18 Saskatoon Public Health Observatory *Evidence, Action, Equity: Making Population Health Information Count.* Saskatoon: Saskatoon Public Health Observatory, 2015. Available at: https://www.saskatoonhealthregion.ca/locations_services/Services/Health-Observatory (October 2015, date last accessed).

19 Pooransingh S, Misir A, Ramdath DD, et al. An assessment of the barriers and facilitators for a national public health observatory in Trinidad and Tobago. *West Indian Medical Journal* 2014;63(Suppl 2):69–70.

20 Kebede D, Zielinski C, Mbondji PE, et al. The African Health Observatory and national health observatories as platforms for strengthening health information systems in sub-Saharan Africa. *J R Soc Med* 2014;107(1): suppl. 6–9.

21 Emerson E, Glover G, Turner S, et al. Improving health and lives: the Learning Disabilities Public Health Observatory. *Adv Mental Health Intellect Disabil* 2012;6:26–32.

22 Young TK, Chatwood S, Rawat R. Data brief from the Circumpolar Health Observatory. *Int J Circumpolar Health* 2010;69:404–8.

23 Gattini CH. Implementing National Health Observatories: Operational Approach and Strategic Recommendations. Santiago de Chile: Office of the Pan-American Health Organization & WHO Representation in Chile, 2009.

24 Foro IberoAmericano de Observatorios de Calud. Report. 2008. Santander, Calabria (Spain). Available at: http://www.opss.pt/sites/opss.pt/files/Informe_Foro_OSPC.pdf (October 2015, date last accessed).

25 National Collaborating Centre for Determinants of Health. *Public Health Observatories: Learning from Our World Neighbours*. Antigonish, NS: NCCDH, 2011.

26 Pourmalek F. National health observatories: need for stepped-up action. *The Health* 2012;3:63–4.

....

....

Self-perceived health in older Europeans: Does the choice of survey matter?

Simone Croezen[1,2], Alex Burdorf[1], Frank J. van Lenthe[1]

1 Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
2 The Netherlands Institute for Social Research | SCP, The Hague, The Netherlands

Correspondence: Frank J. van Lenthe, Department of Public Health, Erasmus MC, University Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands, Tel: +31 10 7044220, Fax: +31 10 7038475, e-mail: f.vanlenthe@erasmusmc.nl

Background: Cross-national comparisons of health in European countries provide crucial information to monitor health and disease within and between countries and to inform policy and research priorities. However, variations in estimates might occur when information from cross-national European surveys with different characteristics are used. We compared the prevalence of very good or good self-perceived health across 10 European countries according to three European surveys and investigated which survey characteristics contributed to differences in prevalence estimates. **Methods:** We used aggregate data from 2004 to 2005 of respondents aged 55–64 years from the European Union Statistics on Income and Living Conditions (EU-SILC), the Survey of Health, Ageing and Retirement in Europe (SHARE) and the European Social Survey (ESS). Across the surveys, self-perceived health was assessed by the same question with response options ranging from very good to very bad. **Results:** Despite a good correlation between the surveys (intraclass correlation coefficient: 0.77), significant differences were found in prevalence estimates of very good or good self-perceived health. The survey response, sample size and survey mode contributed statistically significantly to the differences between the surveys. Multilevel linear regression analyses, adjusted for survey characteristics, showed a higher prevalence for SHARE (+6.96, 95% CIs: 3.14 to 10.8) and a lower prevalence (−3.12; 95% CIs: −7.11 to 0.86) for ESS, with EU-SILC as the reference survey. **Conclusion:** Three important health surveys in Europe showed substantial differences for presence of very good or good self-perceived health. These differences limit the usefulness for direct comparisons across studies in health policies for Europe.

....

Introduction

Existing population-based surveys carried out across several European countries provide crucial information to monitor health and disease,[1,2] to describe health inequalities within and across countries,[3–5] and to inform policy and research priorities.[6,7] Currently, there are three important surveys in Europe, which provide cross-nationally comparable information on health of European adults and have their information publicly available: EU Statistics on Income and Living Conditions (EU-SILC), the Survey of Health, Ageing and Retirement in Europe (SHARE) and the European Social Survey (ESS).[8–10] EU-SILC, SHARE and ESS have been performed in several European countries, and their information on health status has contributed to several important publications.[3,5,11–19] However, differences between surveys in a multitude of factors, such as study size, sampling strategy and collection methods, can affect the representation of the target population and the measurement of the survey estimate, leading to different estimates of health and thus to different interpretations of the population health status.[20]

Self-perceived health is a widely used indicator for health and the question used to assess this indicator was phrased identically in these three European surveys.[21,22] It deals with the subjective assessment that a person makes about one's own health state and serves as an independent predictor for morbidity, health service use and mortality and is often used as an indicator to compute healthy life expectancy.[23,24] Initial comparisons between EU-SILC, SHARE and ESS showed overall systematic differences in prevalence of self-perceived health, but insight into the determinants of these differences is lacking.[25]

Several survey characteristics have been linked to observed differences in health estimates and thus to different interpretations of the population health status, such as sampling strategy, response and survey mode.[20] The influence of survey modes is more frequently discussed than other survey characteristics, and comparisons

between the different modes have been related to prevalence estimates of health.[26] Concerning self-perceived health, telephone interviewing or using a self-completion mode leads to less positive ratings of self-perceived health than face-to-face interviews.[26,27] Differences between surveys in the used indicators may also lead to incomparable results. For instance, measures of self-perceived health are not directly comparable when the exact wording of the question differs, when different response options are used, or when the question order differs between questionnaires.[28–30]

Potential disagreement in prevalence estimates of health between European surveys could limit the usefulness for direct comparisons across studies in health policies for Europe. Therefore, understanding differences between EU-SILC, SHARE and ESS in terms of their estimation of self-perceived health will provide valuable insights in the comparability of health status information in Europe. We aim to investigate the agreement between the surveys by describing systematic differences in self-perceived health across countries, and investigating the role of survey characteristics such as survey response, sample size, sampling strategy and survey mode as possible determinants of observed differences between the surveys.

Methods

Surveys

EU-SILC is the reference source for comparative statistics on income distribution and social inclusion in the European Union (EU).[31] EU-SILC was launched in 2003 and provides annually collected data based on nationally representative probability samples of the population residing in private households. Only household members aged 16 years and older were interviewed. SHARE is a multidisciplinary and cross-national panel database on health, socio-economic status and social and family networks.[8] Since the baseline survey in 2004–05, four additional waves were conducted every 2 years. Based on probability samples in all participating countries, SHARE represents the non-institutionalized population aged 50 years and older (born in 1954 or earlier). ESS is a social survey aimed at investigating Europe's changing institutions and the attitudes, beliefs and behaviour patterns of the European population.[9] It is a biannual cross-sectional survey and covers more than thirty European countries. The first wave was fielded in 2002–03. ESS used random probability samples based on full coverage of the eligible residential populations aged 15 years and older.

To avoid hampering the comparability between the surveys because of potential selection bias due to loss to follow-up in SHARE, we focussed our comparisons on information from the baseline measurement in SHARE (2004–05). This corresponded to the cross-sectional information from ESS wave 2 (2004–05), and EU-SILC 2004 and EU-SILC 2005. EU-SILC 2005 was preferred, as it was implemented in more countries. Across the surveys, 10 countries were commonly included: Austria (AT), Belgium (BE), Germany (DE), Denmark (DK), Greece (GR), Spain (ES), France (FR), Italy (IT), Netherlands (NL) and Sweden (SE) (supplementary files, Table S1). For each country separately, the response, realised sample size, sampling strategy and survey mode are presented in the supplementary files (Tables S2–S5). These study characteristics were based on the total population of the surveys in 2004–05. In brief, all three surveys used probability sampling as their primary sampling strategy. For most countries, the response was between 60 and 69%. The largest sample sizes were obtained in EU-SILC, with at least 5000 respondents for each country, whereas the country-specific sample sizes in SHARE and ESS were between 1000 and 4000 respondents. SHARE and ESS predominantly used computer-assisted personal interview (CAPI) to collect the information from their respondents. In EU-SILC, survey modes varied between and even within countries (table 1).

Self-perceived health

EU-SILC, SHARE and ESS included the same measure of self-perceived health, based on the question: "How is your health in general?". Also, similar response options were used across the surveys: (i) very good; (ii) good; (iii) fair; (iv) bad; (v) very bad. This assessment of self-perceived health was proposed by the WHO[21] and adopted by the Minimum European Health Module (MEHM).[22] For our current analysis and in line with common practice in presenting the prevalence rates in research and policy documents, response options for self-perceived health have been dichotomized across all surveys, with the response categories 'very good' and 'good' into one 'very good or good' category.

Statistical analysis

In each survey information was retrieved on self-perceived health among respondents aged 55–64 years for 10 countries. We focussed on the age group 55–64 years for the comparability of self-perceived health between the surveys. For EU-SILC, population-weighted aggregate data or self-perceived health were only publicly available by 10-year age groups through the Eurostat website.[32] We calculated comparable group-level information based on individual-level data, which were available after registration on their project websites for SHARE (www.share-project.org) and for ESS (www.europeansocialsurvey.org).[8,33–35] Like in EU-SILC, weights were applied in SHARE and ESS to represent the national populations based distributions of age and gender. First, weighted prevalence estimates of very good or good self-perceived health by gender and educational level were described for each survey. All three surveys followed the 1997 International Standard Classification of Education ISCED-97[36] and educational level was presented by three groups: (i) pre-primary, primary and lower secondary education, first and second stage of basic education (ISCED 0-2); (ii) upper and post-secondary education (ISCED 3-4); (iii) first and second stage of tertiary education (ISCED 5-6). Second, we established whether systematic differences in self-perceived health between surveys were present across countries by using Bland and Altman plots. The Bland and Altman technique enabled to quantify the agreement in self-perceived health between surveys for each country separately, by calculating the mean difference between two surveys, and 95% limits of agreement as the mean difference ±1.96 SD.[37] The mean difference reflects the systematic difference between the surveys. The smaller the range between the two limits of agreement, the better the agreement is. Third, the intraclass correlation coefficient (ICC) was calculated as an indication of the correlation of the prevalence estimates of self-perceived health between the surveys within a country. The total variance is reflected by variations between countries and by variations within countries. The ICC is the proportion of total variance that is attributed to differences between countries. As the surveys are nested within the countries, the variance within countries is reflected by variance between the surveys. Thus, a large ICC implies substantial differences between countries and therefore relatively small differences between surveys within countries. Finally, we investigated the role of survey characteristics as possible determinants of disagreement between surveys using multi-level modelling with countries as the higher level, based on the assumption that the surveys were clustered within each country. The prevalence of very good or good self-perceived health (in percentages) was our continuous dependent variable, and regressed on a variable including a category for each survey and several other variables, each representing a specific characteristic of the survey, i.e. response (continuous variable), sample size (continuous variable), sampling strategy (simple random sampling; multi-stage sampling; single or multi-stage sampling using telephone directories; other) and survey mode [CAPI, computer-assisted telephone interview (CATI), paper and pencil interview (PAPI) and self-administered]. Each possible determinant was included in

Table 1 Survey characteristics of EU-SILC, SHARE and ESS in overlapping countries (Austria, Belgium, Germany, Denmark, Greece, Spain, France, Italy, Netherlands, Sweden)

	Sampling strategy	Type of survey sample design and frame	Response per country	Sample size per country	Survey mode
EU-SILC	Household probability sample, aged 16 + (except Germany, that used quota samples)	Simple random sampling (2); multi-stage sampling (6); quota sampling with random sampling (1); systematic sampling (1)	<50% (1); 50–59% (1); 60–69% (3); 70–79% (1); ≥80% (4)	5000–5999 (4); 6000–6999 (1); 9000–9999 (2); >9999 (3)	CAPI exclusively (2), or CAPI supplemented with CATI (2); CATI exclusively (1), or CATI supplemented with self-administered (1); PAPI exclusively (1), PAPI mixed with CATI (1), or a mixture of PAPI, CAPI and CATI (1); Self-administered exclusively (1)
SHARE	Household probability sample, aged ≥50 years	Simple random sampling from national population registers (2); multi-stage sampling using regional/local population registers (6); single or multi-stage sampling using telephone directories followed by screening in the field (2)	<50% (2); 50–59% (3); 60–69% (4); ≥80% (1)	1000–1999 (2); 2000–2999 (4); 3000–4000 (4)	CAPI (10)
ESS	Probability sample, aged ≥15 years	Simple random sampling from national population registers (2); multi-stage sampling using regional/local population registers (6); single or multi-stage sampling using telephone directories followed by screening in the field (1); info not available (1)	<50% (1); 50–59% (2); 60–69% (6); 70–79% (1)	1000–1999 (7); 2000–2999 (3)	CAPI (8) and PAPI (2)

Between brackets the number of countries with that characteristic.

separate models, plus in a fully adjusted model. With three surveys and 10 countries, the sample size included 30 observations (prevalence estimates) of very good or good self-perceived health. All analyses were performed using Stata/SE 13.1.

Results

Across all surveys, very good or good self-perceived health was least prevalent in Austria, Germany, Spain and Italy, with the lowest prevalence found among lower educated German men in EU-SILC (33.6%) and among lower educated Austrian men in SHARE (35.2%) (table 2). The highest prevalence of very good or good health was reported for higher educated Greek men in SHARE (93.2%) and for higher educated Austrian women in EU-SILC (86.9%). For some countries in ESS, there was a low availability of information within strata of gender and educational level. Higher educated respondents reported a higher prevalence of very good or good self-perceived health than the lower educated respondents across all surveys. There was, however, no consistent pattern in gender inequalities for self-perceived health across the surveys.

Figures 1a–1f present Bland and Altman plots showing the agreement in prevalence estimates of very good or good self-perceived health between EU-SILC, SHARE and ESS, separately for men and women. When comparing the surveys, SHARE's prevalence estimates of very good or good self-perceived health are systematically higher than the estimates obtained in EU-SILC or ESS. The mean differences were 9.8% in men and 6.7% in women between SHARE and EU-SILC (figures 1a and 1d) and were 6.3% in men and 10.0% in women between SHARE and ESS (figures 1b and 1e). Between EU-SILC and ESS, the agreement was better,

indicated by a lower prevalence estimate by EU-SILC in men (mean difference = −3.5%) and a higher estimate by EU-SILC in women (mean difference 3.2%) (figures 1c and 1f). When looking at the individual countries, it seems that the agreement between surveys is better for those countries with a relatively high prevalence of very good or good self-perceived health, such as Denmark, Sweden, the Netherlands, Belgium and Greece.

According to the ICC, the overall level of agreement between the surveys within countries was fairly high (ICC = 0.77) (table 3). When taking the survey characteristics into account the ICC increased to 0.84, meaning that 7% of the total variance was explained by the survey characteristics. According to the multilevel analysis, we found a statistically significantly higher prevalence of very good or good self-perceived health in SHARE as compared with EU-SILC (+8.22% points). There was no statistical significant difference between ESS and EU-SILC (+0.16% points). After adjustment for the survey characteristics, we still found a statistical significant difference between SHARE and EU-SILC (+6.96% points, 95% CI: 3.14 to 10.8) and an increased but still not significant difference between ESS and EU-SILC (−3.12% points, 95% CI, −7.11 to 0.86). In the fully adjusted model, the prevalence of self-perceived health was 0.13% points higher when the survey response was increased by 1% point. On the other hand, sample size was negatively associated with the prevalence of self-perceived health; an increase of 1000 respondents resulted in a 0.79% points decrease in the prevalence of very good or good health. When compared with the CAPI survey mode, CATI and PAPI resulted in higher prevalence estimates of self-perceived health (+6.71% points and +5.36% points, respectively). Sampling strategy did not contribute significantly to the self-perceived health prevalence. Results from the multilevel analysis by educational level

Table 2 Weighted prevalence (%) of very good or good self-perceived health by survey, country, gender and educational level in respondents aged 55–64 years

	EU-SILC (2005)			SHARE (2004–05)			ESS (2004–05)		
	Total	Men	Women	Total	Men	Women	Total	Men	Women
Austria	57.0			66.2			66.3		
Lower		35.4	45.3		35.2[b]	53.4		a	a
Medium		54.4	65.0		66.9	70.8		a	a
Higher		68.0	86.9		80.9	80.7		a	a
Belgium	65.4			73.2			63.5		
Lower		55.0	57.1		68.4	67.2		58.8	50.0
Medium		67.0	68.4		75.0	74.7		54.8[b]	69.2[b]
Higher		78.9	73.5		83.1	76.7		78.8[b]	81.5[b]
Germany	45.1			60.3			52.2		
Lower		33.6	39.3		43.1[b]	41.4		a	33.3[b]
Medium		41.1	47.7		54.6	63.3		49.0	52.1
Higher		53.5	55.1		68.3	75.4		58.7	71.8[b]
Denmark	70.0			72.0			68.6		
Lower		64.6	60.7		66.9[b]	48.7		50.0[b]	54.8[b]
Medium		68.4	75.3		74.6	68.1		62.2	68.3
Higher		81.4	80.5		79.1	84.9		90.2[b]	79.5[b]
Greece	67.9			72.0			65.0		
Lower		66.8	60.2		69.3	62.2		a	a
Medium		72.3	77.0		83.0	67.9		a	a
Higher		85.1	80.6		93.2	74.9[b]		a	a
Spain	51.2			62.0			47.9		
Lower		49.2	44.2		62.0	53.3		44.9	40.0
Medium		56.9	59.9		76.7[b]	80.3[b]		a	a
Higher		69.3	66.9		77.8[b]	79.2[b]		a	a
France	59.1			73.0			54.6		
Lower		53.6	51.4		60.1	63.5		a	a
Medium		64.0	61.6		75.3	74.7		a	a
Higher		72.3	71.3		90.1	91.5		a	a
Italy	42.0			58.6			46.7		
Lower		39.6	34.0		62.4	49.7		a	a
Medium		51.7	49.5		68.2	60.6		a	a
Higher		66.8	54.9		79.0[b]	62.0[b]		a	a
Netherlands	67.0			73.0			68.9		
Lower		52.3	64.1		65.3	70.2		67.9	64.9
Medium		67.5	67.4		77.8	70.0		73.0[b]	75.0[b]
Higher		81.2	72.3		87.5	79.0		77.8[b]	58.6[b]
Sweden	70.2			66.8			62.8		
Lower		64.7	65.9		60.4	53.6		a	a
Medium		71.2	66.2		77.9	62.5		a	a
Higher		79.2	76.0		82.8	77.7		a	a

a: Not available, or strata size is <20, or item non-response is >50%.
b: Strata size is between 20 and 49, or item non-response is between 20 and 49%.

and gender are presented in the supplementary files (Tables S6), yielding less precise results, but in line with the pooled analyses.

Discussion

This study showed that the prevalence of very good or good self-perceived health assessed by different surveys varies substantially across the surveys. When taking into account survey characteristics such as survey response, sample size, sampling strategy and survey mode, we found that prevalence estimates of very good or good self-perceived health ranged almost 10% points between the surveys on the country level. Response, sample size and survey mode contributed significantly to the disagreement between the surveys, explaining 7% of the total variance in the prevalence estimates of very good or good self-perceived health.

Survey response, sample size and survey mode were all associated with the prevalence of very good or good self-perceived health. A low response could indicate a selective population taking part in the survey, and if health plays an important role in this selection process, surveys or countries with lower response may have higher prevalence estimates of very good or good health. Furthermore, a lower response may enhance the probability that respondents differ from non-respondents in their characteristics, which may weaken the external validity of the survey results. This may suggest that in our study, where a higher response was associated with a higher prevalence of very good or good self-perceived health, a higher prevalence of very good or good self-perceived health was less likely to suffer from selection bias. Furthermore, it could be that the response bias in our estimates may be differential between surveys, countries or age groups, as factors behind non-response may differ. The response variable in our study was based on the response of the total study population, whereas the analyses only focussed on those aged 55–64 years. If the non-response was age specific, we might be able to explain more variation between the surveys. But, we question whether age-specific non-response would have different effects on the variation across the surveys. Thus, we have no strong arguments to believe that this would have substantially influenced our results. For sample size, we found opposite associations: a higher sample size was associated with lower prevalence estimates of very good or good self-perceived health. Although response and sample size helped explain the observed differences in the prevalence estimates of self-perceived health between the surveys, the impact of non-response bias on our results could be minimal as we based our analyses on the age stratum of 55–64 years, thereby minimizing the influence of age-distribution on the observed differences in prevalence estimates.
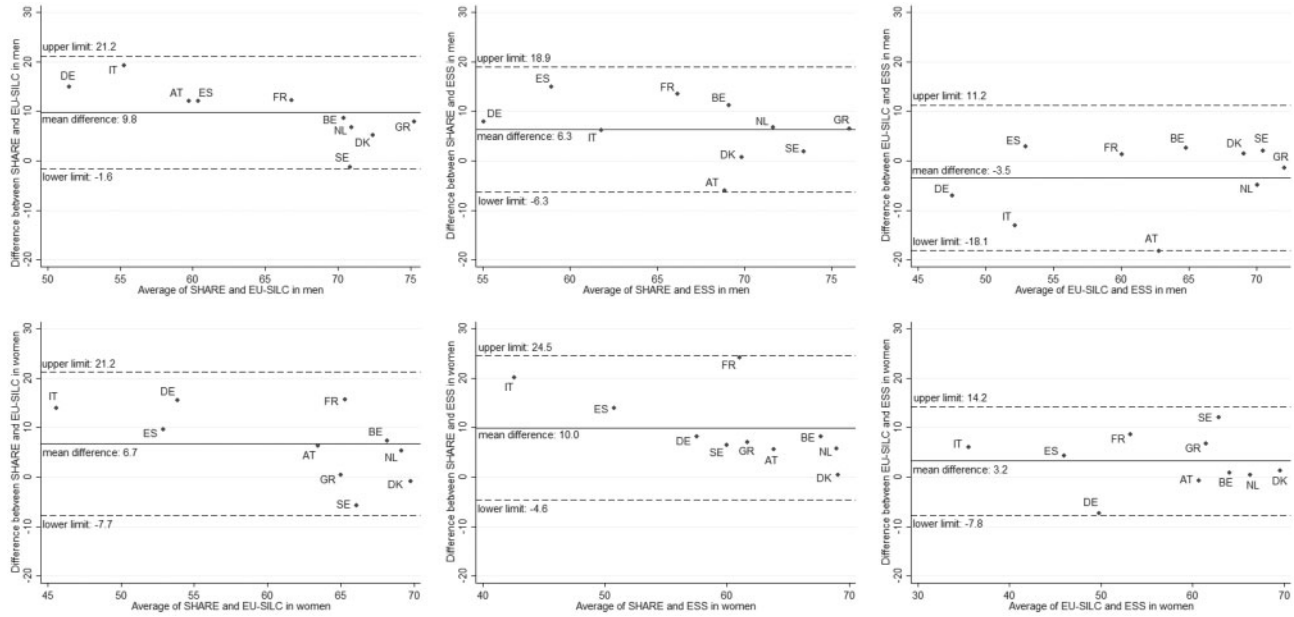
Figure 1 Bland and Altman plots presenting agreement in prevalence estimates of very good or good self-perceived health with mean differences and limits of agreement (1.96 SD) between EU-SILC, SHARE and ESS in men (first row) and women (second row)

Table 3 Multilevel analysis for very good or good self-perceived health (random intercept for country), N = 30

	Model 0		Model 1		Model 2		Model 3		Model 4		Model 5[a]	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Survey												
EU-SILC	1		1		1		1		1		1	
SHARE	8.22	4.83 to 11.61	9.10	5.50 to 12.69	3.80	−0.92 to 8.51	7.20	3.61 to 10.80	9.61	5.84 to 13.39	6.96	3.14 to 10.8
ESS	0.16	−3.23 to 3.55	0.84	−2.63 to 4.32	−4.77	−9.80 to 0.25	−0.13	−3.80 to 3.54	0.69	−2.77 to 4.14	−3.12	−7.11 to 0.86
Response			0.08	−0.05 to 0.20							0.13	0.02 to 0.24
Sample size (∗1000)					−0.65	−1.17 to −0.14					−0.79	−1.19 to −0.40
Sampling strategy												
Simple random, national records							1					
Multi-stage, regional/local records							1.44	−6.27 to 9.15				
Single/multi-stage, telephone records							3.00	−4.90 to 10.89				
Other							−4.19	−11.77 to 3.39				
Survey mode												
CAPI									1		1	
CATI									3.88	−2.56 to 10.33	6.71	1.03 to 12.39
PAPI									4.33	−0.10 to 8.76	5.36	1.63 to 9.10
Self-administered									−6.81	−15.32 to 1.70	−3.03	−6.63 to 3.58
Constant	59.49		54.20		65.70		59.05		58.10		55.40	
Variance (country)	49.07		50.21		36.97		49.86		42.54		31.86	
Variance (residual)	14.99		13.95		13.00		13.66		11.17		6.25	
ICC	0.77		0.78		0.74		0.78		0.79		0.84	

a: Sampling strategy did not statistically significantly contribute to this fully-adjusted model, and was therefore dropped.

Furthermore, subgroup analyses were performed within this age group by gender and educational level, for which similar results were obtained. For survey mode, our results showed that for surveys using CATI, more positive scores were obtained than when using CAPI or when questionnaires were self-administered. Our findings are in line with previous research on health-related quality of life, where telephone administration yielded more positive scores than the self-administered mode, but are contradicting with a study comparing four surveys, which found that telephone interviewing led to a higher prevalence of poor self-perceived health than face-to-face interviews.[27,38,39] As CATI was used only in a few countries in EU-SILC in our study, more

research is needed to replicate our findings and to give insight in the possible impact of CATI on the health estimates.

A major strength of this research is that the same question on health was used in comparable populations of the three surveys: non-institutionalized persons, aged 55–64 years. The comparisons between surveys that were made in this study were based on aggregate data in older adults aged 55–64 years. We deliberately chose to use this age group, as for this age group the numbers of respondents were highest and probably gave most reliable estimates across the surveys. Furthermore, all three surveys used weights to take into account the original national population. Also, all surveys had an overlapping data collection period, 2005, in which the

assessments of self-perceived health took place. However, as the prevalence of ill-health is increasing with age, we should be careful to extrapolate our findings to the general older population and younger age groups.

A different question order may have hampered the comparability between the surveys.[29,30] Especially for a subjective indicator such as self-perceived health, respondents' answers may depend on the questions preceding the self-perceived health question. Respondents may consider a broader view and reflect more thoroughly on the concept of health when more nearby questions are posed.[40] Based on the survey questionnaires, we found that the order was quite consistent across EU-SILC and ESS; the first question of the health-related module. However, the place of the health-related module within the total questionnaire differed between the surveys, as they have different scopes. Only in SHARE, the order within the health-related module was randomly assigned to the respondents, as an experiment. Based on earlier research on this topic, which found that health was negatively influenced by the battery of health questions preceding the assessment,[30] we hypothesized that SHARE would underestimate the level of very good or good self-perceived health when compared with the other surveys, but the opposite is true. Therefore, we do not feel that the order determines the differences found in our study.

This article provides valuable insights in the comparability and the causes of incomparability of health information across European surveys. The goal of this study was not so much to provide absolute estimates of self-perceived health for policy purposes, but to examine the influence of several survey characteristics on the observed differences in estimates of health. As differences in health estimates between three major European population-based surveys seem substantial, they could thus lead to different interpretations of the population health status. When using SHARE information, population health might be estimated more favourably than, when using EU-SILC or ESS. However, we do not know whether SHARE overestimates, or ESS underestimates self-perceived health when compared with the other surveys. Further steps need to be taken to improve understanding of observed disagreement between European surveys and to evaluate whether disagreement in occurrence of other health indicators could be explained by survey characteristics. If the disagreement across health indicators is differential, this might influence the identification of research and policy priorities. To conclude, our findings offer an important reminder that disagreement in health prevalence estimates between surveys may limit the usefulness for direct comparisons across studies in health policies for Europe.

Supplementary data

Supplementary data are available at *EURPUB* online.

Acknowledgements

Funding

Conflicts of interest: None declared.

Key points

- Cross-national comparisons of health in European countries are crucial, but results might differ across surveys.
- This study showed that the prevalence of very good or good self-perceived health assessed by three major European surveys differs substantially between the surveys.
- Several survey characteristics such as the response, sample size and mode contributed to the differences found in very good or good self-perceived health between the surveys.
- Taking survey characteristics into account, the prevalence estimates of very good or good self-perceived health ranged almost 10% points between the surveys on the country level.
- Disagreement in health prevalence estimates between surveys may limit the usefulness for direct comparisons across studies in health policies for Europe.

References

1 OECD. *Health at a Glance: Europe 2012*. OECD Publishing, Paris, France; 2012. doi: 10.1787/9789264183896-en.

2 Tolonen H, Koponen P, Mindell J, et al. European Health Examination Survey-towards a sustainable monitoring system. *Eur J Public Health* 2014;24:338–44.

3 Maki N, Martikainen P, Eikemo T, et al. Educational differences in disability-free life expectancy: a comparative study of long-standing activity limitation in eight European countries. *Soc Sci Med* 2013;94:1–8.

4 Gallo V, Mackenbach JP, Ezzati M, et al. Social inequalities and mortality in Europe–results from a large multi-national cohort. *PLoS One* 2012;7:e39013

5 Eikemo TA, Bambra C, Joyce K, Dahl E. Welfare state regimes and income-related health inequalities: a comparison of 23 European countries. *Eur J Public Health* 2008;18:593–9.

6 Oyebode O, Mindell JS. A review of the use of health examination data from the Health Survey for England in government policy development and implementation. *Arch Public Health* 2014;72:24

7 Bonneux L, Van der Gaag N, Bijwaard G, et al. Demographic epidemiologic projections of long-term care needs in selected European countries: Germany, Spain, the Netherlands and Poland. *Economic Policy* 2012;1–12. ENEPRI Policy Briefs

8 Borsch-Supan A, Brandt M, Hunkler C, et al. Data Resource Profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). *Int J Epidemiol* 2013;42:992–1001.

9 ESS Round 1: European Social Survey. ESS-1 2002 Documentation Report. Edition 6.3. Bergen, Norway; 2012.

10 Santourian A, Ntakou E. Working paper with the description of 'Income and Living conditions dataset'. EU-SILC (European Union Statistics on Income and Living Conditions): Methodological studies and publications. 2014.

11 van der Wel KA, Dahl E, Thielen K. Social inequalities in 'sickness': European welfare states and non-employment among the chronically ill. *Soc Sci Med* 2011;73:1608–17.

12 Dahl E, van der Wel KA. Educational inequalities in health in European welfare states: a social expenditure approach. *Soc Sci Med* 2013;81(0):60–9.

13 Palencia L, Malmusi D, De Moortel D, et al. The influence of gender equality policies on gender inequalities in health in Europe. *Soc Sci Med* 2014;117:25–33.

14 Swift HJ, Vauclair CM, Abrams D, et al. Revisiting the paradox of well-being: the importance of national context. *J Gerontol B Psychol Sci Soc Sci* 2014;69:920–9.

15 Bambra C, Eikemo TA. Welfare state regimes, unemployment and health: a comparative study of the relationship between unemployment and self-reported health in 23 European countries. *J Epidemiol Community Health* 2009;63:92–8.

16 Avendano M, Glymour MM, Banks J, Mackenbach JP. Health disadvantage in US adults aged 50 to 74 years: a comparison of the health of rich and poor Americans with that of Europeans. *Am J Public Health* 2009;99:540–8.

17 Castro-Costa E, Dewey M, Stewart R, et al. Prevalence of depressive symptoms and syndromes in later life in ten European countries: the SHARE study. *Br J Psychiatry* 2007;191:393–401.

18 Jagger C, Weston C, Cambois E, et al. Inequalities in health expectancies at older ages in the European Union: findings from the Survey of Health and Retirement in Europe (SHARE). *J Epidemiol Community Health* 2011;65:1030–5.

19 Hank K. How ''successful'' do older Europeans age? Findings from SHARE. *J Gerontol B Psychol Sci Soc Sci* 2011;66B:230–6.

20 Aromaa A, Koponen P, Tafforeau J, Vermeire C, The HIS/HES Core Group. Evaluation of health interview surveys and health examination surveys in the European Union. *Eur J Public Health* 2003;13(suppl 1):67–72.

21 de Bruin A, Picavet H, Nossikov A. Health interview surveys. Towards international harmonization of methods and instruments. *WHO Reg Publ Eur Ser* 1996;58:i–xiii, 1–161.

22 The EHEMU/EHLEIS team. The Minimum European Health Module background documents. In: EHEMU Technical report 2010_4.6. 2010. http://maryland.mri.cnrs. fr/ehemu/pdf/healthmodule/TR_2010_4.6_Health_module.pdf. (24 February 2016, date last accessed).

23 Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav* 1997;38:21–37.

24 Stiefel M, Perla R, Zell B. A healthy bottom line: healthy life expectancy as an outcome measure for health improvement efforts. *Milbank Quarterly* 2010;88:30–53.

25 Börsch-Supan A, Mariuzzo F. Our sample 50+ in Europe. In: Börsch-Supan A, Brugiavini A, Jürges H, Mackenbach JP, Siegrist J, Weber G, editors. *Health, Ageing and Retirement in Europe - First results from the Survey of Health, Ageing and Retirement in Europe*. Mannheim: Mannheim Research Institute for the Economics of Aging, 2005: 30–4.

26 Tipping S, Hope S, Pickering K, et al. The effect of mode and context on survey results: Analysis of data from the Health Survey for England 2006 and the Boost Survey for London. *BMC Med Res Methodol* 2010;10:84

27 Salomon JA, Nordhagen S, Oza S, Murray CJ. Are Americans feeling less healthy? The puzzle of trends in self-rated health. *Am J Epidemiol* 2009;170:343–51.

28 Jurges H, Avendano M, Mackenbach JP. Are different measures of self-rated health comparable? An assessment in five European countries. *Eur J Epidemiol* 2008;23:773–81.

29 Bowling A, Windsor J. The effects of question order and response-choice on self-rated health status in the English Longitudinal Study of Ageing (ELSA). *J Epidemiol Community Health* 2008;62:81–5.

30 Lumsdaine RL, Exterkate A. How survey design affects self-assessed health responses in the Survey of Health, Ageing, and Retirement in Europe (SHARE). *European Economic Review* 2013;63:299–307.

31 Eurostat. Comparative EU statistics on income and living conditions: Issues and challenges. Proceedings of the EU-SILC conference (Helsinki, 6-8 November 2006). 2007; Methodologies and working papers.

32 Eurostat. Eurostat Database. Eurostat, 2014. http://epp.eurostat.ec.europa.eu/portal/ page/portal/statistics/search_database. Accessed December 2014.

33 ESS Round 2: European Social Survey. ESS-2 2004 Documentation Report. Edition 3.4. Bergen, Norway, 2012.

34 ESS Round 2: European Social Survey Round 2 Data. Data file edition 3.3. Bergen, Norway, 2004.

35 SHARE. Release guide 2.5.0. Waves 1 & 2, 2011. http://www.share_project.org (24 February 2016, date last accessed).

36 United Nations Educational, Scientific, and Cultural Organization. *ISCED 1997: International Standard Classification of Education*. Montreal, Quebec, Canada: UNESCO Institute for Statistics; 2006.

37 Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.

38 Hanmer J, Hays R, Fryback D. Mode of administration is important in US national estimates of health-related quality of life. *Med Care* 2007;45:1171–9.

39 Hays RD, Kim S, Spritzer KL, et al. Effects of mode and order of administration on generic health-related quality of life scores. *Value Health* 2009;12:1035–9.

40 Wikman A. Context effects as an illustration of response uncertainty - a cautionary tale. *Soc Indic Res* 2007;84:27–38.