of causative use.

2.2 VALUATION OF HEALTH STATES

Two technical issues must be addressed as when estimating the weights for HALE construction, as well as for HALYs in general: (1) Whose values should be taken into account, i.e., the general population or only those who experience or experienced a particular disease or health problem; (2) how those values should be collected (Brazier et al., 1999; Helgelsson et al., 2020; Richardson, 2002). The first issue relates back to two types of utility in microeconomics: decision utility and

experienced utility. In HALEs, health state valuations are typically derived from general population surveys, and are therefore based on decision utility, which is the predominant type of utility applied in microeconomics. The rationale for using this approach in health economics is that a population as taxpayers finances health programs that are evaluated and, therefore, has the right to decide which health system investments should be prioritised (Helgelsson et al., 2020; Mukuria and Brazier, 2013; Richardson, 2002). On the other hand, this approach to collecting health state valuations from the general population is often criticised because respondents who have never experienced a specific health problem cannot fully judge how much health problems might impact different aspects of their lives and well-being. Additionally, the general public has been shown to frequently overestimate the effect of a diminished health state on well-being. This is because they underestimate the extent to which they could adapt to a given decreased health state (Dolan, 1996, 2007; Dolan and Green, 1998; Dolan and Kahneman, 2008; Helgelsson et al., 2020; Menzel et al., 2002; Mukuria and Brazier, 2013). Similarly, the original disability weights used in DALE (and the more general DALY) are estimated based on evaluations collected from a panel of health care professionals. Likewise, these have been criticised for not reflecting "individuals' differential ability to cope with their functional limitation" (Anand and Hanson, 1997, p. 689). An additional criticism of DALE weights is that since they are independent of the social context in which disabilities and health limitations occur, they do not accurately reflect the real impact of diseases on an individual's life (Anand and Hanson, 1998). While the GBD weights used in DALY and DALE have not been collected from health professionals since 2010, and instead come from population surveys (Salomon et al., 2015), the abovementioned concerns raised by Anand and Hanson (1997, 1998) still hold.

Rather than collecting information for health states valuation from the public, an alternative involves interviewing individuals who experience a given state of health, which relates to Kahneman's concept of experienced utility in economics (1994). Experienced utility was first proposed as a new standard in health policy evaluations derived from a social survey by Ferrer-i-Carbonell and van Praag (2002) derived from a social survey, as well as by Kahneman and Sugden (2005) and Brazier et al. (2005) according to patient clinical studies.

In consideration of determining whose values should be accounted for in health policy evaluations, no single, universally accepted solution for collecting this type of valuation information currently exists for. Currently, the most frequently used survey methods are visual analogue scaling (VAS), standard gamble (SG), time trade-off (TTO) and willingness-to-pay (WTP) (Brazier et al., 1999; Essink-Bot and Bonsel, 2002; Feeny, 2002). As multi-attribute health-state classification systems describe more states than is practical to directly score in surveys, valuation functions are often estimated from models based on a set of values obtained in a survey for only a selection of states. The experienced utility approach described above also critiques how these methods collect health states' valuations, which led to the development of subjective well-being valuation (SWV). To our knowledge, the first recognised application of SWV in health economics was by Ferrer-i Carbonell and Van Praag (2002), who propose using a method to estimate the valuation function by translating a decrease in subjective well-being caused by a decline in health into monetary terms. This technique estimates the compensating income for the loss in a health state, which refers to the additional income needed to return an individual to their state of well-being from before a given health loss. The approach assumes cardinality of well-being, and the weights are estimated with OLS regressions. Generally speaking, methods used in subjective well-being evaluations either assume the cardinality or only ordinality of the well-being scale (MacKerron, 2012). However, Ferrer-i Carbonell and Frijters (2004) demonstrated that assumptions of the cardinality or ordinality of answers concerning the well-being question do not affect the results in health studies.

According to the subjective well-being evaluation approach, the proposed WAHE indicator quantifies the overall burden of decreased health using health-related well-being based on information collected directly from those who experience specific health states. Like the health-related quality of life utility scores from QALYs (Drummond et al., 2015; Gold et al., 2002; Prieto and Sacristán, 2003; Whitehead and Ali, 2010), WAHE weights are expressed in terms of equivalent years of full health. Hence, a weight for a decreased health status in WAHE is expressed in terms of its equivalency in well-being to that of full health or, using Kaplan and Bush's (1982) terminology, in relation to a *Well-Year*. A *Well-Year* is "the equivalent of a year of completely well life, or a year free of disfunction, symptoms, and health-related problems" (p.64). For example, the reduction in quality of life by half due to a decreased health "will take away 0.5000 *Well-Years* over the course of one year" (Ibidem, p.64) and hence result in this health state being assigned a weight of 0.5.

The generalised ordered probit model is the preferred method for estimating health-related well-being weights in WAHE. Compared to other methods, the model only requires minimal assumptions about the well-being data, which is of ordinal

character. Ordered probit models have been previously used by Cutler et al. (1997) and Cutler and Richardson (1998) to estimate QALY weights to study population health in the United States, whereas we extend the standard ordered probit model to the generalised model in WAHE. The generalised ordered probit model was proposed by King et al. (2004) and used to estimate health-related weights in population studies by Groot (2000), Jürges (2007) and Oksuzyan et al. (2019). Unlike the standard model, the generalised model allows for making the dependent variable thresholds contingent on covariates; for example, country of residence, sex, and age group. This means that the estimated models consider how individuals use different scales to rate their well-being depending on these characteristics.

To adapt the well-being valuation function across contextual factors, different thresholds were set for study countries. This helped capture the variation in cross-country norms about how people evaluate health and well-being. A similar approach was used by Jürges (2007), which investigates how cultural differences across European countries affect personal health ratings. Likewise, decreased health has a different impact on well-being among women and men, who Oksuzyan et al. (2019) and Di Lego et al. (2020) show use different evaluation scales. Similarly, the effect of health on well-being depends on the respondent's age (Ulloa et al., 2013). The WAHE weights for a given health state are estimated as a standardised coefficient for this health state from the model. This standardisation follows Cutler et al.'s (1997) approach of dividing a corresponding coefficient by the difference between the threshold for excellent health and that for poor health. The thresholds for latent health are country-, sex-, and age-specific. A full description of the model and methods to estimate the weights are included in the Appendix.

Like other SMPHs, the WAHE health state can be measured in a single or multiple simultaneous health dimensions. The only methodological issue concerning the single dimension is its defined number of exclusive health states. For multiple dimensions, the health states and the corresponding weights can be modelled multiplicatively, in interaction, multiplicatively with an interaction correction—as in DALE (see GBD, 2015), or as the worst state of health in all dimensions. Given that the prevalence of WAHE health states and well-being weights are estimated from the same survey, the most straightforward solution that does not require any additional assumptions is to define the health of an individual as a complete interaction of health states across all of the health dimensions under study. This means that health and corresponding weights are measured simultaneously on a grid comprised of all health state combinations across the specified health dimensions.

2.3 METHOD TO ESTIMATE HEALTH (ADJUSTED) LIFE EXPECTANCY

Well-being adjusted life years are straightforward to compute using the Sullivan method (Sullivan 1971), which is the most common formula for estimating HE or DALE from cross-sectional data. It combines information about the number of years lived from the period life table with health state prevalence from cross-sectional survey data. The Sullivan method's major advantage in estimating SMPH over the alternative double-decrement life table method or multistate life-table methods lies in its lower data requirements. By contrast, the latter two methods require longitudinal data on transition rates between the various morbidity states obtained from panel surveys (Jagger and Robine, 2011; Rogers et al., 1990).

The Sullivan method is used to estimate the WAHE with the discrete approximation of formula (1). The prevalence of a specific health state is weighted by the well-being weights for a given state, sex, age, and country of residence, and then combined with information on the number of years lived from a period life table as:

$$WAHE_x = \frac{1}{l_x} \sum_{k=1}^n \sum_{a=x}^{\omega} \pi_{k,a} h_{k,a} L_a \quad (2)$$

where $WAHE_x$ is the well-being adjusted health expectancy at age x ; l_x is the number of persons that survived to age x ; L_a is the number of years lived at age a ; $h_{k,a}$ is the share of the population in health state k at age a , and $\pi_{k,a}$ is the health-related well-being weight assigned to the health state k at age a with the weight for full health being $\pi_{k,a} = 1.0$. The values for l_x and L_a stem from a period life table, while the values $h_{k,a}$ and $\pi_{k,a}$ are derived from survey data.

3 EMPIRICAL APPLICATION

3.1 DATA AND METHODS

Our empirical analysis compares WAHE estimates with corresponding values for LE, HE, and DALE. WAHE and HE are estimated using data from the EU Statistics on Income and Living Conditions Survey in 2018 (Eurostat, 2021c) and 2018 life tables come from Eurostat (2021a). When estimating the population share in selected health states, we apply individual cross-sectional weights from the EU-SILC dataset that correct the sample for non-response to the entire survey, but not to specific questions (Osier et al., 2006). The DALE estimates are taken from GBD (2020), while LE values are from Eurostat (2021).

Both HE and WAHE initially specify health state estimates in a single dimension, and then simultaneously in the three health dimensions from the Minimum European Health Module (MEHM): chronic morbidity, limitations in activities of daily living (GALI), and self-rated health. Although these dimensions are strongly correlated at the individual level, none have been shown to be a fully comparable, objective measure of health across countries and different socio-economic groups (Au and Johnston, 2014; Berger et al., 2015; Lazarevic and Brandt, 2020). These three health dimensions are also strongly dependent on contextual factors—and not only on differences in physical health. For example, Jürges (2007) finds significant cross-country differences in health reporting for self-rated health. The effect of morbidity and disability on the limitations represented in GALI is affected by the availability of publicly financed facilities and services for those with health limitations (Anand and Hanson, 1998). Additionally, chronic morbidity reporting depends on respondents' knowledge, which is influenced by health insurance systems, health utilisation, and medical technology levels (Saito et al., 2014). The third health dimension, self-rated health, has been shown to be a good predictor of health outcomes like mortality and hospitalisation (Jylhä, 2011; Lee, 2014). People tend to evaluate their health holistically, but nevertheless according to specific contextual factors. Although the most significant determinants of an individual's poor self-rated health are physical functioning (GALI) and presence and knowledge about chronic morbidity, self-rated health reflects essential dimensions that the other two measures are less likely to include. In particular, this includes mental health and bodily pain (Au and Johnston, 2014; Hardy et al., 2014; Jylhä et al., 1998; Lazarevič, 2019; Lazarevič and Brandt, 2020; Saito et al., 2014; Simon et al., 2005; Singh-Manoux et al., 2006). Likewise, the observed that the effects of decreased health on well-being depends on the specification of the health dimension applied in the study. For example, Wu et al. (2014) studied the effect of health conditions across the EQ-5D composite measure (including mobility, self-care, usual activities of daily living, pain/discomfort, and depression, along with chronic morbidity) on well-being. They reported that physical health had no significant effect on well-being, while the most considerable effects were anxiety/depression and limitations to self-care.

Building on the subjective valuation approach, we derive information on well-being associated with different health states directly from respondents who assess their health across the study dimensions. Similar to Dolan and Metcalfe (2012) and Dolan et al. (2012), subjective well-being is measured from answers to the question, "[o]verall, how satisfied are you with your life?" where options range from 1 ("not satisfied") to 10 ("completely satisfied"). Responses to this question have also been labelled as "happiness," "general satisfaction," and "subjective well-being" (Ferrer-i-Carbonell and Frijters, 2004), or "evaluative well-being" (Steptoe, 2019). They are commonly used to estimate weights for health states for subjective valuation in economic studies (for example, Dolan and Metcalfe, 2012) and were recently applied to the *Years of Good Life* index by Lutz et al. (2021) in their study on sustainable development.

Health dimensions are specified in this study using: (1) The absence or presence of self-reported chronic morbidity; (2) the three GALI levels of health (not limited, limited but not severe, and severely limited), and (3) self-rated health in five categories, ranging from "excellent" to "poor" health states. Full health is defined as the absence of chronic morbidity or no limitations according to GALI, or excellent self-rated health. When studied simultaneously across the three health dimensions, full health combines each of these states. The Appendix includes Figure A1, which shows the prevalence of

age-standardised health states across these studied dimensions, where the standard population is the total population of the 29 European countries.

In total, there were 528,697 respondents from the 29 countries comprising the EU-SILC 2018 dataset; however, Germany and Malta were excluded from this study because of different age coverage. Appendix Tables A1 and A2 present the distribution of respondents per country, as well, as the number of respondents included in the separate analyses for each of the health dimension and for estimating the well-being weights. Since no imputation method was applied for missing data, and observations with missing values for a state across a studied health dimension are removed, there are small differences in the sample size for each analysis. EU-SILC population data is available for all study countries in single-year intervals between 17 and 79 years, as well as an open age interval group for 80+ years. The SMPHs have been estimated at age 15 to make them comparable with DALE estimates available at this age. Like Eurostat (2021b), we assume that the prevalence of decreased health at ages 15 and 16 years is equal to that at age 17 years. The prevalence of health states across the study dimensions is estimated for age groups 15–29, 30–39, 40–49; in five-year age groups between ages 50 and 79, and open-ended for 80+ years. The same age groups are used in the generalised probit models to estimate the well-being weights for WAHE.

Because this empirical work also aims to assess if WAHE is a good candidate for an SMPH, we compare the distribution of WAHEs in the 29 European countries to estimates from other commonly used SMPHs: LE, HE, and DALE. As originally done by Murray and Lopez (1997), we refer to the HALE from the GBD as DALE to distinguish the index from the overall group of HALEs. We study correlation and agreement pairwise between the indices, and also assess our new indicator's reliability. Additionally, we examine a monotonic relationship between the SMPHs pairs using Spearman correlation. While Spearman correlation quantifies the degree to which two rankings are identical, agreement assesses comparability between results obtained by different measurement methods (Giavarina, 2015). Agreement between SMPHs is assessed in this study by Bland-Altman plots with 95 percent limits of agreement for a linear relationship between variables according to Bland and Altman's methodology (1999). Additionally, we assess the reliability of the WAHE measure across different specifications of health.

When comparing the indicators, we examine interrater reliability, which refers to “the degree to which two or more raters are able to differentiate among subjects or objects under similar assessment conditions” (Kottner et al., 2011, p.104). Hence, SMPH reliability is designated by how well this group of indicators can differentiate between population health across the study countries. High reliability signifies that the systematic differences in health levels between countries constitute a high share of the sample's total variability and that the measurement error, which is the sum of differences between values obtained for the indices for each country, is small. Following the guidelines created by Kottner et al. (2011) for estimating and reporting reliability and agreement studies, the intraclass correlation coefficient (ICC) is the preferred statistical method to assess interrater reliability for continuous variables such as the SMPH. In particular, we assess reliability using a two-way mixed effects consistency model (Koo and Li, 2016; McGraw and Wong, 1996; Shrout and Fleiss, 1979). Thus, we assess the reliability of using the group of SMPHs to quantify population health across countries and assess if adding WAHE, as well as other measures, increases the group's reliability. A detailed description of the methods used to assess agreement and reliability is provided in the Appendix A3. However, it should be noted that our results on correlation, agreement, and the reliability of measures only refer to this particular empirical application and are not a universal feature of the indices under study (Streiner et al., 2015). All estimations were carried out in R, and we used the *psych* (Revelle, 2015) and *ordinal* (Christensen, 2019a; 2019b) packages.

3.2 RESULTS

Given that health is recognised as one of the most important determinants of well-being (Graham, 2008; Steptoe, 2019; Steptoe et al., 2015; Wu et al., 2014), it is no surprise that the well-being weights for a year lived in decreased health fall below one (Table 1). The health-related well-being value for a year spent in decreased health drops significantly with the severity of the health limitations. Across all sexes, study age-groups, and countries, a year lived in very good self-rated health combined with chronic morbidity and some limitations of activities across GALI is equivalent in well-being to about 10.5 months of full health, as indicated by the well-being weight of 88 for the health state of activity limitations and 0.89 for

the health state of chronic morbidity and very good self-rated health. Good self-rated health and severe limitations across GALI further decreases this value to 9 months (well-being weights ranging from 0.75 to 0.77). The most considerable effect that decreased health has well-being is for those with fair and poor self-rated health: A year lived in fair self-rated health is equivalent to 7.5 months (well-being weight of 0.63) lived in excellent self-rated health, while a year of poor self-rated health is equivalent to only half a year (well-being weight of 0.50) in well-being. Although the thresholds for the well-being latent variable in the estimated models were significantly different for men and women and the selected age groups, the well-being weights are similar across these characteristics (Table 1 and additional Tables A3–A6 in the Appendix). However, the health-related well-being weights are quite different between countries, even though they show a similar pattern for all the health dimensions under study or simultaneously across the health dimensions (Figure 1). This result serves as empirical evidence to the theoretical assumption that contextual factors associated with country of residence have different effects on how decreased health affects well-being, which should therefore be taken into account in the SMPH. The strongest effect of decreased health on well-being is reflected in Denmark's lowest well-being weights, while the weakest effects can be found in Belgium. For example, a year spent in poor self-rated health is equivalent to 5 months of full health in Denmark (well-being weight of 0.43) and 7 months of full health in Belgium (0.60).

TABLE 1: MEAN VALUES OF HEALTH-RELATED WELL-BEING WEIGHTS AND MEAN AGE-STANDARDISED PREVALENCE OF HEALTH STATES IN 29 EUROPEAN COUNTRIES, BY SEX, 2018

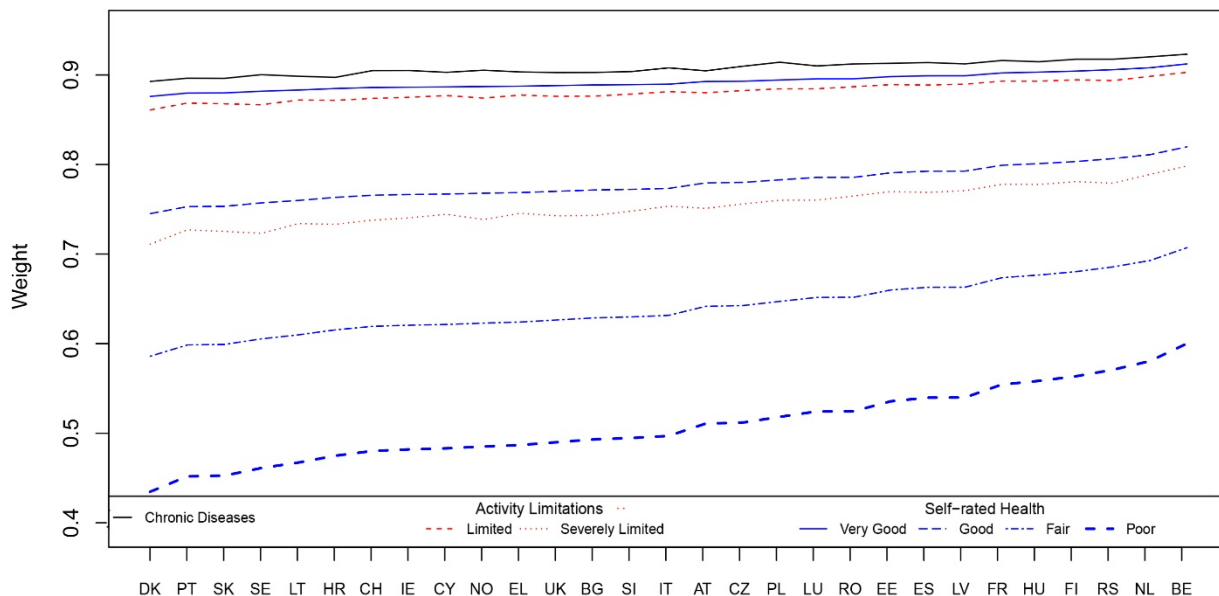
Health Dimension		Weight		Prevalence	
		Men	Women	Men	Women
Chronic Morbidity	No	1.00	1.00	0.71	0.66
	Yes	0.89	0.89	0.29	0.34
Activity Limitations	Not Limited	1.00	1.00	0.78	0.73
	Limited	0.88	0.88	0.15	0.19
	Severely Limited	0.75	0.74	0.07	0.08
Self-rated Health	Excellent	1.00	1.00	0.27	0.23
	Very Good	0.89	0.89	0.46	0.44
	Good	0.78	0.77	0.20	0.23
	Fair	0.64	0.63	0.06	0.08
	Poor	0.50	0.50	0.01	0.02
Multiple (Chronic Morbidity, Activity Limitations Self-rated Health)	No, Not, Excellent	1.00	1.00	0.23	0.19
	No, Not, Very Good	0.89	0.89	0.36	0.34
	No, Not, Good	0.79	0.78	0.06	0.06
	Yes, Not, Very Good	0.91	0.91	0.06	0.06
	Yes, Not, Good	0.80	0.80	0.04	0.05
	Yes, Limited, Good	0.76	0.76	0.07	0.09

Notes: Mean values of well-being weights across age-groups and countries; health states in the multiple dimensions selected if their mean prevalence is higher than 0.03.

Source: Authors' estimations based on Eurostat (2021c).

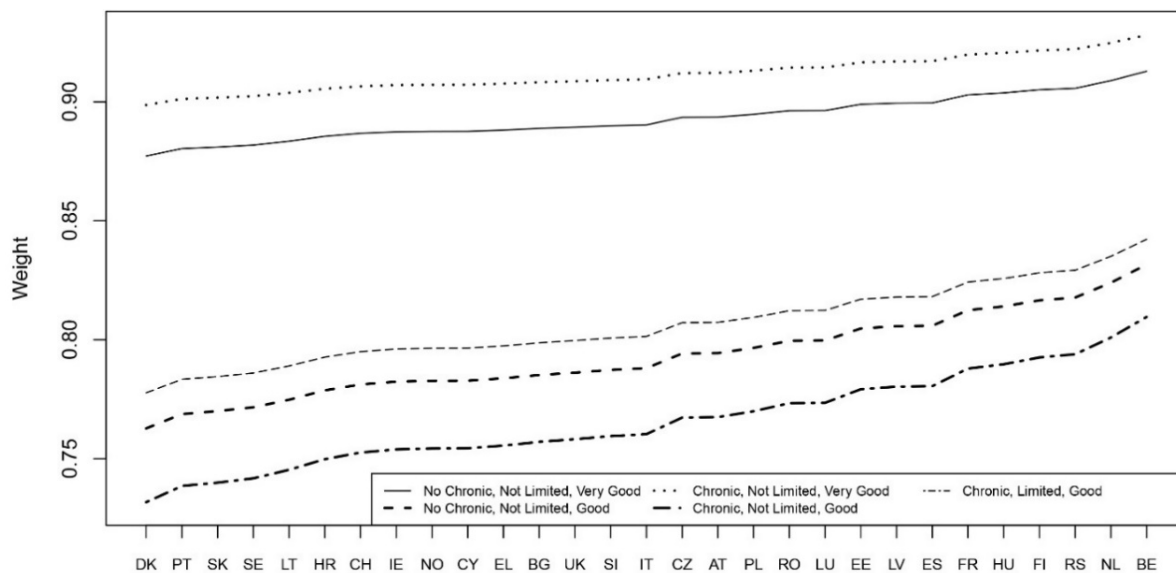
FIGURE 1: WELL-BEING WEIGHTS FOR HEALTH STATES IN 29 EUROPEAN COUNTRIES, 2018, BOTH SEXES

SINGLE HEALTH DIMENSIONS



Note: Countries sorted in ascending order by the level of well-being weights for poor self-rated health.

MULTIPLE HEALTH DIMENSIONS



Notes: Health states in the multiple dimensions (chronic diseases, activity limitations, self-rated health) selected if their mean prevalence is higher than 0.03; countries ordered in ascending order by the level of weights for health states: chronic morbidity, activity limitations, good self-rated health.

Source: Authors' estimations based on Eurostat (2021c).

Table 2 presents summary statistics for the distribution of SMPHs across the 29 European countries. The statistics are estimated based on values for single countries and hence not weighted by the population size. Indices for individual countries are included in the Appendix (Tables A7 and A8). Under stationary assumptions, the average male LE at age 15 in European countries was 63 years, of which 17 to 44 years were spent in full health (HE), 55 to 61 years were spent equivalent to full health (WAHE), and 54 years were disability-free equivalent (DALE). For women, the corresponding numbers are an LE of 68 years, HE of 15 to 61 years, WAHE of 59 to 66 years, and DALE of 56 years (Table 2). The lowest HE and WAHE values appear for measures that are based on self-rated health—both when the health state is specified across self-rated health as a single dimension or one of the multiple health dimensions. Additionally, there are considerable differences in HE between countries based on self-rated health: The coefficient of variation equals to up to 36 percent (37 percent in case of multiple dimensions) for men and 39 percent (40 percent) for women. Low HE based on self-rated health and its substantial variation across the study countries result from the fact that full health is defined by only a single level, which is the highest possible level of “excellent” self-rated health, and the variation of the prevalence of this single health state across countries is large (see also: Figures A1 and A2 in the Appendix). This result demonstrates that when based on a subjective threshold for full and decreased health, HE results are unreliable, which is the main point of criticism for this indicator (Murray et al., 2002). As shown by Jürges (2007) and acknowledged in Eurostat’s accompanying technical guidelines for official HE estimates (2021b), the wide variation in the prevalence of any specific level of self-rated health in social surveys arises from the substantial cultural differences in the comprehension and interpretation of the health rating scales. The WAHE indicator overcomes this issue because it avoids relying on a single threshold between full and decreased health. Instead, WAHE simultaneously accounts for all levels of self-rated health.

The country rankings according to the four WAHE measures are highly correlated, as indicated by the Spearman correlation coefficients above 0.9 (Table 3). In particular, country rankings for WAHE based on multiple dimensions and self-rated health are almost identical, with a correlation coefficient close to one. This means that the information contained in both indicators is almost equal and only subject to a monotonic transformation. Additionally, WAHEs are strongly correlated with LE and DALE. However, it should be noted that country rankings according to DALE and LE are also nearly identical. Moreover, this pattern was also observed for men in country rankings for all studied measures, while the correlation coefficients for the women’s pairs of measures are also high and significant. However, they were lower than for men, indicating some degree of disagreement. The most striking result is that the ranking countries based on DALE are not correlated to rankings according to any of the HE measures. On the other hand, this is not the case for the WAHE indicators. The correlation coefficients are statistically significant for most pairs of WAHEs and HEs, with chronic morbidity among men as the only exception.

It should also be stressed that the WAHEs and DALE rankings are more similar to LE rankings, which only summarises mortality across the countries, but excludes HE values. These similarities reflect the HALE formula, which is used to estimate both WAHE and DALE. The HALE formula applies weights to the prevalence of decreased health; however, in empirical studies, it assigns relatively low weights to health states with the highest prevalence (e.g., chronic morbidity in WAHE), while the opposite is true for rare health states (poor self-rated health). Consequently, decreased quality of population health only leads to a small decrease in quality-adjusted life years and, therefore, the health-adjusted indicators remain close to the value of LE.

TABLE 2: SUMMARY STATISTICS FOR LIFE EXPECTANCY (LE), DISABILITY-ADJUSTED LIFE EXPECTANCY (DALE), HEALTH EXPECTANCY (HE) AND WELLBEING-ADJUSTED HEALTH EXPECTANCY (WAHE) FOR THREE DIMENSIONS OF HEALTH AT AGE 15 IN 29 EUROPEAN COUNTRIES, 2018.

Statistic			Chronic Morbidity		Activity Limitations		Self-rated Health		Multiple Dim.	
	LE	DALE	HE	WAHE	HE	WAHE	HE	WAHE	HE	WAHE
Men										
Mean	62.7	53.7	43.7	60.7	47.8	60.4	17.9	54.9	16.5	54.9
Standard Deviation	3.6	2.8	5.1	3.4	5.0	3.5	6.5	4.1	6.1	4.0
Variation Coefficient (in %)	6	5	12	6	10	6	36	7	37	7
Women										
Mean	68.3	56.4	43.9	65.7	48.3	65.1	17.0	58.7	15.4	58.6
Standard Deviation	2.2	1.3	5.5	2.1	4.8	2.1	6.6	3.1	6.2	3.0
Variation Coefficient (in %)	3	2	12	3	10	3	39	5	40	5

Note: Variation coefficient $V=SD/mean *100\%$.

Source: Authors' estimations based on Eurostat (2021c), Eurostat (2021a) and GBD (2020).

TABLE 3: SPEARMAN CORRELATION COEFFICIENTS FOR THE SUMMARY MEASURES OF POPULATION HEALTH IN 29 EUROPEAN COUNTRIES, 2018.

		LE	DALE	Chronic Morbidity		Activity Limitations		Self-rated Health		Multiple Dim.
				HE	WAHE	HE	WAHE	HE	WAHE	HE
Men										
	DALE	0.94***	1							
Chronic Morbidity	HE	0.35*	0.35*	1						
	WAHE	0.96***	0.93***	0.52**	1					
Activity Limit.	HE	0.59***	0.54***	0.57***	0.63***	1				
	WAHE	0.97***	0.93***	0.41**	0.97***	0.71***	1			
Self-rated	HE	0.54***	0.45**	0.39**	0.50***	0.44*	0.51***	1		
	WAHE	0.91***	0.85***	0.44**	0.91***	0.62***	0.92***	0.70***	1	
Multiple Dim.	HE	0.45**	0.36*	0.46**	0.44**	0.43**	0.43**	0.97***	0.64***	1
	WAHE	0.92***	0.87***	0.44**	0.92***	0.63***	0.93***	0.68***	0.99***	0.61***
Women										
	DALE	0.76***	1							
Chronic Morbidity	HE	0.05	0.01	1						
	WAHE	0.91***	0.70***	0.36	1					
Activity Limit.	HE	0.16	0.08	0.43*	0.23	1				
	WAHE	0.90***	0.69***	0.21	0.90***	0.47**	1			
Self-rated	HE	0.31*	0.15	0.29	0.40*	0.39*	0.43*	1		
	WAHE	0.77***	0.56***	0.28	0.85***	0.34*	0.86***	0.64***	1	
Multiple	HE	0.20	0.10	0.38*	0.32*	0.43*	0.33*	0.98***	0.55***	1
	WAHE	0.77***	0.57***	0.28	0.86***	0.36*	0.87***	0.61***	1.00***	0.52***

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Source: Authors' estimations based on Eurostat (2021c), Eurostat (2021) and GBD (2020).

TABLE 4: INTRACLASS CORRELATION COEFFICIENTS (ICC) FOR ALL SMPHS AND AFTER EXCLUDING SINGLE MEASURES IN 29 EUROPEAN COUNTRIES, 2018.

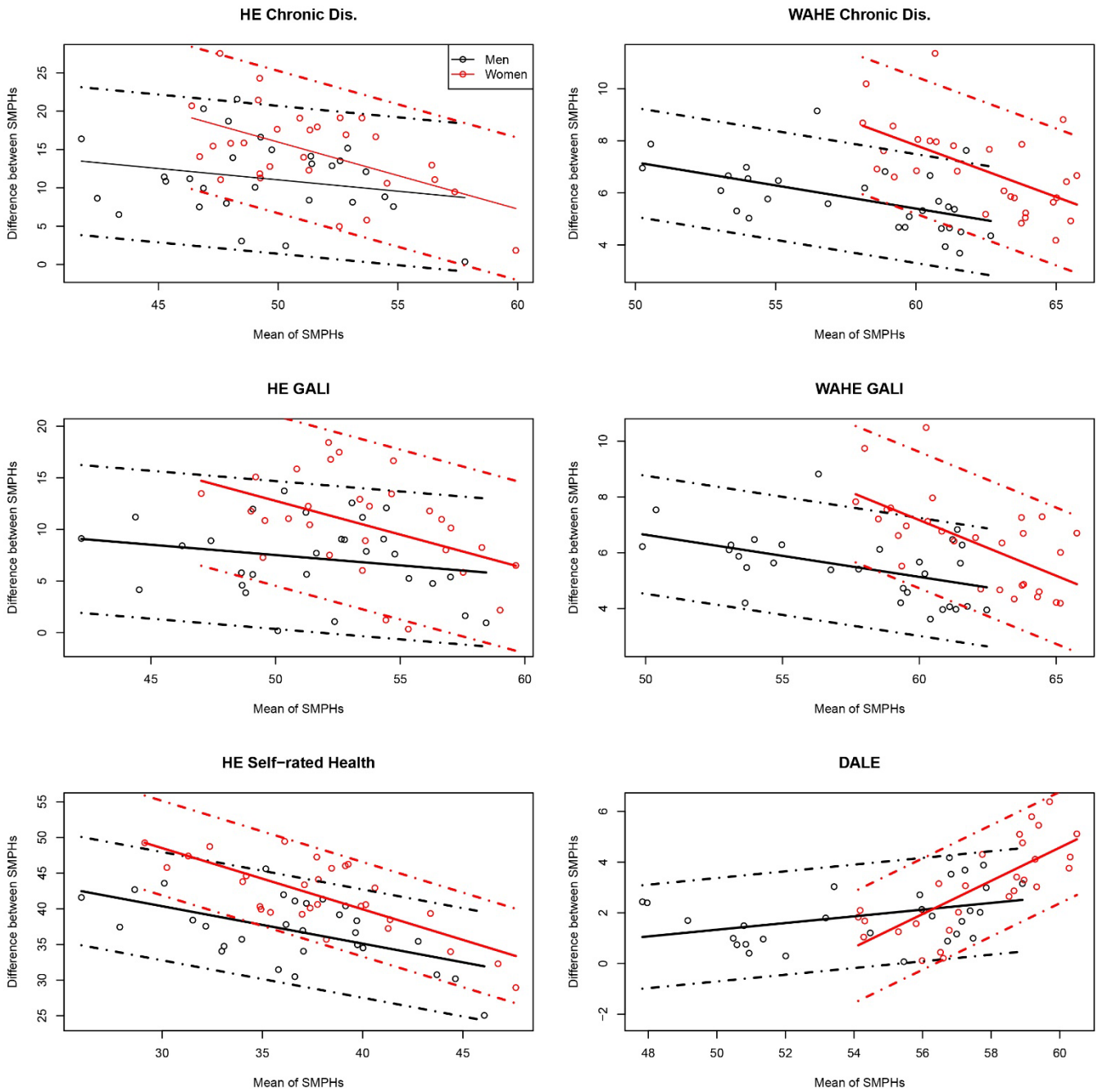
	All	Excluding								
		DALE	Chronic Morbidity		GALI		Self-rated Health		Multiple	
			HE	WAHE	HE	WAHE	HE	WAHE	HE	WAHE
Men										
ICC	0.92	0.92	0.93	0.91	0.92	0.91	0.92	0.90	0.92	0.90
Women										
ICC	0.84	0.85	0.85	0.83	0.84	0.83	0.81	0.81	0.80	0.81

Source: Authors' estimations based on Eurostat (2021C), Eurostat (2021) and GBD (2020).

Table 4 shows the intraclass correlation coefficients for all studied SMPHs combined, as well as the same group with single indicators omitted for the 29 included European populations. The group of SMPHs is visibly characterised by its high reliability for measuring population health across the European countries. This reliability is higher for men, where the intraclass correlation coefficient equals 0.92, compared to 0.84 for women (Table 4). Only minor changes occurred in the measurement's reliability when a single measure was excluded from or added to the group. For most measures, their inclusion in the group of indices does not change the group's reliability—or increases it only mildly. For both sexes, however, excluding HE based on chronic morbidity increases reliability of this group of measures. The same effect is observed for women in case of DALE. Both effects were only minor. Moreover, the small sample size led to the very wide confidence intervals in the correlation coefficients (Streiner et al., 2015), making these changes to reliability insignificant. Figure 2 shows Bland-Altman plots for the agreement between WAHE based on self-rated health and the other SMPHs. Appendix Figures A3 and A4 depict nearly identical Bland-Altman plots for WAHE based on chronic morbidity or GALI. Figure 3 shows Bland-Altman plots of the agreement between DALE and the remaining indices, which demonstrates agreement between the pairwise SMPH measurements: With very few exceptions, the measurements are closer to each other in the Bland-Altman plots than the 95 percent limits of agreement. The most frequent exception in the pairwise comparisons between the HE and other measures was Bulgaria. Here, the lack of agreement between the indices was repeatedly indicated by differences between the two values that occurred outside the limits of agreement. Because there is low prevalence of poor self-reported health in Bulgaria, relatively high HE values coincide with low LE values. By contrast, low Bulgarian LE is connected with low WAHE and DALE, as the two measures put a greater emphasis on mortality, as described above.

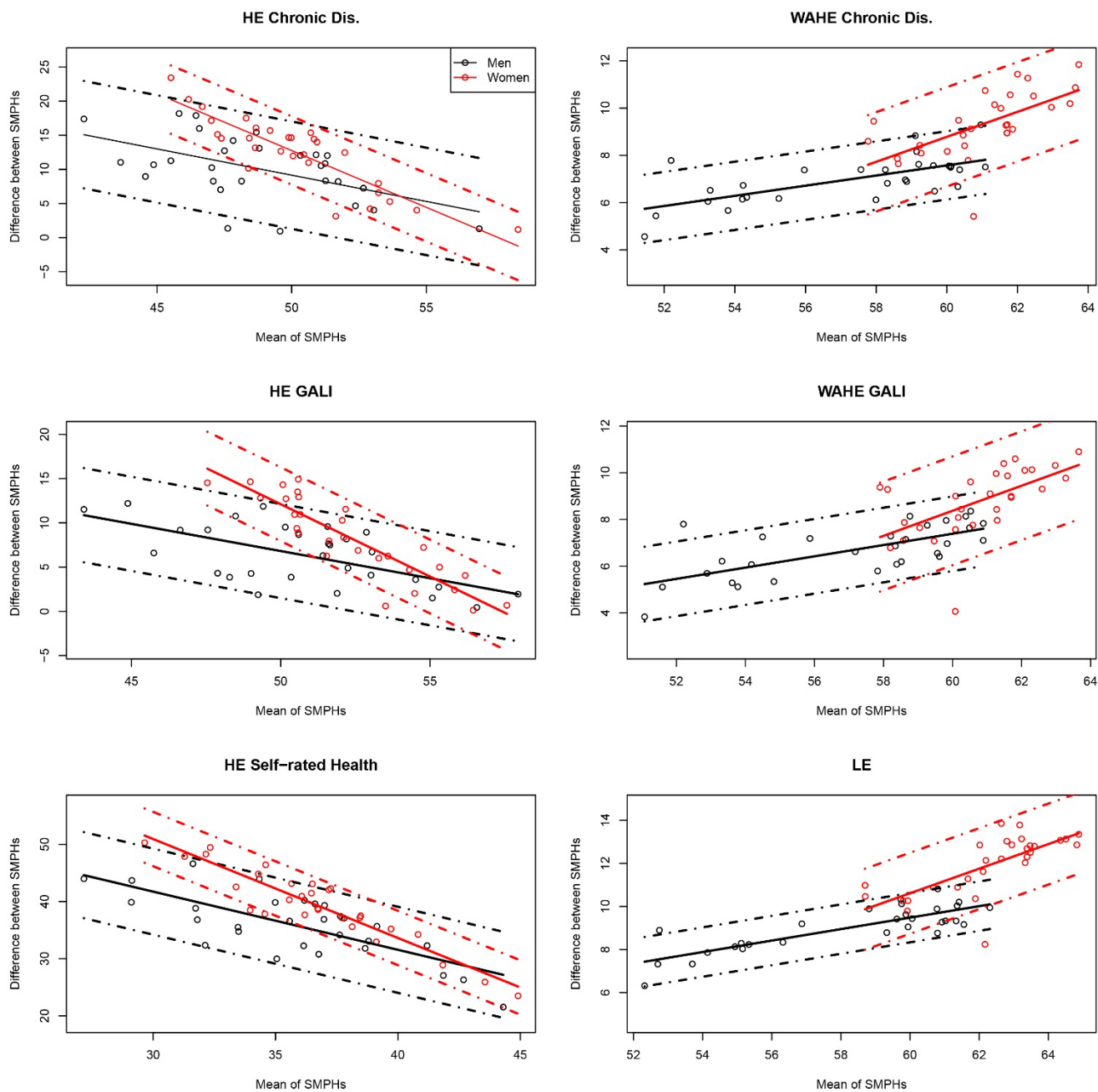
The limits of agreement in the Bland-Altman plots are wide and range between 2 and 10 years. The widest limits of agreement, which indicate the largest absolute differences between indicator values, are between HE based on chronic morbidity and WAHEs, independent of the health dimension. These large absolute differences, however, do not signify a large measurement error in the SMPH. Rather, they simply occur because health is a complex, multidimensional phenomenon, and the indicators quantify its different dimensions. An interesting SMPH feature revealed by the agreement plots is the linear relationship between the mean value of the pairs of indicator values and their absolute differences. Aside from the pairwise comparisons of WAHEs and DALE, each linear relationship between WAHEs and the remaining SMPHs is negative. Combined with the positive correlation between the measures described above, this means that as the indicator values increase, the absolute differences between them decrease. In other words, the indicators report very similar values in countries with a higher level of population health and diverge in countries with a lower level of population health. Interestingly, the slope of the fitted linear models for all pairs of indices are much higher for women than for men. The higher women's slopes result from larger differences between indices at their low values, while the gap between measures at high index values is similar for both sexes. However, our empirical findings show that the opposite is true for the relationships between DALE and WAHE: As the indicator values increase, the differences between them become larger—especially for women. Altogether, these results indicate that, despite their high correlation, there are important differences between the SMPHs in quantifying population health and mortality across the countries under study.

FIGURE 2: BLAND-ALTMAN PLOTS FOR AGREEMENT BETWEEN WAHE BASED ON SELF-RATED HEALTH AND OTHER SMPHS IN 29 COUNTRIES BY SEX, 2018.



Note: Dots represent country observations; solid lines represent the linear model; dashed lines represent the 95% confidence intervals.
Source: Authors' estimations based on Eurostat (2021c), Eurostat (2021a) and GBD (2020).

FIGURE 3: BLAND-ALTMAN PLOTS FOR AGREEMENT BETWEEN DALE AND OTHER SMPHs IN 29 COUNTRIES BY SEX, 2018.



Note: Dots represent country observations; solid lines represent the linear model; dashed lines represent the 95% confidence intervals.

Source: Authors' estimations based on Eurostat (2021c), Eurostat (2021a) and GBD (2020).


4 SUMMARY AND DISCUSSION

In this article, we propose a new SMPH with descriptive and causative use for public policy to join the family of HALE indicators. The Well-being Adjusted Health Expectancy (WAHE) indicator combines information on mortality, health, and health-related well-being. It is constructed using the Sullivan method, which first divides the total number of years lived according to the prevalence of selected health states and then combines them into a particular life-years value by applying health-related weights. These weights quantify the well-being associated with a given health state as compared to well-being in full health. Analogous to QALYs in the economic cost-effectiveness studies and DALEs from the GBD, WAHE measures the expected health-related equivalent length of life under stationary assumptions in the cross-sectional context. The empirical part of the study estimated four WAHE indicators for 29 European countries in 2018. The indicators are based on health state definitions across the three dimensions from the Minimum Health Module: self-rated health, limitations in activities of daily living according to GALI and chronic morbidity. Health is measured both across the three single dimensions and simultaneously across the dimensions. WAHE estimates in the study countries were compared to LE and the two other commonly used SMPHs, HE and DALE. We analysed the correlation between country rankings according to the pairs of indicators, agreement between WAHEs and other SMPHs, and the reliability of this group of measures.

The empirical component demonstrated that WAHE is a good candidate for a SMPH. The ranking of population health across the study countries according to WAHE was similar to other commonly used SMPHs. WAHE performs well, independent of the health dimensions applied to specify health, as demonstrated by its agreement with all other SMPHs and the high reliability of this group of indices. Ranking countries according to WAHE was also highly correlated with the ranking based on LE and DALE, but less so with HE. The largest absolute differences between WAHE and HEs exist at low bounds of these indicators, while the opposite is true for the gap between WAHE and DALE. We also concluded that both measures, WAHE and DALE, reflect greater differences between countries in the total number of years lived than in the prevalence of decreased health. This is because both indices are estimated according to the HALE formula; the most prevalent, relatively good health states have high weights, while the rare, relatively poor health states have low weights. The weights reflect the effect that decreased health has on disability in DALE and on well-being in WAHE.

Compared to DALE, WAHE is based on easily accessible social survey data and has a straightforward valuation function, which accounts for the consequences of decreased health among those who actually experience it. Moreover, WAHE accounts for the fact that decreased health occurs in a social context, which not only shapes differences in the prevalence of specific health conditions, but also their consequences. Although DALE accounts for the consequences of decreased health in its disability weights, it does not differentiate those weights between countries. Using the extreme example of HE, where full health was defined as excellent self-rated health, we also demonstrated that WAHE is a superior indicator, as it summarises information on population health without being driven by the subjective dual distinction between full and decreased health. When compared to HE, WAHE's inclusion of more levels of health states is an essential contribution to using health measurement indicators of descriptive and causative use for public policy. Thus, the WAHE indicator provides a more comprehensive summary of complex health and mortality information than the commonly used HEs.

The main limitation of the study's empirical component concerns the EU-SILC survey methodology. First, while a proxy can answer the health question, the well-being question cannot. This means the WAHE weights are probably too high, because they miss (at least some) information on the well-being for the most severely disabled and those in the worst health state, while their health is still taken into account in the prevalence. Second, the EU-SILC survey is conducted in private households, and thus excludes institutionalised populations. Due to this sample selection method and the fact that the share of those in institutions and with decreased health states is high amongst older adults (Schanze and Zins, 2019), the estimated SMPHs do not depict the actual health situation of the total population. Furthermore, since the share of the institutionalised elderly population differs between countries (Schanze and Zins, 2019), these limitations hinder the comparability of the results across the selected European countries. Note, however, that these limitations apply equally to HE, WAHE, and other measures of health based on the EU-SILC data.



In conclusion, there is no simple answer to the question of which of the SMPH is the best measure for analysing population health across countries and between subpopulations. WAHE proved, however, its potential for a universal measure: We have demonstrated that WAHE performs at least as well as other SMPHs, and in some cases, it can be considered superior to any of the commonly used indices. Since each index, including WAHE, measures different aspects of health, the results from empirical studies based on these indicators should be interpreted in light of these differences and the specificities each index.

REFERENCES

- Anand, S., & Hanson, K. (1997). Disability-adjusted life years: A critical review. *Journal of Health Economics*, 16(6), 685-702.
- Anand, S., & Hanson, K. (1998). DALYs: Efficiency versus equity. *World Development*, 26(2), 307-310.
- Au, N., & Johnston, D. W. (2014). Self-assessed health: What does it mean and what does it hide?. *Social Science & Medicine*, 121, 21-28.
- Asada, Y. (2007). *Health inequality: morality and measurement*. University of Toronto Press.
- Asada, Y., & Ohkusa, Y. (2004). Analysis of health-related quality of life (HRQL), its distribution, and its distribution by income in Japan, 1989 and 1998. *Social Science & Medicine*, 59(7), 1423-1433.
- Berger, N., Van Oyen, H., Cambois, E., Fouweather, T., Jagger, C., Nusselder, W., & Robine, J. M. (2015). Assessing the validity of the Global Activity Limitation Indicator in fourteen European countries. *BMC Medical Research Methodology*, 15(1), 1-8.
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135-160.
- Brazier, J., R. Akehurst, A. Brennan, P. Dolan, K. Claxton, C. McCabe, M. Sculpher, and A. Tsuchiya (2005). Should patients have a greater role in valuing health states? *Applied Health Economics and Health Policy* 4(4), 201-208.
- Brazier, J., Deverill, M., & Green, C. (1999). A review of the use of health status measures in economic evaluation. *Journal of Health Services Research & Policy*, 4(3), 174-184.
- Broome, J. (2002). Measuring the Burden of Disease by Aggregating Well-Being, in: Murray, C. J., Salomon, J. A., Mathers, C. D., Lopez, A. D., & World Health Organization [eds.], *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*. Geneva: World Health Organization.
- Christensen RHB (2019a). Cumulative Link Models for Ordinal Regression with the R Package ordinal. R package vignette version 2019.12-10, <https://CRAN.R-project.org/package=ordinal>.
- Christensen RHB (2019b). "ordinal: Regression Models for Ordinal Data." R package version 2019.12-10, <https://CRAN.R-project.org/package=ordinal>.
- Cutler, D. M., & Richardson, E. (1998). The value of health: 1970-1990. *The American Economic Review*, 88(2), 97-100.
- Cutler, D. M., Richardson, E., Keeler, T. E., & Staiger, D. (1997). Measuring the health of the US population. *Brookings Papers on Economic Activity. Microeconomics*, 1997, 217-282.
- Di Lego, V., Di Giulio, P., & Luy, M. (2020). Gender differences in healthy and unhealthy life expectancy. In *International Handbook of Health Expectancies* (pp. 151-172). Springer, Cham.
- Dolan, P. (1996). Modelling valuations for health states: the effect of duration. *Health Policy*, 38(3), 189-203.
- Dolan, P. (2007). Finding a NICER way to value health: From hypothetical preferences to real experiences. *Social Market Foundation*.

-
- Dolan, P., & Green, C. (1998). Using the person trade-off approach to examine differences between individual and social values. *Health Economics*, 7(4), 307-312.
- Dolan, P., & Kahneman, D. (2008). Interpretations of utility and their implications for the valuation of health. *The Economic Journal*, 118(525), 215-234.
- Dolan, P., Lee, H., & Peasgood, T. (2012). Losing Sight of the Wood for the Trees. *Pharmacoeconomics*, 30(11), 1035-1049.
- Dolan, P., & Metcalfe, R. (2012). Valuing health: A brief report on subjective well-being versus preferences. *Medical Decision Making*, 32(4), 578-582.
- Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press.
- Erickson, P. (1998). Evaluation of a population-based measure of quality of life: The Health and Activity Limitation Index (HALex). *Quality of Life Research*, 7(2), 101-114.
- Erickson, P., R. Wilson, I. I. Shannon, et al. (1995). Years of healthy life. Statistical Notes. National Center for Health Statistics.
- Essink-Bot, M. L., Bonsel (2002). How to derive disability weights. In: Murray, C. J., Salomon, J. A., Mathers, C. D., Lopez, A. D., & World Health Organization [eds.] *Summary measures of population health: concepts, ethics, measurement and applications*, Geneva: World Health Organization, 449-465.
- Eurostat (2021a). Life-tables in European countries, retrieved on 17/03/2021 from https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_mlifetable&lang=en
- Eurostat (2021b). Healthy life expectancy based on self-perceived health (hlth_silc_17), retrieved on 02/04/2021 from https://ec.europa.eu/eurostat/cache/metadata/en/hlth_silc_17_esms.htm.
- Eurostat (2021c). EU Statistics on Income and Living Conditions 2018, data accessed on 03/04/2021.
- Feeny, D. (2002). The utility approach to assessing population health. In: Murray, C. J., Salomon, J. A., Mathers, C. D., Lopez, A. D., & World Health Organization [eds.] *Summary measures of population health: concepts, ethics, measurement and applications*, Geneva: World Health Organization, 515-528.
- Ferrer-i-Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness?. *The Economic Journal*, 114(497), 641-659.
- Ferrer-i-Carbonell, A., & Van Praag, B. M. (2002). The subjective costs of health losses due to chronic diseases. An alternative model for monetary appraisal. *Health Economics*, 11(8), 709-722.
- GBD (2015). Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, 388(10053), 1545-1602.
- GBD (2020). Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019. Life Expectancy and Healthy Life Expectancy 1990-2019. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME).
- Giavarina, D. (2015). Understanding Bland-Altman analysis. *Biochimica Medica*, 25(2), 141-151.
- Gold, M. R., Stevenson, D., & Fryback, D. G. (2002). HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population health. *Annual Review of Public Health*, 23(1), 115-134.

- Graham, C. (2008). Happiness and health: Lessons—and questions—for public policy. *Health Affairs*, 27(1), 72-87.
- Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics*, 19(3), 403-420.
- Hardy, M. A., Acciai, F., & Reyes, A. M. (2014). How health conditions translate into self-ratings: A comparative study of older adults across Europe. *Journal of Health and Social Behavior*, 55(3), 320-341.
- Hausman, D. M. (2015). *Valuing health: Well-being, freedom, and suffering*. Oxford University Press.
- Helgelsson, G., Ernstsson, O., Åström, M., & Burström, K. (2020). Whom should we ask? A systematic literature review of the arguments regarding the most accurate source of information for valuation of health states. *Quality of Life Research*, 29(6), 1465-1482.
- Jagger, C. and J.-M. Robine (2011). Healthy life expectancy. In R. G. Rogers and E. M. Crimmins (Eds.), *International handbook of adult mortality*, pp. 551–568. Springer.
- Jürges, H. (2007). True health vs response styles: exploring cross-country differences in self-reported health. *Health Economics*, 16(2), 163-178.
- Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science & Medicine*, 69(3), 307-316.
- Jylhä, M. (2011). Self-rated health and subjective survival probabilities as predictors of mortality. In *International handbook of adult mortality* (pp. 329-344). Springer, Dordrecht.
- Jylhä, M., Guralnik, J. M., Ferrucci, L., Jokela, J., & Heikkinen, E. (1998). Is self-rated health comparable across cultures and genders?. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 53(3), S144-S152.
- Kahneman, D. (1994). New challenges to the rationality assumption. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, 18-36.
- Kahneman, D., & Sugden, R. (2005). Experienced utility as a standard of policy evaluation. *Environmental and Resource Economics*, 32(1), 161-181.
- Kaplan, R. M., & Bush, J. W. (1982). Health-related quality of life measurement for evaluation research and policy analysis. *Health Psychology*, 1(1), 61.
- King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1), 191-207.
- Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., C. Roberts, M. Shoukri, & Streiner, D. L. (2011). Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *International Journal of Nursing Studies*, 48(6), 661-671.
- Lazarevič, P. (2019). Was misst Self-Rated Health? Die Basis subjektiver Gesundheit und Unterschiede nach Geschlecht, Alter und Kohorte in Europa und Kanada [What Does Self-Rated Health Measure? The Basis of Subjective Health and Differences by Gender, Age, and Cohort in Europe and Canada]. Wiesbaden: Springer.
- Lazarevič, P., & Brandt, M. (2020). Diverging ideas of health? Comparing the basis of health ratings across gender, age, and country. *Social Science & Medicine*, 267, 112913.
- Lee, S. (2014). Self-rated health in health surveys. *Health Survey Methods*, 193-216.

- Lutz, W., Striessnig, E., Dimitrova, A., Ghislandi, S., Lijadi, A., Reiter, C., Spitzer, S. & Yildiz, D. (2021). Years of good life is a well-being indicator designed to serve research on sustainability. *Proceedings of the National Academy of Sciences*, 118(12).
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163.
- MacKerron, G. (2012). Happiness economics from 35 000 feet. *Journal of Economic Surveys*, 26(4), 705-735.
- Manuel, D. G., Schultz, S. E., & Kopec, J. A. (2002). Measuring the health burden of chronic disease and injury using health adjusted life expectancy and the Health Utilities Index. *Journal of Epidemiology & Community Health*, 56(11), 843-850.
- Mathers, C. D. (2002). Health expectancies: an overview and critical appraisal. In: Murray, C. J., Salomon, J. A., Mathers, C. D., Lopez, A. D., & World Health Organization [eds.] *Summary measures of population health: concepts, ethics, measurement and applications*, Geneva: World Health Organization, 177-204.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30.
- Menzel, P., Dolan, P., Richardson, J., & Olsen, J. A. (2002). The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Social Science & Medicine*, 55(12), 2149-2158.
- Mukuria, C., & Brazier, J. (2013). Valuing the EQ-5D and the SF-6D health states using subjective well-being: A secondary analysis of patient data. *Social Science & Medicine*, 77, 97-105.
- Murray, C. J., & Lopez, A. D. (1997). Regional patterns of disability-free life expectancy and disability-adjusted life expectancy: Global Burden of Disease Study. *The Lancet*, 349(9062), 1347-1352.
- Murray, C. J., Salomon, J. A., & Mathers, C. (2000). A critical examination of summary measures of population health. *Bulletin of the World Health Organization*, 78, 981-994.
- Murray, C. J., Salomon, J. A., & Mathers, C. (2002). A critical examination of summary measures of population health. . In: Murray, C. J., Salomon, J. A., Mathers, C. D., Lopez, A. D., & World Health Organization [eds.] *Summary measures of population health: concepts, ethics, measurement and applications*, Geneva: World Health Organization, 205-211.
- Oksuzyan, A., Daňko, M. J., Caputo, J., Jasilionis, D., & Shkolnikov, V. M. (2019). Is the story about sensitive women and stoical men true? Gender differences in health after adjustment for reporting behavior. *Social Science & Medicine*, 228, 41-50.
- Osier, G., Museux, J. M., Seoane, P., & Verma, V. (2006). Cross-sectional and longitudinal weighting for the EU-SILC rotational design. *Methodology of Longitudinal Surveys*, Essex, UK.
- Prieto, L., & Sacristán, J. A. (2003). Problems and solutions in calculating quality-adjusted life years (QALYs). *Health and Quality of Life Outcomes*, 1(1), 1-8.
- Revelle, W. (2015). Package 'psych'. *The comprehensive R archive network*, 337:338.
- Richardson, J. (2002). Evaluating summary measures of population health. In: Murray, C. J., Salomon, J. A., Mathers, C. D., Lopez, A. D., & World Health Organization [eds.] *Summary measures of population health: concepts, ethics, measurement and applications*, Geneva: World Health Organization, 147-160.
- Rogers, A., Rogers, R. G., & Belanger, A. (1990). Longer life but worse health? Measurement and dynamics. *The Gerontologist*, 30(5), 640-649.

- Saito, Y., Robine, J. M., & Crimmins, E. M. (2014). The methods and materials of health expectancy. *Statistical Journal of the IAOS*, 30(3), 209-223.
- Salomon, J. A., Haagsma, J. A., Davis, A., de Noordhout, C. M., Polinder, S., Havelaar, A. H., ... & Murray, C. J. (2015). Disability weights for the Global Burden of Disease 2013 study. *The Lancet Global Health*, 3(11), e712-e723.
- Schanze, J. L., & Zins, S. (2019). Undercoverage of the elderly institutionalized population: The risk of biased estimates and the potentials of weighting. *Survey Methods: Insights from the Field*, 1-19.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420.
- Simon, J. G., De Boer, J. B., Joung, I. M. A., Bosma, H., & Mackenbach, J. P. (2005). How is your health in general? A qualitative study on self-assessed health. *The European Journal of Public Health*, 15(2), 200-208.
- Singh-Manoux, A., Martikainen, P., Ferrie, J., Zins, M., Marmot, M., & Goldberg, M. (2006). What does self rated health measure? Results from the British Whitehall II and French Gazel cohort studies. *Journal of Epidemiology & Community Health*, 60(4), 364-372.
- Solberg, C. T., Sørheim, P., Müller, K. E., Gamlund, E., Norheim, O. F., & Barra, M. (2020). The Devils in the DALY: Prevailing Evaluative Assumptions. *Public Health Ethics*, 13(3), 259-274.
- Steptoe, A. (2019). Happiness and health. *Annual review of public health*, 40, 339-359.
- Steptoe, A., Deaton, A., & Stone, A. A. (2015). Subjective well-being, health, and ageing. *The Lancet*, 385(9968), 640-648.
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: a practical guide to their development and use*. Oxford University Press, USA.
- Sullivan, D. F. (1971). A single index of mortality and morbidity. *HSMHA Health Reports*, 86(4), 347.
- Ulloa, B. F. L., Møller, V., & Sousa-Poza, A. (2013). How does subjective well-being evolve with age? A literature review. *Journal of Population Ageing*, 6(3), 227-246.
- Van der Maas, P. (2002). Applications of summary measures of population health. In: Murray, C. J., Salomon, J. A., Mathers, C. D., Lopez, A. D., & World Health Organization [eds.] *Summary measures of population health: concepts, ethics, measurement and applications*, Geneva: World Health Organization, 53-66.
- Voigt, K., & King, N. B. (2014). Disability weights in the global burden of disease 2010 study: two steps forward, one step back? *Bulletin of the World Health Organization*, 92, 226-228.
- Whitehead, S. J., & Ali, S. (2010). Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin*, 96(1), 5-21.
- Wolfson, M. C. (1996). Health-adjusted life expectancy. *Health Reports-Statistics Canada*, 8, 41-45.
- Wu, M., Brazier, J., Relton, C., Cooper, C., Smith, C., & Blackburn, J. (2014). Examining the incremental impact of long-standing health conditions on subjective well-being alongside the EQ-5D. *Health and Quality of Life Outcomes*, 12(1), 1-8.

APPENDIX

A.1 TABLES AND FIGURES

TABLE A1: NUMBER OF RESPONDENTS IN THE 29 COUNTRIES IN EU-SILC 2018, NUMBER OF RESPONDENTS INCLUDED IN THE ANALYSES FOR EACH HEALTH DIMENSION AND IN THE ESTIMATION OF THE WELL-BEING WEIGHTS, MEN.

Country	Total	Chronic Diseases	GALI	Self-rated Health	Multiple Health States	Well-being
Austria	4,991	4,986	4,988	4,989	4,984	4,511
Belgium	5,382	5,330	5,330	5,329	5,329	4,782
Bulgaria	7,012	7,008	7,008	7,008	7,008	5,120
Switzerland	6,007	5,144	5,141	4,839	4,829	4,839
Cyprus	4,283	4,280	4,280	4,280	4,280	3,701
Czech Rep.	7,495	4,122	4,122	4,122	4,122	4,116
Denmark	4,815	2,626	2,613	2,627	2,609	2,616
Estonia	5,717	4,160	4,159	4,156	4,155	4,162
Greece	23,480	23,480	23,480	23,480	23,480	22,000
Spain	13,536	13,434	13,434	13,434	13,434	13,263
Finland	9,739	5,050	5,046	4,705	4,702	4,712
France	9,475	9,151	9,166	9,149	9,134	6,118
Croatia	8,747	8,681	8,681	8,681	8,679	4,393
Hungary	6,387	6,369	6,353	6,376	6,342	5,252
Ireland	4,103	4,103	4,103	4,103	4,103	2,404
Italy	18,874	18,596	18,587	18,645	18,584	13,379
Lithuania	4,288	4,035	4,025	2,230	2,167	2,242
Luxembourg	4,161	4,133	4,134	4,134	4,128	2,661
Latvia	4,593	4,376	4,376	4,376	4,376	2,913
Netherlands	11,171	5,588	5,530	5,630	5,496	5,641
Norway	5,839	3,103	3,073	3,104	3,073	3,097
Poland	15,624	12,750	12,731	12,777	12,702	7,527
Portugal	13,687	13,663	13,664	13,639	13,638	7,841
Romania	7,437	7,437	7,437	7,437	7,437	6,411
Serbia	6,835	6,835	6,835	6,835	6,835	6,010
Sweden	5,724	2,901	2,888	2,858	2,838	2,846
Slovenia	10,738	4,076	4,076	4,076	4,076	3,424
Slovakia	6,278	6,178	6,168	6,247	6,121	6,278
UK	15,004	14,687	14,686	14,692	14,682	7,876
Total	251,422	216,282	216,114	213,958	213,343	170,135

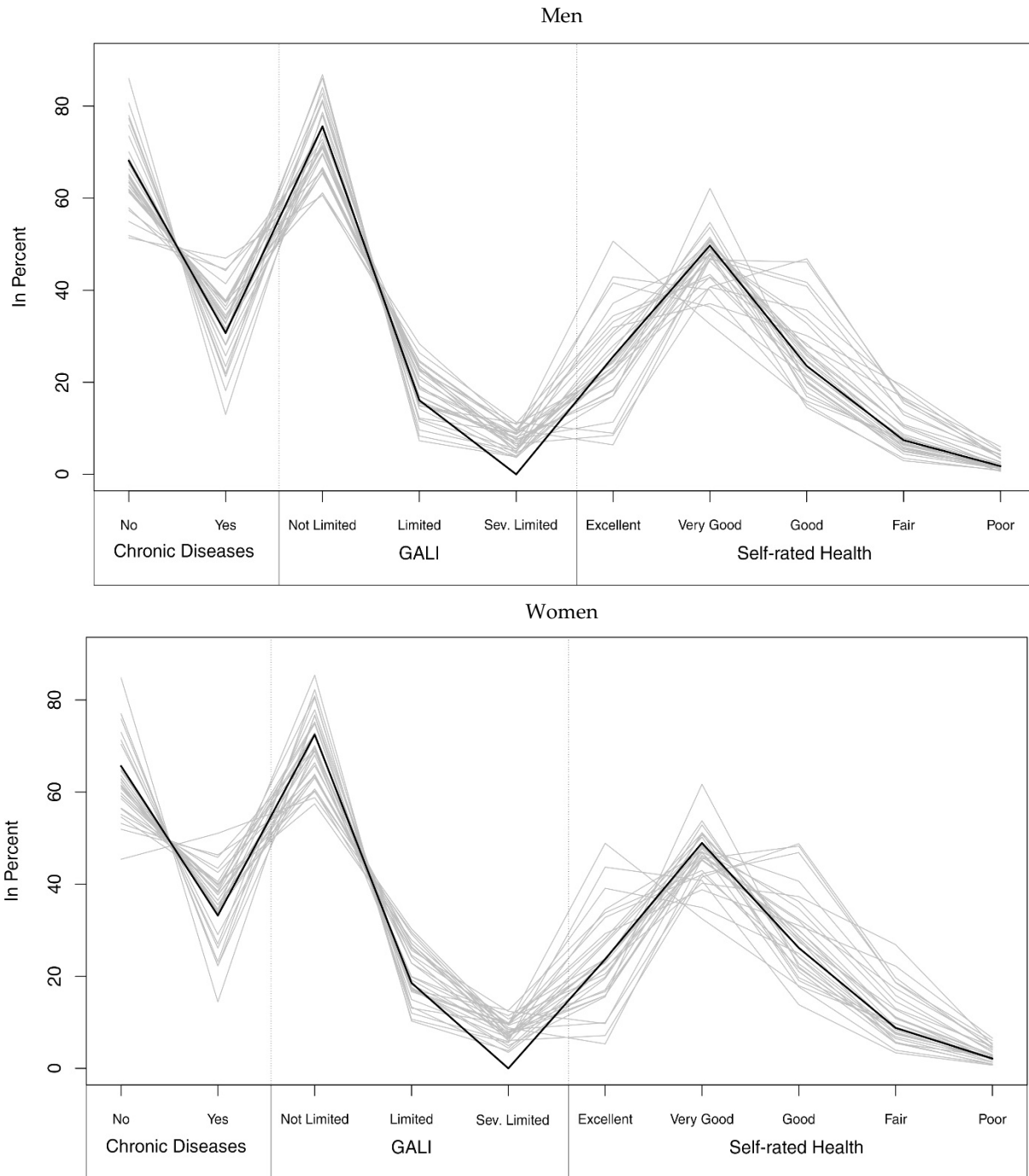
Source: Authors' estimations based on Eurostat (2021c).

TABLE A2: NUMBER OF RESPONDENTS IN THE 29 COUNTRIES IN EU-SILC 2018, NUMBER OF RESPONDENTS INCLUDED IN THE ANALYSES FOR EACH HEALTH DIMENSION AND IN THE ESTIMATION OF THE WELL-BEING WEIGHTS, WOMEN.

Country	Total	Chronic Diseases	GALI	Self-rated Health	Multiple Health States	Well-being
Austria	5,642	5,642	5,639	5,641	5,638	5,288
Belgium	5,700	5,650	5,650	5,650	5,649	5,162
Bulgaria	7,927	7,924	7,924	7,924	7,924	6,812
Switzerland	6,534	5,756	5,748	5,450	5,427	5,448
Cyprus	4,919	4,917	4,917	4,917	4,917	4,309
Czech Rep.	8,559	6,566	6,566	6,566	6,566	6,562
Denmark	5,134	2,927	2,918	2,922	2,911	2,917
Estonia	6,573	5,698	5,698	5,697	5,696	5,699
Greece	25,423	25,423	25,423	25,423	25,423	24,090
Spain	14,836	14,726	14,726	14,726	14,726	14,552
Finland	9,485	4,672	4,667	4,431	4,428	4,437
France	10,477	10,216	10,230	10,217	10,197	8,435
Croatia	9,625	9,536	9,535	9,535	9,534	6,297
Hungary	7,978	7,962	7,948	7,972	7,936	7,464
Ireland	4,387	4,387	4,387	4,387	4,387	3,125
Italy	21,095	20,735	20,736	20,798	20,720	16,167
Lithuania	5,467	5,271	5,265	4,116	4,039	4,131
Luxembourg	4,365	4,343	4,350	4,348	4,340	3,317
Latvia	6,192	6,044	6,044	6,044	6,044	5,155
Netherlands	11,942	6,776	6,682	6,837	6,628	6,852
Norway	5,793	2,836	2,799	2,836	2,798	2,822
Poland	17,756	15,945	15,901	15,971	15,873	13,299
Portugal	15,673	15,647	15,649	15,633	15,631	10,962
Romania	8,100	8,100	8,100	8,100	8,100	7,349
Serbia	7,152	7,152	7,152	7,152	7,152	6,474
Sweden	5,680	2,922	2,907	2,881	2,860	2,870
Slovenia	11,186	4,593	4,593	4,593	4,593	5,245
Slovakia	7,310	7,200	7,209	7,287	7,152	7,310
UK	16,365	16,140	16,136	16,138	16,130	9,331
Total	277,275	245,706	245,499	244,192	243,419	211,881

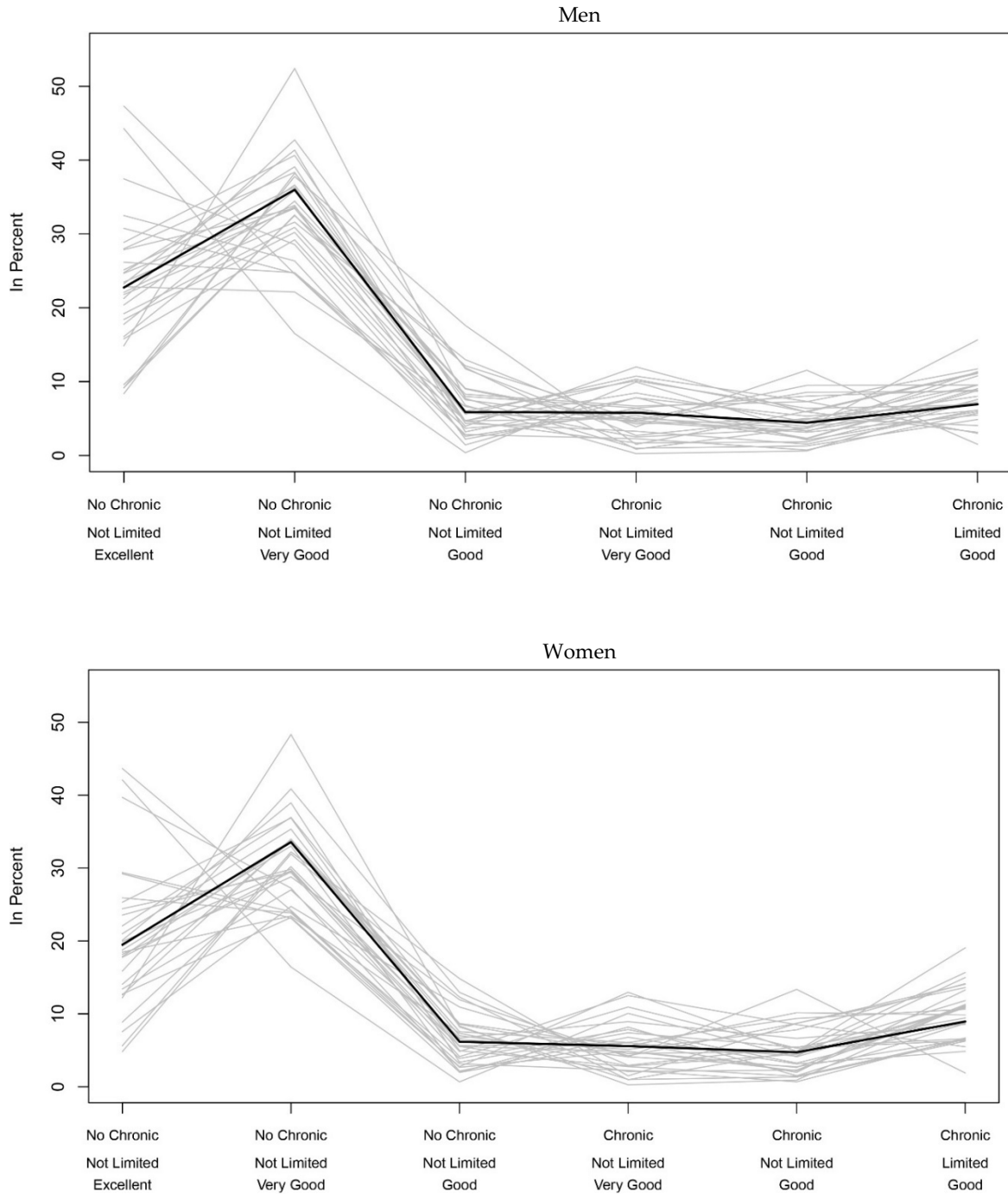
Source: Authors' estimations based on Eurostat (2021c).

FIGURE A1: PREVALENCE OF HEALTH STATES ACROSS THE STUDY HEALTH DIMENSIONS IN ALL 29 EUROPEAN COUNTRIES (SOLID LINE) AND ACROSS THE SINGLE COUNTRIES (GREY LINES) BY SEX.



Source: Authors' estimations based on Eurostat (2021c).

FIGURE A2: PREVALENCE OF SELECTED HEALTH STATES WITH MINIMUM 3 PERCENT PREVALENCE ACROSS THE MULTIPLE HEALTH DIMENSIONS IN ALL 29 EUROPEAN COUNTRIES (SOLID LINE) AND ACROSS THE SINGLE COUNTRIES (GREY LINES) BY SEX.



Source: Authors' estimations based on Eurostat (2021c).

TABLE A3: COEFFICIENTS FOR THE GENERALISED ORDERED PROBIT MODEL FOR CHRONIC MORBIDITY

		Estimate	Std.Error									
Main effect	Chronic Morbidity	-0.8	0.01									
Threshold Coefficients		Threshold										
		1	2	3	4	5	6	7	8	9	10	
	Intercept	-6.40	-6.01	-5.48	-4.73	-4.11	-2.72	-2.27	-1.37	0.05	0.97	
Age Group	15-29	0.21	0.09	0.15	0.13	0.13	0.13	0.15	0.15	0.12	0.09	
Ref: 65-70	30-39	0.44	0.41	0.43	0.40	0.38	0.33	0.31	0.26	0.16	0.04	
	40-49	0.12	0.06	0.07	0.03	0.01	-0.07	-0.03	-0.01	0.00	0.02	
	50-54	0.28	0.24	0.21	0.17	0.15	0.06	0.10	0.10	0.17	0.23	
	55-59	0.48	0.36	0.31	0.25	0.23	0.10	0.16	0.19	0.24	0.36	
	60-64	0.35	0.30	0.24	0.16	0.17	0.03	0.11	0.17	0.27	0.43	
	70-74	0.14	0.11	0.10	0.06	0.05	-0.09	0.00	0.07	0.16	0.37	
	74-79	-0.08	-0.15	-0.18	-0.18	-0.14	-0.29	-0.16	-0.05	0.00	0.24	
	80+	-0.27	-0.35	-0.32	-0.35	-0.33	-0.50	-0.34	-0.22	-0.21	0.00	
Country	Belgium	-0.08	0.22	0.53	0.46	0.43	0.15	0.33	0.57	1.10	1.73	
Ref: Austria	Bulgaria	2.49	2.90	3.16	3.18	3.19	2.74	2.77	2.57	2.10	1.93	
	Switzerland	0.33	0.30	0.29	0.13	-0.01	-0.19	-0.18	-0.07	-0.03	-0.05	
	Cyprus	1.44	1.45	1.44	1.34	1.16	0.95	1.01	0.98	0.82	0.89	
	Czechia	0.52	0.80	1.00	1.02	0.90	0.92	0.92	0.82	0.61	0.59	
	Denmark	1.08	1.29	1.47	1.29	1.07	0.47	0.43	0.27	-0.11	-0.25	
	Estonia	0.58	0.80	0.94	1.12	1.09	1.05	1.16	1.18	1.08	0.98	
	Greece	1.98	2.05	2.01	1.87	1.82	1.49	1.86	1.78	1.49	1.46	
	Spain	0.52	0.93	1.07	0.94	0.84	0.62	0.76	0.83	0.80	1.01	
	Finland	-0.41	-0.24	-0.28	-0.39	-0.44	-0.73	-0.72	-0.50	-0.16	0.69	
	France	0.39	0.48	0.59	0.68	0.74	0.55	0.71	0.89	1.07	1.19	
	Croatia	2.02	2.21	2.39	2.29	2.05	1.78	1.76	1.60	1.29	1.02	
	Hungary	1.10	1.38	1.64	1.82	1.86	1.60	1.76	1.73	1.68	1.71	
	Ireland	-0.03	0.40	0.66	0.63	0.54	0.18	0.23	0.07	-0.12	-0.22	
	Italy	1.23	1.32	1.30	1.21	1.34	0.83	1.18	1.43	1.35	1.12	
	Lithuania	1.90	2.04	2.19	2.16	2.04	1.74	1.74	1.61	1.26	1.05	
	Luxembourg	0.68	0.84	0.90	0.66	0.49	0.39	0.47	0.60	0.67	0.78	
	Latvia	0.72	1.03	1.28	1.48	1.48	1.35	1.51	1.57	1.43	1.24	
	Netherlands	-0.06	0.33	0.43	0.35	0.22	-0.35	-0.03	0.37	0.84	1.23	

Source: Authors' estimations based on Eurostat (2021c).

TABLE A3 (CONT.): COEFFICIENTS FOR THE GENERALISED ORDERED PROBIT MODEL FOR CHRONIC MORBIDITY

Threshold Coefficients		Threshold									
		1	2	3	4	5	6	7	8	9	10
	Norway	0.29	0.56	0.45	0.41	0.38	0.12	0.11	0.13	0.05	-0.06
	Poland	-0.41	0.14	0.40	0.38	0.33	0.41	0.49	0.43	0.21	0.17
	Portugal	1.88	2.05	2.11	1.91	1.67	1.53	1.46	1.25	1.03	0.80
	Romania	0.85	1.03	1.25	1.20	1.23	0.73	0.86	0.92	0.92	1.18
	Serbia	1.44	2.24	2.45	2.66	2.73	2.49	2.61	2.51	2.25	2.35
	Sweden	0.77	1.18	1.07	0.76	0.56	0.24	0.19	0.31	0.21	0.00
	Slovenia	1.19	1.37	1.23	1.17	0.96	0.87	0.94	0.91	0.87	0.77
	Slovakia	1.52	1.69	2.05	1.85	1.65	1.40	1.37	1.12	0.72	0.43
	United Kingdom	0.78	1.07	1.10	0.89	0.76	0.33	0.31	0.29	0.16	0.21
Sex	Women	0.06	0.01	0.02	0.03	0.04	0.05	0.05	0.01	-0.01	-0.03

Source: Authors' estimations based on Eurostat (2021c).

TABLE A4: COEFFICIENTS FOR THE GENERALISED ORDERED PROBIT MODEL FOR GALI

		Estimate	Std.Error
Main effect	Limited	-0.47	0
Activity Limitations	Severely Limited	-0.97	0.01

Threshold Coefficients		Threshold									
		1	2	3	4	5	6	7	8	9	10
	Intercept	-3.3	-3.15	-2.94	-2.63	-2.35	-1.65	-1.4	-0.86	0.04	0.6
Age Group	15-29	0.04	-0.01	0.03	0.03	0.03	0.04	0.06	0.06	0.04	0.02
Ref: 65-70	30-39	0.04	0.04	0.07	0.08	0.09	0.09	0.09	0.06	-0.01	-0.09
	40-49	0.08	0.06	0.06	0.04	0.03	-0.03	0.00	0.00	0.01	0.02
	50-54	0.17	0.15	0.14	0.12	0.11	0.05	0.08	0.07	0.10	0.13
	55-59	0.27	0.22	0.2	0.17	0.17	0.08	0.11	0.13	0.14	0.19
	60-64	0.24	0.22	0.19	0.15	0.14	0.04	0.08	0.12	0.16	0.23
	70-74	0.17	0.16	0.14	0.11	0.09	-0.03	0.01	0.04	0.09	0.20
	74-79	0.11	0.06	0.03	0.00	-0.01	-0.14	-0.08	-0.04	0.00	0.13
	80+	0.07	0.01	-0.01	-0.06	-0.09	-0.24	-0.18	-0.14	-0.13	0.00

Source: Authors' estimations based on Eurostat (2021c).

TABLE A4 (CONT.): COEFFICIENTS FOR THE GENERALISED ORDERED PROBIT MODEL FOR GALI

Threshold Coefficients		Threshold									
		1	2	3	4	5	6	7	8	9	10
Country	Belgium	-0.02	0.08	0.21	0.19	0.19	0.07	0.18	0.34	0.65	0.89
Ref: Austria	Bulgaria	1.16	1.38	1.58	1.71	1.82	1.70	1.75	1.63	1.25	1.04
	Switzerland	0.19	0.19	0.18	0.11	0.05	-0.06	-0.06	-0.01	0.01	0.00
	Cyprus	0.65	0.67	0.69	0.68	0.64	0.61	0.67	0.68	0.57	0.55
	Czechia	0.28	0.38	0.47	0.5	0.47	0.56	0.58	0.55	0.41	0.35
	Denmark	0.40	0.50	0.62	0.59	0.53	0.27	0.25	0.18	-0.06	-0.13
	Estonia	0.22	0.31	0.37	0.49	0.52	0.59	0.70	0.74	0.66	0.53
	Greece	0.83	0.88	0.9	0.88	0.92	0.83	1.10	1.09	0.85	0.75
	Spain	0.31	0.47	0.54	0.51	0.49	0.43	0.53	0.59	0.55	0.61
	Finland	-0.08	-0.03	-0.06	-0.1	-0.13	-0.3	-0.32	-0.22	-0.03	0.45
	France	0.22	0.25	0.3	0.34	0.4	0.35	0.47	0.61	0.70	0.69
	Croatia	0.81	0.92	1.07	1.11	1.05	1.04	1.06	1.00	0.77	0.55
	Hungary	0.48	0.61	0.75	0.92	1.01	1.00	1.14	1.16	1.05	0.94
	Ireland	0.08	0.23	0.34	0.35	0.32	0.16	0.19	0.10	-0.02	-0.08
	Italy	0.55	0.59	0.59	0.58	0.67	0.44	0.67	0.85	0.77	0.58
	Lithuania	0.85	0.93	1.05	1.11	1.12	1.08	1.11	1.06	0.79	0.60
	Luxembourg	0.26	0.32	0.36	0.28	0.22	0.20	0.25	0.35	0.39	0.40
	Latvia	0.32	0.44	0.55	0.69	0.73	0.77	0.91	0.97	0.84	0.65
	Netherlands	-0.04	0.11	0.15	0.14	0.09	-0.19	-0.02	0.23	0.52	0.67
	Norway	0.23	0.34	0.30	0.29	0.28	0.17	0.17	0.18	0.11	0.04
	Poland	0.00	0.19	0.28	0.27	0.26	0.31	0.36	0.34	0.19	0.15
	Portugal	0.80	0.90	0.97	0.94	0.88	0.93	0.92	0.81	0.65	0.46
	Romania	0.38	0.45	0.55	0.55	0.60	0.38	0.48	0.54	0.52	0.61
	Serbia	0.77	1.12	1.26	1.47	1.59	1.60	1.71	1.65	1.38	1.26
	Sweden	0.50	0.65	0.62	0.50	0.43	0.28	0.25	0.32	0.24	0.10
	Slovenia	0.48	0.56	0.52	0.53	0.46	0.49	0.56	0.57	0.54	0.43
	Slovakia	0.58	0.67	0.87	0.84	0.79	0.77	0.79	0.66	0.41	0.21
	United Kingdom	0.29	0.42	0.46	0.40	0.38	0.21	0.21	0.23	0.15	0.17
Sex	Women	0.01	0.00	0.00	0.00	0.01	0.02	0.03	0.00	-0.01	-0.02

Source: Authors' estimations based on Eurostat (2021c).

TABLE A5: COEFFICIENTS FOR THE GENERALISED ORDERED PROBIT MODEL FOR SELF-RATED HEALTH

		Estimate	Std.Error									
Main Effect	Very Good	-0.44	0.00									
Self-rated	Good	-0.91	0.01									
Ref: Excellent	Fair	-1.48	0.01									
	Poor	-2.02	0.01									
Threshold Coefficients		Threshold										
		1	2	3	4	5	6	7	8	9	10	
	Intercept	-3.88	-3.73	-3.49	-3.14	-2.85	-2.10	-1.84	-1.27	-0.33	0.25	
Age Group	15-29	0.02	-0.03	0.00	0.00	0.01	0.02	0.04	0.04	0.02	0.00	
Ref: 65-70	30-39	-0.01	0.00	0.03	0.04	0.05	0.06	0.06	0.03	-0.04	-0.11	
	40-49	0.11	0.08	0.09	0.07	0.05	-0.01	0.02	0.02	0.03	0.04	
	50-54	0.21	0.20	0.19	0.16	0.15	0.09	0.12	0.11	0.15	0.17	
	55-59	0.35	0.31	0.28	0.25	0.24	0.15	0.19	0.20	0.22	0.27	
	60-64	0.36	0.34	0.31	0.26	0.26	0.15	0.19	0.22	0.27	0.34	
	70-74	0.36	0.34	0.33	0.29	0.27	0.14	0.18	0.21	0.25	0.36	
	74-79	0.38	0.34	0.30	0.26	0.25	0.10	0.15	0.19	0.22	0.35	
	80+	0.44	0.38	0.34	0.27	0.24	0.07	0.12	0.16	0.16	0.29	
Country	Belgium	-0.07	0.05	0.18	0.15	0.15	0.02	0.13	0.29	0.61	0.86	
Ref: Austria	Bulgaria	1.01	1.23	1.44	1.56	1.66	1.54	1.59	1.46	1.07	0.86	
	Switzerland	0.29	0.28	0.27	0.19	0.12	0.00	0.00	0.04	0.06	0.04	
	Cyprus	0.78	0.79	0.80	0.77	0.71	0.67	0.73	0.72	0.59	0.56	
	Czechia	0.20	0.31	0.39	0.41	0.37	0.44	0.46	0.41	0.27	0.21	
	Denmark	0.36	0.47	0.59	0.56	0.48	0.21	0.20	0.11	-0.13	-0.20	
	Estonia	0.11	0.20	0.26	0.37	0.39	0.45	0.54	0.57	0.47	0.34	
	Greece	0.97	1.01	1.02	0.99	1.01	0.91	1.17	1.14	0.89	0.78	
	Spain	0.23	0.39	0.46	0.42	0.39	0.33	0.43	0.49	0.44	0.49	
	Finland	-0.10	-0.06	-0.10	-0.15	-0.19	-0.37	-0.39	-0.29	-0.10	0.40	
	France	0.19	0.22	0.26	0.29	0.34	0.28	0.39	0.52	0.60	0.59	
	Croatia	0.69	0.81	0.95	0.98	0.92	0.91	0.93	0.87	0.63	0.41	
	Hungary	0.30	0.44	0.59	0.75	0.85	0.83	0.97	0.97	0.85	0.75	
	Ireland	0.17	0.33	0.41	0.41	0.37	0.21	0.24	0.13	0.00	-0.06	
	Italy	0.57	0.61	0.60	0.56	0.63	0.38	0.60	0.77	0.66	0.45	
	Lithuania	0.65	0.72	0.82	0.87	0.86	0.80	0.83	0.77	0.49	0.31	
	Luxembourg	0.18	0.26	0.29	0.19	0.12	0.10	0.15	0.25	0.28	0.30	
	Latvia	0.14	0.26	0.37	0.50	0.54	0.56	0.70	0.75	0.59	0.40	
	Netherlands	-0.03	0.13	0.18	0.15	0.10	-0.20	-0.03	0.23	0.51	0.66	
	Norway	0.16	0.27	0.21	0.19	0.18	0.07	0.07	0.08	0.02	-0.04	

Source: Authors' estimations based on Eurostat (2021c).

TABLE A5 (CONT.): COEFFICIENTS FOR THE GENERALISED ORDERED PROBIT MODEL FOR SELF-RATED HEALTH

Threshold Coefficients		Threshold									
		1	2	3	4	5	6	7	8	9	10
	Poland	-0.16	0.03	0.11	0.08	0.05	0.09	0.14	0.11	-0.05	-0.09
	Portugal	0.65	0.74	0.81	0.76	0.68	0.71	0.70	0.58	0.40	0.20
	Romania	0.39	0.47	0.56	0.55	0.58	0.34	0.42	0.47	0.44	0.51
	Serbia	0.46	0.82	0.97	1.18	1.31	1.34	1.46	1.41	1.15	1.04
	Sweden	0.40	0.56	0.51	0.38	0.30	0.15	0.13	0.21	0.15	0.02
	Slovenia	0.46	0.55	0.50	0.49	0.41	0.42	0.49	0.49	0.44	0.33
	Slovakia	0.53	0.62	0.83	0.78	0.72	0.69	0.70	0.56	0.30	0.09
	United Kingdom	0.33	0.46	0.49	0.43	0.40	0.22	0.22	0.23	0.15	0.16
Sex	Women	0.00	-0.02	-0.02	-0.01	0.00	0.01	0.02	-0.01	-0.02	-0.03

Source: Authors' estimations based on Eurostat (2021c).

TABLE A6: COEFFICIENTS FOR THE GENERALISED ORDERED PROBIT MODEL FOR MULTIPLE DIMENSIONS OF HEALTH

Main effect		Self-rated Health				
		Excellent	Very Good	Good	Fair	Poor
No chronic morbidity						
Activity Limitations	No Limitations	ref.	-0.44	-0.86	-1.31	-0.69
	Limited	-0.14	-0.58	-0.97	-1.42	-1.62
	Severely Limited	-0.35	-0.56	-1.09	-1.48	-2.21
chronic morbidity						
Activity Limitations	No Limitations	0.03	-0.37	-0.80	-1.32	-1.78
	Limited	-0.27	-0.53	-0.97	-1.45	-1.78
	Severely Limited	-0.72	-0.74	-1.11	-1.60	-2.13

Threshold coefficients		Threshold									
		1	2	3	4	5	6	7	8	9	10
	Intercept	-3.93	-3.77	-3.53	-3.18	-2.88	-2.13	-1.86	-1.29	-0.35	0.24
Age Group	15-29	0.01	-0.04	0.00	0.00	0.00	0.02	0.03	0.04	0.02	-0.01
Ref: 65-70	30-39	-0.03	-0.03	0.01	0.01	0.03	0.03	0.03	0.01	-0.07	-0.14
	40-49	0.12	0.09	0.09	0.07	0.06	0.00	0.02	0.03	0.03	0.04
	50-54	0.22	0.21	0.20	0.17	0.16	0.09	0.12	0.12	0.15	0.18
	55-59	0.36	0.31	0.29	0.26	0.25	0.16	0.19	0.21	0.22	0.27
	60-64	0.37	0.35	0.32	0.27	0.27	0.16	0.20	0.23	0.27	0.35
	70-74	0.37	0.35	0.34	0.30	0.28	0.15	0.19	0.21	0.26	0.37

TABLE A6 (CONT.): COEFFICIENTS FOR THE GENERALISED ORDERED PROBIT MODEL FOR MULTIPLE DIMENSIONS OF HEALTH

Threshold coefficients		Threshold									
		1	2	3	4	5	6	7	8	9	10
	74-79	0.39	0.35	0.31	0.27	0.26	0.11	0.16	0.20	0.22	0.36
	80+	0.45	0.39	0.36	0.28	0.25	0.08	0.13	0.16	0.16	0.30
Country	Belgium	-0.06	0.06	0.19	0.16	0.16	0.03	0.14	0.30	0.62	0.87
Ref: Austria	Bulgaria	1.07	1.29	1.50	1.61	1.72	1.59	1.63	1.50	1.10	0.89
	Switzerland	0.30	0.29	0.28	0.20	0.12	0.01	0.00	0.05	0.06	0.05
	Cyprus	0.79	0.81	0.81	0.79	0.73	0.68	0.74	0.74	0.60	0.57
	Czechia	0.23	0.34	0.42	0.44	0.40	0.47	0.49	0.44	0.29	0.23
	Denmark	0.37	0.48	0.61	0.57	0.50	0.22	0.21	0.13	-0.12	-0.18
	Estonia	0.12	0.22	0.27	0.38	0.40	0.46	0.56	0.58	0.48	0.34
	Greece	0.98	1.03	1.03	1.00	1.02	0.92	1.18	1.15	0.89	0.78
	Spain	0.27	0.43	0.50	0.45	0.43	0.36	0.46	0.52	0.46	0.51
	Finland	-0.09	-0.05	-0.09	-0.14	-0.18	-0.36	-0.38	-0.28	-0.08	0.41
	France	0.21	0.24	0.28	0.31	0.36	0.30	0.41	0.54	0.62	0.61
	Croatia	0.72	0.83	0.98	1.01	0.95	0.93	0.95	0.89	0.65	0.42
	Hungary	0.33	0.47	0.62	0.79	0.88	0.86	1.00	1.00	0.88	0.77
	Ireland	0.18	0.34	0.43	0.42	0.39	0.22	0.25	0.15	0.02	-0.05
	Italy	0.59	0.63	0.62	0.58	0.65	0.39	0.61	0.78	0.68	0.47
	Lithuania	0.70	0.77	0.87	0.92	0.91	0.84	0.87	0.80	0.53	0.33
	Luxembourg	0.20	0.27	0.30	0.21	0.14	0.11	0.17	0.26	0.29	0.31
	Latvia	0.17	0.29	0.40	0.52	0.56	0.58	0.71	0.76	0.61	0.41
	Netherlands	-0.03	0.13	0.16	0.14	0.09	-0.21	-0.03	0.22	0.51	0.67
	Norway	0.20	0.31	0.26	0.23	0.22	0.09	0.09	0.11	0.05	-0.02
	Poland	-0.11	0.07	0.15	0.13	0.10	0.13	0.18	0.14	-0.02	-0.07
	Portugal	0.69	0.78	0.85	0.80	0.72	0.75	0.73	0.61	0.43	0.23
	Romania	0.41	0.48	0.57	0.56	0.59	0.35	0.43	0.47	0.44	0.52
	Serbia	0.55	0.91	1.06	1.26	1.39	1.40	1.52	1.47	1.19	1.09
	Sweden	0.47	0.63	0.57	0.44	0.35	0.20	0.17	0.25	0.19	0.05
	Slovenia	0.48	0.57	0.51	0.50	0.42	0.43	0.50	0.50	0.45	0.34
	Slovakia	0.53	0.63	0.83	0.78	0.72	0.69	0.70	0.56	0.30	0.09
	United Kingdom	0.32	0.46	0.49	0.43	0.40	0.22	0.22	0.23	0.15	0.16
Sex	Women	0.00	-0.02	-0.02	-0.01	0.00	0.01	0.01	-0.01	-0.02	-0.03

Source: Authors' estimations based on Eurostat (2021c).

TABLE A7: LIFE EXPECTANCY (LE), DISABILITY ADJUSTED HEALTH EXPECTANCY (DALE), HEALTH EXPECTANCY (HE) AND WELL-BEING ADJUSTED HEALTH EXPECTANCY (WAHE) AT AGE 15 FOR THE THREE DIMENSIONS OF HEALTH, MEN, 2018

Country	LE	DALE	Chronic Diseases		Activity Limitations		Self-rated Health		Multiple Dimensions	
			HE	WAHE	HE	WAHE	HE	WAHE	HE	WAHE
Latvia	55.5	49.2	38.1	53.7	37.6	53.0	5.2	46.8	6.2	46.8
Lithuania	56.4	49.0	40.1	54.5	42.4	54.1	9.2	46.6	6.8	46.4
Bulgaria	57.2	48.3	46.9	56.1	50.2	56.1	15.9	50.0	16.0	50.3
Romania	57.4	50.0	49.1	56.6	45.7	55.7	20.1	51.5	19.1	51.4
Hungary	58.1	50.2	39.5	56.3	46.4	56.4	13.4	51.0	13.3	51.2
Serbia	59.0	50.9	43.8	57.6	52.9	58.1	16.1	51.8	15.7	52.1
Poland	59.2	51.1	39.9	57.3	46.9	57.2	12.3	50.7	13.3	51.1
Estonia	59.3	51.0	33.6	56.6	38.8	56.3	7.3	50.0	6.1	49.9
Slovakia	59.5	51.2	43.0	57.4	42.0	56.3	15.7	50.5	15.3	50.3
Croatia	60.5	52.2	41.9	58.3	43.0	57.5	21.4	51.9	21.8	52.0
Czechia	61.5	52.3	44.0	59.7	48.4	59.5	20.0	54.1	17.9	53.9
Portugal	63.7	54.9	40.7	61.1	46.3	60.7	8.3	51.9	8.9	52.7
Slovenia	63.8	53.9	41.1	61.3	43.1	60.5	17.3	55.1	14.8	54.7
Finland	64.3	54.9	36.7	61.7	45.4	61.8	15.1	57.0	11.0	56.7
Denmark	64.5	55.4	47.1	62.3	47.8	61.6	18.5	55.5	16.2	55.2
Greece	64.7	55.1	51.0	63.2	51.0	62.2	33.5	58.6	33.1	58.5
Austria	64.8	55.3	42.3	62.3	43.5	61.4	22.3	57.2	20.3	57.2
Belgium	64.8	54.7	50.0	63.5	49.8	62.8	20.5	58.9	19.7	58.8
United Kingdom	64.9	54.6	38.6	62.0	48.3	61.9	22.8	57.3	18.7	56.7
France	65.2	56.4	41.0	62.9	49.7	62.8	16.8	57.6	14.8	57.7
Luxembourg	65.6	56.3	49.0	63.8	48.1	62.8	16.1	57.2	15.2	57.3
Netherlands	65.7	56.4	45.8	63.9	46.8	63.3	18.9	59.4	16.6	59.1
Ireland	65.7	55.8	47.6	63.4	54.3	63.8	29.5	59.7	27.4	59.7
Spain	66.0	56.7	45.8	64.1	53.9	64.3	17.3	58.7	16.3	58.7
Sweden	66.1	57.0	44.8	63.6	58.9	64.8	21.3	58.0	18.1	58.2
Cyprus	66.2	55.4	37.5	63.0	47.9	63.1	28.3	59.1	26.9	59.1
Norway	66.4	56.3	44.3	63.9	56.8	64.7	19.3	58.4	16.5	58.5
Italy	66.5	56.3	57.6	65.6	52.7	64.5	12.4	58.0	11.9	58.0
Switzerland	67.3	57.3	45.3	64.8	48.4	64.4	25.1	60.5	21.6	60.2

Note: Countries ordered by LE

Source: Authors' estimations based on Eurostat (2021c), Eurostat (2021) and GBD (2020).

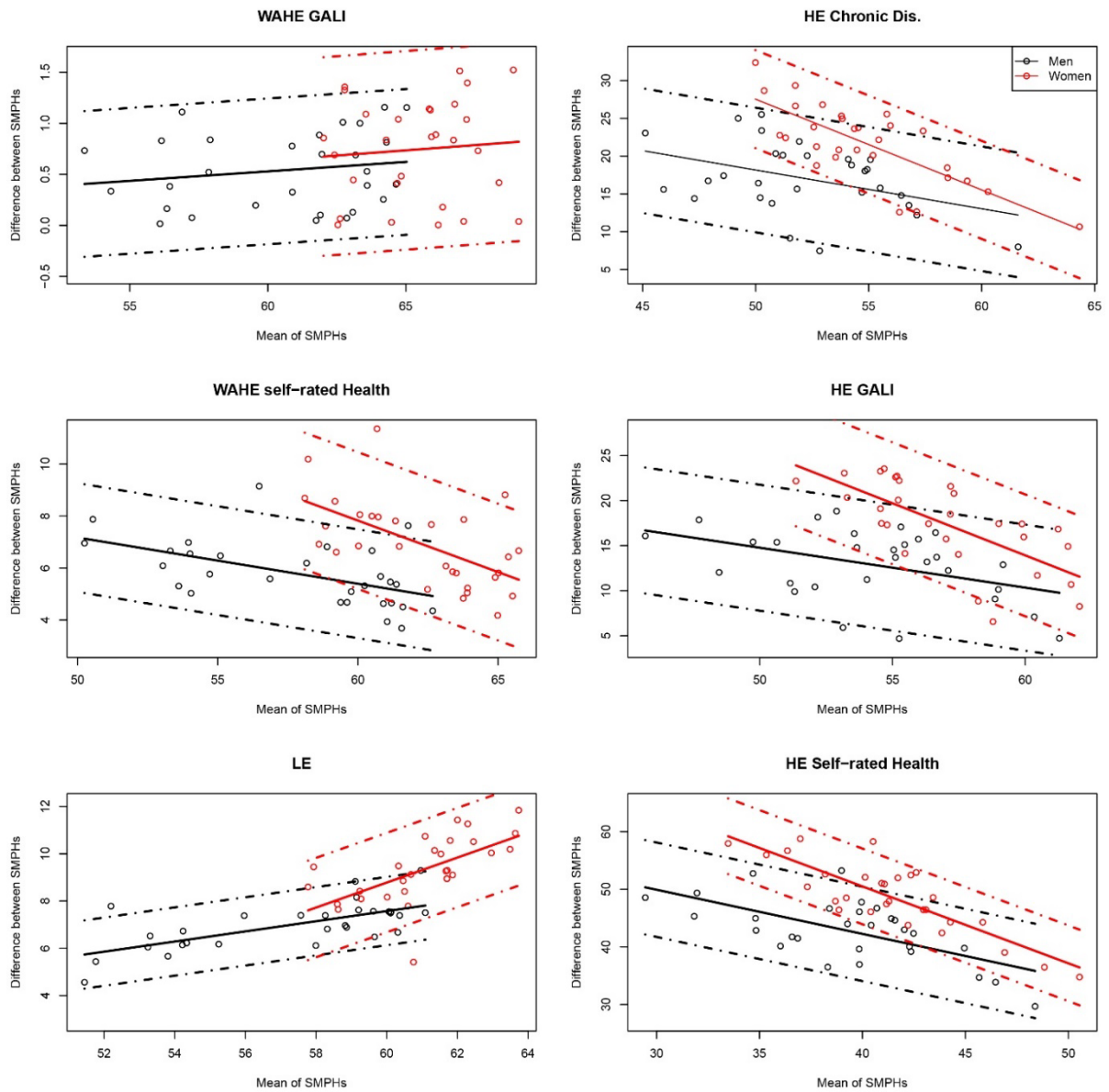
TABLE A8: LIFE EXPECTANCY (LE), DISABILITY ADJUSTED HEALTH EXPECTANCY (DALE), HEALTH EXPECTANCY (HE) AND WELL-BEING ADJUSTED HEALTH EXPECTANCY (WAHE) AT AGE 15 FOR THE THREE DIMENSIONS OF HEALTH, WOMEN, 2018

Country	LE	DALE	Chronic Diseases		Activity Limitations		Self-rated Health		Multiple Dimensions	
			HE	WAHE	HE	WAHE	HE	WAHE	HE	WAHE
Serbia	63.9	53.5	43.3	62.1	55.5	62.8	15.7	55.2	14.8	55.6
Bulgaria	64.2	53.2	50.1	62.7	53.8	62.6	14.7	55.0	14.4	55.3
Romania	64.8	55.0	50.8	63.4	46.1	62.1	17.4	56.6	16.9	56.5
Hungary	64.9	54.7	40.1	62.5	48.4	62.5	12.1	55.9	12.2	56.1
Latvia	65.1	54.8	39.7	62.4	40.3	61.6	4.5	53.8	4.8	53.7
Lithuania	66.1	55.2	42.1	63.3	45.9	62.9	7.3	53.1	5.6	53.0
Slovakia	66.3	58.0	43.6	63.5	43.1	62.1	15.0	54.9	14.7	54.7
Croatia	66.8	55.9	43.3	64.1	45.0	63.0	20.3	56.0	19.3	55.9
Poland	67.1	56.7	40.6	64.5	50.5	64.5	11.9	56.5	11.3	56.7
Czechia	67.3	56.0	44.1	64.9	49.1	64.5	17.4	58.1	15.7	57.8
Estonia	67.9	56.3	36.0	64.7	41.7	63.9	8.0	56.7	7.2	56.5
Denmark	68.3	56.1	45.1	65.2	45.2	64.2	17.3	57.4	14.2	57.0
United Kingdom	68.4	55.6	38.4	65.1	47.6	64.6	22.6	59.9	18.2	59.2
Netherlands	68.7	56.5	44.4	66.5	43.8	65.6	15.6	61.3	13.3	60.9
Belgium	69.3	56.3	51.0	67.7	50.3	66.7	19.1	62.1	18.5	61.9
Austria	69.3	57.3	42.7	66.4	43.8	65.3	22.1	60.6	20.4	60.5
Ireland	69.4	56.5	49.9	67.1	56.4	67.1	30.6	62.9	27.6	62.8
Sweden	69.5	57.2	42.5	66.2	57.9	67.4	19.7	60.1	16.8	60.2
Slovenia	69.6	55.7	41.1	66.5	42.9	65.3	15.4	58.8	12.3	58.3
Norway	69.7	57.0	41.4	66.3	54.6	67.1	19.9	60.5	17.1	60.5
Finland	69.7	57.2	33.8	66.2	42.9	66.2	14.1	61.3	9.8	60.9
Greece	69.8	56.7	52.6	67.9	52.0	66.5	33.2	62.1	32.7	61.9
Portugal	69.9	57.1	39.6	66.4	44.1	65.5	7.6	55.0	8.3	55.4
Luxembourg	70.0	57.2	49.2	67.7	46.9	66.2	15.7	59.8	13.3	59.6
Cyprus	70.1	56.3	37.1	66.4	47.9	66.2	27.4	61.4	25.5	61.3
Italy	70.9	57.8	59.0	69.7	52.8	68.1	11.4	60.8	11.3	60.8
Switzerland	71.1	57.9	43.9	68.0	46.4	67.2	23.7	63.1	20.0	62.7
France	71.3	58.4	43.0	68.6	51.2	68.2	16.1	62.2	13.9	62.0
Spain	71.6	58.2	45.7	69.1	54.2	69.1	16.2	62.4	15.2	62.5

Note: Countries ordered by LE

Source: Authors' estimations based on Eurostat (2021c), Eurostat (2021) and GBD (2020).

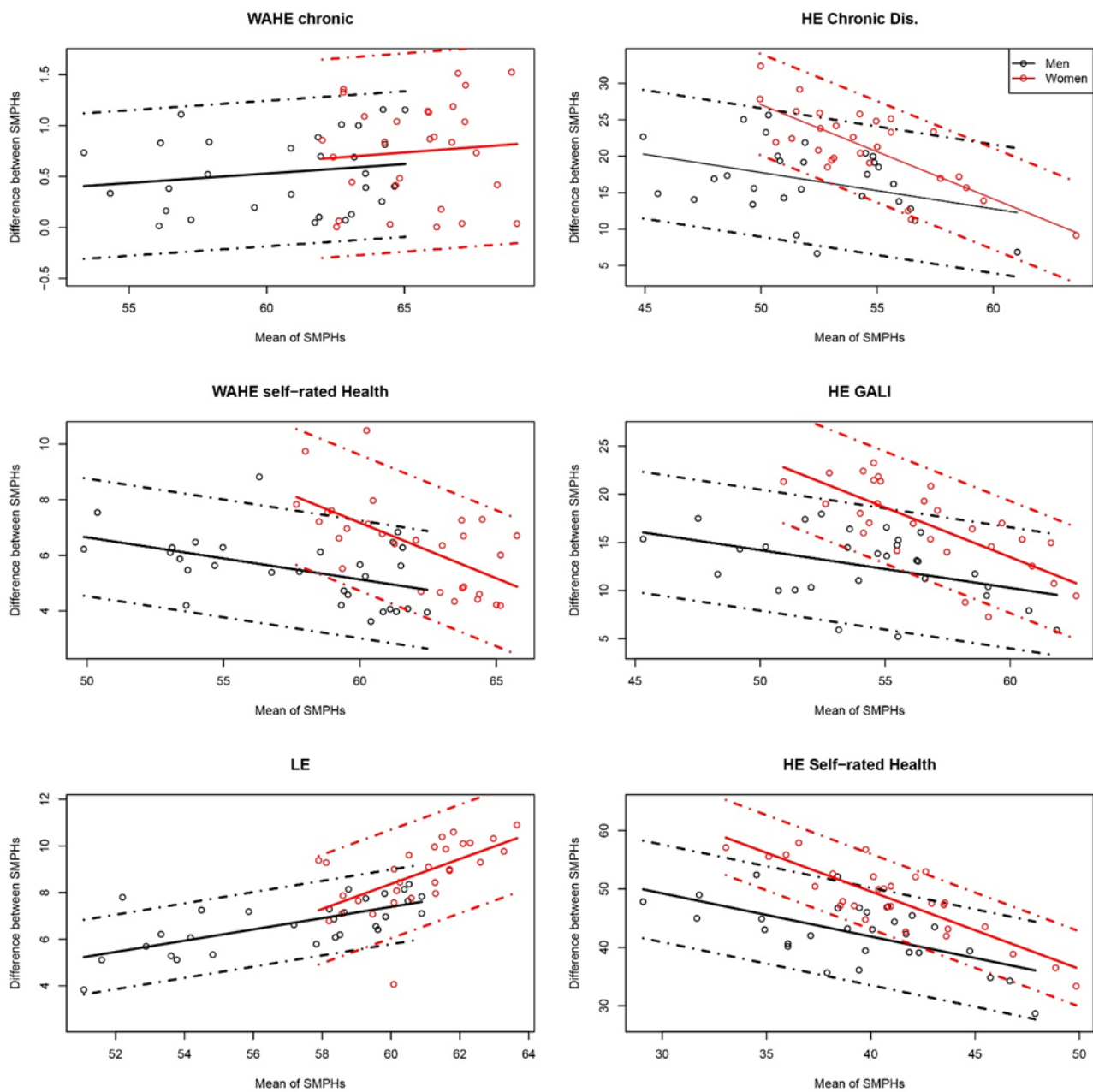
FIGURE A3: BLAND-ALTMAN PLOTS FOR AGREEMENT BETWEEN WAHE BASED ON CHRONIC MORBIDITY AND OTHER SMPHS IN 29 COUNTRIES BY SEX, 2018.



Note: Dots represent country observations; solid lines represent the linear model; dashed lines represent the 95% confidence intervals.
Source: Authors' estimations based on Eurostat (2021c), Eurostat (2021) and GBD (2020).

○

FIGURE A4: BLAND-ALTMAN PLOTS FOR AGREEMENT BETWEEN WAHE BASED ON GALI AND OTHER SMPHs IN 29 COUNTRIES BY SEX, 2018.



Note: Dots represent country observations; solid lines represent the linear model; dashed lines represent the 95% confidence intervals.

Source: Authors' estimations based on Eurostat (2021c), Eurostat (2021) and GBD (2020).

A.2 GENERALISED ORDERED PROBIT MODEL FOR THE WELL-BEING WEIGHTS

Well-being is a latent continuous variable, denoted by WB^* . When answering the question "Overall, how satisfied are you with your life?" respondents assess their well-being using a discrete scale, ranging from 1 (maximum unsatisfied) to 10 (maximum satisfied). The collected responses measure observed well-being, which is denoted by WB^o .

Latent well-being is assumed to be related to health status and other individual- and contextual-level characteristics as:

$$WB^* = \beta_0 + H\beta_1 + \mathbf{X}\beta_2 + \varepsilon$$

where H is the health status of an individual; \mathbf{X} is a vector of other individual characteristics; β are vectors of coefficients, and ε is a normal distributed random error term.

Since an ordered scale is applied to the well-being question, observed well-being (WB^o) is a categorical ordered response variable. It is assumed to be related to the latent well-being variable WB^* as:

$$WB^o = i \leftrightarrow c_{i-1} < WB^* \leq c_i, \quad i = 1, \dots, 10$$

where i denotes response categories (from 1 to 10 for observed well-being in this study); c_i are threshold levels, where $c_0 = -\infty$ and $c_n = \infty$, and the remaining threshold levels are specified as:

$$c_i = \delta_i \text{ for } i = 1, \dots, 10$$

where δ_i are coefficients to be estimated in the ordered link model (with the link being, e.g., logit or probit). While thresholds are constant in most well-known ordered link models, generalised ordered models allow thresholds to be dependent on covariates (i.e., sex, age, country of residence) and, hence, they are modelled separately (Greene et al., 2014). Thresholds are also ordered, which means that a higher threshold will always have a higher value than a lower threshold (King et al., 2004).

Following the notations above, the probability of observing an i level of well-being WB^o for individual j is given by:

$$\begin{aligned} P(WB_j^o = i) &= P(c_{i-1} \leq WB^* \leq c_i) = P(c_{i-1} - \beta_0 + H_j\beta_1 + \mathbf{X}_j\beta_2 \leq \varepsilon_j \leq c_i - \beta_0 + H_j\beta_1 + \mathbf{X}_j\beta_2) \\ &= F(c_i - \beta_0 + H_j\beta_1 + \mathbf{X}_j\beta_2) - F(c_{i-1} - \beta_0 + H_j\beta_1 + \mathbf{X}_j\beta_2) \end{aligned}$$

where F denotes the cumulative distribution function for ε_j , $c_0 = c_{i-1} - \infty$ and $c_n = \infty$.

In the logit link model, the cumulative distribution function for ε_j is specified by cumulative logistic distribution. In the probit link model, ε_j are assumed to be independent and follow a normal distribution with $E(\varepsilon_j) = 0$ and $D(\varepsilon_j) = \sigma$. Hence, based on the above probability for the probit link, we can specify the log-likelihood function as:

$$\log L = \sum_{j=1}^N \sum_{i=1}^9 z_{ij} \log[\Phi(c_i - \beta_0 + H_j\beta_1 + \mathbf{X}_j\beta_2) - \Phi(c_{i-1} - \beta_0 + H_j\beta_1 + \mathbf{X}_j\beta_2)]$$

where $z_{ij} = 1$ if $WB_j^o = i$ and 0 otherwise. Φ denotes a cumulative function of a normal distribution. The log-likelihood function is maximised to derive the estimators of coefficients β and the cutpoints for the latent distribution of well-being c_i are derived.

Well-being weights for each decreased health state are equal to the respective, normalised regression coefficient (β_1). We follow the approach by Cutler et al. (1997), who estimated QALY from a normal ordered probit model as a proportional decrease in self-rated health resulting from a health limitation. QALY weights for disease i are calculated as:

$$QALY_i = \frac{\hat{\beta}_i}{\delta_1 - \delta_n}$$

where $\hat{\beta}_i$ is the coefficient for the disease i and δ_1, δ_n are the lowest and the highest thresholds for the latent health variable from the estimated models. The health-related well-being weights in WAHE are based on the estimated coefficients from the generalised ordered probit model:

$$w_i^{jkl} = 1 - \frac{\hat{\beta}_i}{\delta_1^{jkl} - \delta_{10}^{jkl}}$$

where w_i^{jkl} is a health-related well-being weight for the state of health i for sex j , age group k and country l ; $\hat{\beta}_i$ is the coefficient for the health state i ; $\delta_1^{jkl}, \delta_{10}^{jkl}$ are the lowest and the highest thresholds for the latent health variable for sex j ; age group k and country l estimated as:

$$\delta_1^{jkl} = \delta_1 + \delta_1^j + \delta_1^k + \delta_1^l \text{ and } \delta_{10}^{jkl} = \delta_{10} + \delta_{10}^j + \delta_{10}^k + \delta_{10}^l$$

where δ_1 and δ_{10} are the lowest and the highest base thresholds; δ_1^j and δ_{10}^j are the corresponding shifts in the lowest and the highest thresholds for women (sex $j=2$, for men the latent variable thresholds are the base thresholds); δ_1^k and δ_{10}^k are the corresponding shifts in the lowest and the highest thresholds for the age group k (the reference age group in the study is 65 to 70 years); δ_1^l and δ_{10}^l are the corresponding shifts in the lowest and the highest thresholds for the country l (Austria is the study reference country).

A.3 BLAND-ALTMAN PLOTS WITH A LINEAR TREND TO ASSESS AGREEMENT BETWEEN WAHE AND OTHER SMPHS

Following the guidelines developed by Kottner et al. (2011) for reporting reliability and agreement studies, Bland-Altman plots with limits of agreement are a recommended method for analysing the agreement of continuous variables.

In Bland-Altman plots the difference between two variables for each observation point is plotted against the mean. If the differences are normally distributed, $(1 - \alpha)\%$ of the observations will lie between the limits of agreement of $\bar{d} - z_\alpha s$ and $\bar{d} + z_\alpha s$, where \bar{d} is the mean and s is the standard deviation of differences between variables, and z_α is the upper $(1 - \frac{\alpha}{2})$ critical value for the standard normal distribution (Bland and Altman, 1986).

In this study, a linear regression model with the dependent variable being the difference between the measurements and the independent variable being the mean of the measurements can be fitted. Assuming that the residuals follow a normal distribution with mean zero and variance σ^2 , the expected value of the difference between the methods can be estimated as:

$$\hat{d} = b_0 + b_1 a.$$

In this case, the $(1 - \alpha)$ limits of agreement are $\hat{d} \pm z_\alpha \sqrt{\frac{2}{\pi}} \hat{R}$, where $\hat{R} = c_0 + c_1 a$ are the absolute values of residuals ($\hat{R} = d - \hat{d}$) regressed on the mean of the measurements (a). In case no significant relationship between \hat{R} and a is observed, \hat{R} is simply estimated as a standard deviation of the residuals (Bland and Altman, 1999).

A.4 INTRACLASS CORRELATION COEFFICIENTS TO ASSESS RELIABILITY OF WAHE AND OTHER SMPHS

Following the guidelines developed by Kottner et al. (2011) for reporting reliability and agreement studies, the intraclass correlation coefficient (ICC) is the preferred statistical method for analysing interrater reliability for continuous variables. No single ICC estimate exists, and the preferred method depends on the study design and unit of analysis. For example, Shrout and Fleiss (1979) outline six different ICC estimates, while McGraw and Wong (1996) describe eight types. For assessing the reliability of the SMPHs, we apply a two-way mixed effects model: The SMPHs are considered a fixed effect, because population health is only assessed by this group of measures. However, we are interested in the reliability of this given group and do not extend our results to any additional population health index. However, the list of countries can be extended, so countries are considered a random effect in our model. Next, we need to choose a type of measurement between the mean value of the indices and single indices measurement; in this study, we apply the former. Using the mean measurement across all SMPHs to study reliability allows us to answer the research question of whether this group of measures can reliably discern population health levels across the study countries. Finally, by excluding the indices one by one and comparing the new ICC to that of the whole group, we examine whether incorporating this additional information to a given SMPH facilitates distinguishing between population health levels in the studied countries. The method described above used to derive ICC in this study is denoted by $ICC(C, k)$, as classified by McGraw and Wong (1996) and $ICC(3,k)$, as classified by Shrout and Fleiss (1979). Regarding the relationship between these two standards of the classification, see also Koo and Li (2016). In Shrout and Fleiss' (1979) classification (1979), $ICC(3, k)$ are applied to assess reliability in situations where, "each target is rated by each of the same k judges, who are the only judges of interest" (p.421). In our study "judges" denotes SMPHs and "targets" denotes countries. Hence, the value of index i ($i=1,2,\dots,k$) in the two-way mixed model for country j ($j=1,2,\dots,n$) is equal to:

$$x_{ij} = \mu + a_i + b_j + v_{ij}$$

where μ is the overall mean of all values of the indices across the countries; b_j is the difference from μ of the j^{th} country true score (e.g., mean of all scores), and is assumed to be normally distributed with mean=0 and variance= σ_j^2 ; a_i is the distance in the mean of i^{th} index from μ .

In the $ICC(3, k)$ model:

$$\sum_{i=1}^k a_i = 0;$$

v_{ij} denotes the random error in the i^{th} index for country j .

As demonstrated by Shrout and Fleiss (1979) and McGraw and Wong (1996), and based on the above model specification, the estimator for the ICC is:

$$\widehat{ICC}(3, k) = \frac{MS_R - MS_E}{MS_R}$$

where MS_R is the mean square of differences in the index values between the countries, and MS_E is the mean residual differences, i.e., the mean square error of measurement from the repeated measures analysis of variance (ANOVA).

REFERENCES

- Berger, N., Van Oyen, H., Cambois, E., Fouweather, T., Jagger, C., Nusselder, W., & Robine, J. M. (2015). Assessing the validity of the Global Activity Limitation Indicator in fourteen European countries. *BMC Medical Research Methodology*, *15*(1), 1-8.
- Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*(8476), 307-310.
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135-160.
- Cutler, D. M., Richardson, E., Keeler, T. E., & Staiger, D. (1997). Measuring the health of the US population. *Brookings Papers on Economic Activity. Microeconomics*, *1997*, 217-282.
- Greene, W., Harris, M. N., Hollingsworth, B., & Weterings, T. A. (2014). Heterogeneity in ordered choice models: A review with applications to self-assessed health. *Journal of Economic Surveys*, *28*(1), 109-133.
- King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, *98*(1), 191-207.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163.
- Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., C. Roberts, M. Shoukri, & Streiner, D. L. (2011). Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *International Journal of Nursing Studies*, *48*(6), 661-671.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*(1), 30.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420.