International Journal of Epidemiology, 2019, 14–14j
doi: 10.1093/ije/dyy227
Advance Access Publication Date: 1 December 2018
Data Resource Profile

Data Resource Profile

Data Resource Profile: Cross-national and cross-study sociodemographic and health-related harmonized domains from SAGE plus CHARLS, ELSA, HRS, LASI and SHARE (SAGE+ Wave 2)

Nadia Minicuci,[1] Nirmala Naidoo,[2,3] Barbara Corso,[1] Ilaria Rocco,[1] Somnath Chatterji[2,3] and Paul Kowal[2,4,5]*

[1]National Research Council, Neuroscience Institute, Padova, Italy, [2]World Health Organization Study on global AGEing and adult health (SAGE), Geneva, Switzerland, [3]Surveys, Measurement and Analysis Unit, Department of Information, Evidence and Research, World Health Organization, Geneva, Switzerland, [4]Chiang Mai University Research Institute for Health Sciences (CMU-RIHES), Chiang Mai, Thailand and [5]Research Centre for Generational Health and Ageing, University of Newcastle, Callaghan, NSW, Australia

*Corresponding author. Research Institute for Health Sciences, Chiang Mai University, 110 Intavaroros Road, Sriphum, Muang, Chiang Mai 50200, Thailand. E-mail: paul.r.kowal@rihes.org

Editorial decision 13 September 2018; Accepted 21 November 2018

Data resource basics

The effort to create SAGE+ Wave 2 (2010) follows from the harmonized dataset generated for SAGE+ Wave 1 (2004) whose details can be found in Minicuci *et al.*[1] The SAGE+ Wave 2 harmonization included the previous four studies, the Study on global AGEing and adult health (SAGE), English Longitudinal Study on Ageing (ELSA), US Health and Retirement Study (HRS) and Survey of Health, Ageing and Retirement in Europe (SHARE), and introduced two additional studies: Wave 1 of the China Health and Retirement Longitudinal Study (CHARLS) and the pilot of the Longitudinal Aging Study in India (LASI). The outcome of the harmonization work leading to SAGE+ Wave 1 and SAGE+ Wave 2 is two common datasets that will allow a wide range of cross-country comparisons and evaluation of the predictors of transitions in various domains, from subjective health status and risk factors shifts, to evaluation in relationships between wealth and health in countries at different levels of development. Details about the decisions and methods used for the harmonization process for SAGE+ Wave 1 are available in Minicuci *et al.* 2016.[1] The aim of this article is to present the analogous results for SAGE+ Wave 2 which extends and adds unique data to SAGE+ Wave 1.

The harmonization process for SAGE+ Wave 2 builds on SAGE+ Wave 1 processes and follows the same methodology. SAGE+ Wave 2 included the two additional studies, CHARLS and LASI, whose variables were subjected to the same harmonization process as used in SAGE+ Wave 1 without encountering major deviations. The inclusion of these studies allowed the creation of some extra harmonized variables, such as verbal fluency, depression, emotional/psychiatric problems, cataracts and hip fracture, for some studies.

Data resource area and population coverage

Detailed descriptions of SAGE, ELSA, HRS and SHARE were presented in our previous paper.[1] A brief recapitulation of these four studies is included below, along with more detailed descriptions of the two additional studies that were included in the SAGE+ Wave 2 dataset.

SAGE consists of standardized longitudinal panel studies in China, Ghana, India, Mexico, the Russian Federation and South Africa [www.who.int/healthinfo/sage]. In-person interviewers were used to collect the data. SAGE Wave 1 (2008/10) was considered for SAGE+ Wave 2 harmonization and had an individual response rate ranging from 53.0% in Mexico to 93.0% in China.[2]

ELSA is a representative sample of the older English population, which began in 2002 with seven waves of data collection completed through 2017 [www.elsa-project.ac.uk]. Face-to-face interviews followed by a self-completion questionnaire every 2 years and a nurse assessment every 4 years are used to collect data. The dataset from Wave 5 (2010) was considered for harmonization, with an individual response rate of 78.0%.[3]

The US HRS is a biennial survey which started in 1992 and is representative of the US population aged 50 years and older [hrsonline.isr.umich.edu]. Face-to-face interviews and phone interviews were used to capture data. Data from Wave 10 (2010) were included in this analysis, with an individual response rate of 88.6%.[4]

SHARE consists of cross-national panel surveys using in-person interviews to collect data on individuals aged 50+ [www.share-project.org] in European countries. SHARE Wave 4 (2010), here considered for the harmonization, covered six more countries than SAGE+ Wave 1 (Czech Republic, Estonia, Hungary, Poland, Portugal and Slovenia) but did not include Greece in this wave. Wave 4 individual response rates ranged from 74.0% in Spain to 93.0% in France.[5]

The two new studies included in SAGE+ Wave 2, were CHARLS and LASI.

CHARLS is a nationally representative longitudinal survey of persons in China aged 45 years or older and their spouses [http://charls.pku.edu.cn/en]. The national baseline survey (Wave 1) for the study was conducted between June 2011 and March 2012. Face-to-face interviews were used for data collection. The national baseline survey individual response rate was 80.5%.[6]

LASI completed a pilot study and have recently begun a full national-level study [http://www.hsph.harvard.edu/pgda/lasi.html]. LASI (Pilot, 2008) was designed to collect information conceptually comparable to HRS. The individual response rate of this pilot was of 90.9%.[7]

Data collected and harmonization process

Details of the harmonization process were described in Minicuci *et al.*[1] In summary, steps included: documenting the study design and variables; determining the harmonizable domains and the group of core variables targeted for available domains; and, data processing and data quality assessments using statistical indicators. Details of the harmonization process are documented in STATA codes. This documentation meets the Data Documentation Initiative (DDI) standards and the international standards for data and metadata exchange.[8]

Based on SAGE+ Wave 1, the new wave of harmonized variables was categorized into nine domains as follows:

i. sociodemographic and economic: sex, age, marital status, living arrangements, educational level, occupation, age at retirement;
ii. health states describing physical functioning: Activities of Daily Living (ADL), Instrumental Activities of Daily Living (IADL), mobility, near and distance vision, hearing, pain;
iii. overall self-report of health and mental state: self-reported health, cognition/memory, depression, sleep, loneliness;
iv. health examinations: measured blood pressure;
v. physical and mental performance tests: normal-pace walking test, grip strength, cognition/memory tests;
vi. risk factors: body mass index (BMI), tobacco consumption, alcohol consumption, physical activity;
vii. chronic conditions: self-reported chronic conditions, disease treatment;
viii. social network: social network index; and,
ix. subjective well-being: quality of life, life satisfaction, well-being.

This paper presents the characteristics of the two newly introduced surveys (CHARLS and LASI) in order to show their compatibility with the harmonization pursued in SAGE+ Wave 1. For the other studies, questions that have been posed differently from the previous wave or new ones have been introduced are noted, and any deviations from the questions used for the SAGE+ Wave 1 harmonization are described below.

Sociodemographic and economic domains

In the harmonization of the sociodemographic and economic variables, CHARLS and LASI consistently met the SAGE+ Wave 2 harmonization criteria. A few notes on specific studies: for LASI, age at retirement was not available; and for SAGE Wave 1, the current marital status variable made no distinction between 'separated' and 'divorced', so it was decided to set all these respondents in the 'divorced' category.

Physical functioning domain

Supplementary Table A (available as Supplementary data at *IJE* online) consists of a checklist of the physical functioning items from each study. Regarding ADLs and IADLs, CHARLS Wave 1 had only five ADLs (missing:

walking across a room) and five IADL (missing: using a map to figure out how to get around in a strange place; and making telephone calls) variables; LASI requested only information on ADL. Regarding the mobility section, CHARLS had seven out of the 10 mobility variables available for harmonization, whereas LASI did not include any such variables in its questionnaire. SAGE Wave 1 introduced the questions on all ADLs and eight questions regarding mobility activities; only one question could be categorized as an IADL.

Self-reported health, and physical and mental health domains

Supplementary Table B (available as Supplementary data at *IJE* online) shows the variables available for these domains, by study. For overall general self-rated health, the single question in CHARLS was posed to a subset of respondents at the beginning of the questionnaire and to a different subset of respondents at the end of the questionnaire. Both CHARLS and LASI used the same five response categories, ranging from excellent to poor.

For cognition, CHARLS asked respondents to rate their memory from excellent to poor, in concordance with SAGE+ Wave 1, but in LASI this question was not available.

Self-reported depression and sadness were based on a single question about feeling depressed or sad both in CHARLS and LASI, rated from 'rarely or none of the time (<1 day)' to 'most or all of the time (5–7 days)'. The harmonized variable was dichotomized into no [rarely or none of the time (<1 day)] and yes (the remaining categories). CHARLS and LASI asked another question on depression: 'Overall in the last month, how much of a problem did you have with feeling sad, low, or depressed?' This was rated from none to extreme, but it was not considered because of the high number of missing values (79.0% and 67.0%, respectively).

In CHARLS and LASI, self-reported trouble with sleeping and feelings of loneliness used the same response categories as self-reported depression; so again, a dichotomous variable (yes/no) was generated for harmonization.

Compared with SAGE Wave 0, SAGE Wave 1 asked respondents to rate their memory from very good to very bad and introduced questions about loneliness. ELSA did not ask this question in their Wave 5, and SHARE administered the question in their Wave 4.

Health examination

Both CHARLS and LASI included three blood pressure measurement variables for respondents in their datasets. For ELSA, blood pressure measurements were available only in Wave 4 (2008); therefore its values refer to this wave. In SAGE Wave 1, blood pressure measurements were conducted, whereas this was not done in Wave 0. Three measurements were taken for each respondent, with the variables available in the datasets. In SHARE, the three measurements were performed only for the German sample, so the variable was not harmonized.

Physical and cognitive performance tests domain

Gait speed eligibility criteria differed across surveys; SAGE administered the test to all subjects; CHARLS, ELSA and LASI to individuals aged 60+ years; HRS to those aged 65+ years; and SHARE to those aged 75+ years. Moreover, two different distances were considered: 2.5 m by CHARLS, ELSA and HRS and 4 m by LASI and SAGE. The test had missing, refusal or not-applicable values: CHARLS (valid $n = 5695$ from $n = 7689$); ELSA (valid $n = 6413$ from $n = 7535$); LASI (valid $n = 348$ from $n = 614$); HRS (valid $n = 3841$ from $n = 10\ 937$) and SAGE (valid $n = 30\ 179$ from $n = 35\ 146$). SHARE did not administered the walking test in its 2010 survey round.

SAGE Wave 1 introduced the measurement of grip strength, and for ELSA it was available for only in Wave 4 (2008), hence the harmonized variable refers to this wave. The grip strength variable in the harmonized data set reflects the mean of two measurements in the respondent's dominant hand.

Supplementary Table C (available as Supplementary data at *IJE* online) includes descriptions of the cognition tests used in each study. A verbal fluency test variable was added to SAGE+ Wave 2, which was not present in SAGE+ Wave 1. This test was administered in all surveys (HRS added this variable in 2010) except for CHARLS. HRS administered the time orientation test only to participants aged 65+, whereas all the other studies did not apply this age criterion. SAGE Wave 1 introduced immediate and delayed recall, verbal fluency and digit span tests.

Risk factors domain

In the harmonization process for alcohol consumption, tobacco and physical activity variables, CHARLS and LASI fit the SAGE+ Wave 2 harmonization criteria quite well. For alcohol consumption, the question concerning the number of drinks refers to different time frames for each survey (CHARLS: '. . in the last year'; LASI: '. . in the last 30 days'; HRS and SHARE: '. . . in the past 3 months'; ELSA and SAGE: '. . . in the past 7 days'). SAGE Wave 1 added information on years since quitting smoking for ex-smokers, and the information collected on physical activity

levels was divided into two separate sets of questions: one regarding effort at work and one regarding effort in fitness/recreational/leisure time activities. The height and weight variables in ELSA were measured in Wave 4 (2008), and hence the harmonized variable in the dataset refers to this wave. SHARE Wave 4 did not have information about the number of cigarettes, but included the number of drinks per day, so the computation of frequent/infrequent heavy drinkers was now possible in SAGE+ Wave 2.

Chronic conditions domain

Supplementary Table D (available as Supplementary data at *IJE* online) lists the set of self-reported chronic health conditions recorded in each study through the question, 'Have you ever been diagnosed with/told by a doctor that you have…?' No substantial deviations were present for CHARLS and LASI. Neither study recorded information about angina, and LASI did not ask about asthma. SAGE Wave 1 added questions on chronic lung diseases, hypertension and stroke. SHARE Wave 4 dropped the question about asthma (although the question on current asthma treatment was posed to all subjects) for all countries, except for Austria, Belgium and The Netherlands, so asthma was not harmonizable for SHARE.

Multimorbidity

Multimorbidity was computed based on six common conditions across all surveys (arthritis, cancer, chronic lung disease, diabetes, hypertension and stroke), although SAGE did not include questions about cancer. Hence, the multimorbidity harmonized variable for SAGE Wave 1 in SAGE+ Wave 2 was based on just five conditions.

Social network index

The Social Network Index (SNI) was computed according to SAGE+ Wave 1 for LASI and it ranges from 0 (no ties) to 4 (high level of ties). CHARLS did not have information on religious attendance so SNI construction for this study was not possible.

Subjective well-being: quality of life, life satisfaction and well-being domain

A life satisfaction question was harmonized for CHARLS and LASI as well, but the question on self-perceived social well-being using a response 'ladder' was available only for LASI. SAGE Wave 1 added a question on life satisfaction, as well as the set of questions constituting the brief WHO Quality of Life instrument.

Data resource use

Weights, clustering and stratum variables

Each study provided its own weighting variables, along with the stratum and cluster variables, as part of the complex survey design. This information was used to create a weighted dataset for cross-sectional analysis for SAGE+ Wave 2. Cluster and strata variables were not available for CHARLS. SHARE release 5.0.0 provided these variables; however, since missing values ranged between 20% and 35%, they have not been used in the present harmonization study.

Statistical analyses

Descriptive statistics were expressed as mean and standard error (SE) for quantitative variables and as prevalence for qualitative variables. Weights, strata and cluster information from the SAGE+ Wave 2 harmonized weighted dataset for cross-sectional analysis was used in the *proc surveyreg* or *surveymeans* procedure (SAS v9.4) which produces estimates from complex sample survey data. All means and prevalence rates were age-standardized based on 50–64, 65–74 and 75+ age groups using the WHO standard population,[9] except for the variable on walking test since each study administered the test to individuals with different age ranges.

Differences in the sociodemographic characteristics

Large samples of adults aged 50+ years were available: SAGE, $n = 35\,125$; CHARLS, $n = 13\,817$; ELSA, $n = 8985$; HRS, $n = 20\,337$; LASI, $n = 1109$; and SHARE, $n = 56\,755$. Table 1 shows national differences in sociodemographic characteristics, with more women in all surveys and slightly higher prevalence in HRS (53.5%) and SHARE (53.8%) compared with the other studies. The mean age was comparable across all surveys, varying from 62 years in LASI to 64.4 years in ELSA. The majority of respondents were married/cohabiting, with prevalence from 62.9% in HRS to 81.6% in CHARLS. Education level varied considerably across studies, with ELSA, HRS and SHARE reporting on average more than 10 years of education. These patterns are consistent with results obtained in SAGE+ Wave 1.

The mean household size was around two people for ELSA, HRS and SHARE, whereas in SAGE and CHARLS it was about 3.5 members. LASI had the greatest mean number of household members (5.2). In SAGE, the mean household size varied from 1.2 (Mexico) to 5.5 (India). Similar to what was found in SAGE+ Wave 1, Russia was

Table 1. Sociodemographic and economic characteristics: age-standardized prevalence ratios or mean value for SAGE+ Wave 2 (year 2010), by survey

	SAGE (W1) ($n = 35\,125$)	CHARLS (W1) ($n = 13\,817$)	ELSA (W5) ($n = 8985$)	HRS (W10) ($n = 20\,337$)	LASI (Pilot) ($n = 1109$)	SHARE (W4) ($n = 56\,755$)
Sex						
Male	47.5	49.0	47.7	46.5	49.9	46.2
Female	52.5	51.0	52.3	53.5	50.1	53.8
Mean age, years (SE)	62.6 (0.05)	63.2 (0.04)	64.4 (0.04)	63.3 (0.06)	62.0 (0.14)	63.4 (0.04)
Age groups (years, weighted)						
50–60	44.8	43.7	26.8	39.7	45.4	37.2
60–70	30.5	32.6	35.9	31.1	31.5	29.2
70–80	18.6	17.0	23.5	17.8	16.6	21.6
80+	6.1	6.7	13.8	11.4	6.5	12.0
Marital status						
Never married	2.2	1.2	6.6	7.5	1.5	6.9
Currently married/cohabiting	74.7	81.6	68.8	62.9	74.3	68.9
Separated/divorced/widowed	23.1	17.2	24.6	29.6	24.2	24.2
Years of education						
0	28.9	31.3	0.3	0.5	49.4	1.4
1–5	20.5	19.3	0.0	1.8	16.3	11.4
6–8	17.1	21.5	0.0	4.2	11.5	19.7
9–13	24.9	24.6	68.4	47.6	18.5	44.1
14+	8.6	3.3	31.3	45.9	4.3	23.4
Mean number of years of education (SE)	5.8 (0.19)	4.9 (0.05)	12.3 (0.03)	13.0 (0.08)	4.0 (0.22)	10.5 (0.04)
Living arrangements						
Household with one person	28.2	8.7	30.5	32.6	3.9	29.3
Household with two people	27.7	29.9	49.1	39.8	14.3	43.1
Household with three people	26.9	44.9	19.8	25.0	43.8	25.6
Household with four people	11.3	14.8	0.6	2.4	27.1	1.9
Household with five people or more	5.9	1.7	0.0	0.2	10.9	0.1
Mean number of people residing in the household, respondent included (SE)	3.5 (0.08)	3.6 (0.03)	2.0 (0.01)	2.2 (0.02)	5.2 (0.17)	2.2 (0.02)
Occupation						
Employed	56.2	48.1	39.7	44.8	11.7	31.4
Unemployed	3.9	35.6	1.5	5.5	19.3	4.3
Retired	27.7	14.2	45.7	32.4	3.6	46.8
Homemaker	2.0	1.6	6.1	6.1	39.9	11.0
Ill health/disabled	7.7	0.2	6.2	10.5	3.9	4.8
Other	2.5	0.3	0.8	0.7	21.6	1.7
Mean age at retirement, years (SE)	54.0 (0.20)	52.9 (0.19)	58.2 (0.17)	58.0 (0.17)	NA	54.6 (0.16)

W, Wave; NA, not available.

aligned with ELSA, HRS and SHARE in SAGE+ Wave 2, with a mean value of 2.2 persons per household.

Considerable variability was found in the percentage employed. CHARLS reported an unemployment rate as high as 35.6% and LASI reached almost 20%, whereas the other surveys were below 6%. ELSA and SHARE had similar 'retired' employment status (around 46%), followed by HRS (32.4%) and SAGE (27.7%). Mean age at retirement was similar for ELSA and HRS, as well for SAGE and SHARE; CHARLS had the lowest value (52.9). Compared with SAGE+ Wave 1, SAGE showed a higher percentage of employed (56.2% vs 40.3%) and unemployed (3.9% vs 1.0%), balanced by the lower percentage of homemakers (2.0% vs 18.8%) and retired (27.7% vs 34.0%) in this more current wave of the study. ELSA, HRS and SHARE showed similar percentages between the two waves.

Differences in health status and health habits characteristics

Table 2 reports national differences in health status and habits. Poor/bad/very bad health was reported by 27.4%

Table 2. Health status and health habit characteristics: age-standardized prevalence ratios or mean value for SAGE+ Wave 2 (year 2010), by survey

	SAGE (W1) ($n = 35\,125$)	CHARLS (W1) ($n = 13\,817$)	ELSA (W5) ($n = 8985$)	HRS (W10) ($n = 20\,337$)	LASI (Pilot) ($n = 1109$)	SHARE (W4) ($n = 56\,755$)
Self-reported health						
Excellent/very good/good	29.7	27.2	74.4	75.3	82.9	60.4
Fair/moderate	49.1	45.4	17.5	17.7	13.3	27.9
Poor/bad/very bad	21.2	27.4	8.1	7.0	3.8	11.7
Health conditions (ever been diagnosed/told by a doctor)						
Angina	10.9	NA	7.6	7.9	NA	NA
Arthritis	21.5	34.3	34.7	51.8	9.3	23.9
Asthma	4.1	4.2	11.2	NA	NA	NA
Cancer	NA	0.9	5.5	12.6	0.3	4.8
Diabetes	7.7	6.5	10.6	20.2	10.4	12.5
Hypertension	28.7	27.6	38.4	54.0	20.0	36.7
Lung diseases	7.5	11.3	4.9	9.3	6.0	6.6
Stroke	3.1	3.0	3.7	5.5	0.9	3.3
Measured hypertension	36.1	59.2	39.9[a]	57.8	29.2	NA
Among hypertensives:						
Hypertensive not on treatment	20.4	19.2	2.0	7.1	12.2	
Normotensive on treatment	21.1	4.8	54.3	59.5	43.2	
Hypertensive on treatment	58.5	76.0	43.7	33.4	44.6	
Number of chronic conditions (multimorbidity)[b]						
None	53.1	42.4	37.4	21.4	67.8	42.4
1	30.5	37.0	36.2	30.8	21.1	34.4
2+	16.4	20.6	26.4	47.8	11.1	23.2
Able to remember			NA		NA	
Yes, with no difficulties	2.4	5.6		35.1		24.5
With some difficulties	31.4	13.7		41.3		47.0
With difficulty	48.3	43.7		19.6		22.7
With a lots of difficulties	17.9	37.0		4.0		5.8
Time orientation (number of correct items)[c]	NA					
4		47.5	82.2	80.7	47.6	86.7
2–3		38.8	16.4	18.3	21.2	11.8
0–1		13.7	1.4	1.0	31.2	3.5
Immediate recall (number of words correctly recalled)						
10	0.6	0.1	0.9	0.7	1.2	0.5
7–9	12.7	4.8	38.9	28.7	15.5	22.9
4–6	61.7	42.1	52.7	61.2	58.6	61.1
1–3	23.8	31.5	6.2	8.6	22.1	13.7
None	1.2	21.5	1.3	0.8	2.6	1.8
Delayed recall (number of words recalled)						NA
10	1.3	0.1	0.7	0.5	0.2	0.4
7–9	20.8	2.7	18.0	14.4	8.2	9.7
4–6	52.3	28.3	57.0	59.4	38.7	47.7
1–3	20.5	46.7	18.3	20.9	44.0	33.0
None	5.1	22.2	6.0	4.8	8.9	9.2
Feeling depressed/sad						
Yes	38.4	54.0	25.7	20.4	35.6	43.1
Trouble sleeping						
Yes	51.6	52.3	42.1	46.6	48.8	36.2
Felt lonely						NA
Yes	11.7	31.2	12.9	15.3	32.3	
Able to read						
Yes, with no difficulties	43.0	31.0	56.6	41.8	53.9	36.1

(Continued)

Table 2. Continued

	SAGE (W1) ($n = 35\,125$)	CHARLS (W1) ($n = 13\,817$)	ELSA (W5) ($n = 8985$)	HRS (W10) ($n = 20\,337$)	LASI (Pilot) ($n = 1109$)	SHARE (W4) ($n = 56\,755$)
With some difficulties	31.2	42.8	34.5	41.1	34.9	37.1
With difficulty	18.3	25.2	7.0	11.3	9.0	16.6
With a lots of difficulties	7.5	1.0	1.9	5.8	2.2	10.2
Able to seeing things at a distance						
Yes, with no difficulties	51.1	35.8	61.7	50.5	54.5	44.2
With some difficulties	25.8	38.3	31.6	38.9	33.0	39.1
With difficulty	14.8	24.9	5.2	7.9	10.1	11.4
With a lots of difficulties	8.4	1.0	1.5	2.7	2.4	5.3
Hearing	NA				NA	
Excellent		1.1	18.7	20.2		15.6
Very good		12.3	29.9	28.9		23.2
Good		29.6	31.9	32.9		40.5
Fair		40.2	15.2	14.0		16.7
Poor		16.8	4.3	4.0		4.0
Often troubled with pain					NA	
Yes	61.0	33.7	39.7	35.4		57.6
BMI						
Underweight	15.4	8.4	0.9[a]	1.1	26.0	1.2
Normal	46.3	63.3	25.7	27.3	53.0	36.8
Overweight	25.5	24.0	41.3	37.6	15.7	41.9
Obese	12.8	4.3	32.1	34.0	5.3	20.1
BMI						
Mean value (SE)	24.2 (0.16)	23.2 (0.05)	28.5 (0.08)[a]	28.6 (0.07)	21.8 (0.23)	26.7 (0.05)
Tobacco consumption						
Not current smoker	66.2	71.1	84.5	83.8	92.5	80.3
Current smoker	33.8	28.9	15.5	16.2	7.5	19.7
Mean number of cigarettes (SE)	6.5 (0.49)	17.4 (0.24)	14.1 (0.31)	13.9 (0.29)	16.1 (3.52)	NA
Number of years smoking			NA			
Current smoker	31.5 (0.43)	39.9 (0.20)		44.1 (0.18)	34.8 (1.60)	37.2 (0.32)
Ex-smoker	NA	30.7 (0.45)		23.1 (0.23)	21.7 (2.90)	20.4 (0.27)
Number of years quit smoking			NA			
Ex-smoker	17.4 (0.62)	11.2 (0.42)		23.0 (0.23)	20.9 (2.37)	20.6 (0.57)
Alcohol abstainer						
Yes	64.1	67.9	12.3	38.3	85.3	32.5
No	35.9	32.1	87.7	61.7	14.7	67.5
Frequent heavy drinker						
Yes	6.5	19.5	8.2	4.9	27.1	4.5
No	93.5	80.5	91.8	95.1	72.9	95.5
Infrequent heavy drinker						
Yes	2.8	8.0	6.3	3.9	15.1	4.9
No	97.2	92.0	93.7	96.1	84.9	95.1
Number days per week with alcoholic drinks						NA
Less than once a week/once a week	69.6	24.5	40.3	54.6	26.2	
Twice/three times a week	10.5	13.0	27.2	22.3	26.4	
Four/five times a week	4.1	5.1	14.7	9.5	17.8	
Six/seven times a week	15.8	57.4	17.8	13.6	29.6	
Mean number of drinks per day (SE)	1.4 (0.09)	2.1 (0.05)	3.0 (0.04)	2.2 (0.03)	4.9 (0.38)	2.3 (0.03)
Frequency of participation in vigorous physical activity						
Less than once a week	96.0	69.5	69.6	61.5	72.1	52.5
Once a week	0.3	1.2	10.1	10.8	3.6	13.4
More than once a week	3.7	29.3	20.3	27.7	24.3	34.1

(Continued)

Table 2. Continued

	SAGE (W1) ($n = 35\,125$)	CHARLS (W1) ($n = 13\,817$)	ELSA (W5) ($n = 8985$)	HRS (W10) ($n = 20\,337$)	LASI (Pilot) ($n = 1109$)	SHARE (W4) ($n = 56\,755$)
Frequency of participation in moderate/mild physical activity						
Less than once a week	28.4	15.1	9.4	11.3	60.4	19.8
Once a week	2.7	2.2	8.0	17.9	3.9	14.4
More than once a week	68.9	82.7	82.6	70.8	35.7	65.8
Sedentary						
Yes	27.8	12.8	7.1	5.7	49.5	12.1
No	72.2	87.2	92.9	94.3	50.5	87.9

W, Wave; NA, not available.

[a]ELSA have been measured in Wave 4 (2008).

[b]Multimorbidity for SAGE, since it is based on five conditions (cancer is not available), is not comparable to the other countries (based on six conditions).

[c]HRS did not administer the time orientation test to participants aged ≤64.

in CHARLS and 21.2% in SAGE, compared with 11.7% in SHARE, 8.1% in ELSA, 7.0% in HRS and 3.8% in LASI. As in SAGE+ Wave 1, HRS respondents had the highest prevalence of all chronic health conditions, except for angina which was highest in SAGE, and for lung diseases which was higher in CHARLS. Hypertension prevalence based on measured blood pressure was higher than self-report in all surveys, with hypertension better controlled on treatment in ELSA and HRS. In SAGE and CHARLS, the majority of respondents classified as hypertensive did not have their condition controlled, but a considerable proportion of respondents were hypertensive at measurement, even with treatment, in the other surveys also.

Regarding the three mental health status variables (ability to remember, feelings of depression/sadness and trouble sleeping), CHARLS had the highest percentages of poor outcomes. Compared with SAGE+ Wave 1, SAGE respondents in SAGE+ Wave 2 felt less depressed/sad (38.4% vs 57.9%) and reported fewer problems with sleeping (51.6% vs 57.0%); for ELSA, HRS and SHARE the percentages were similar.

Near and distance vision showed substantial differences across studies, with SAGE and SHARE participants reporting the highest percentage of having 'a lot' of difficulties in reading and seeing things at a distance, whereas ELSA and LASI subjects had the least difficulties. Percentages were comparable to SAGE+ Wave 1.

ELSA, HRS and SHARE had prevalence of overweight varying from 37.6% (HRS) to 41.9% (SHARE) and obesity from 20.1% (SHARE) to 34.0% (HRS). SAGE and LASI had the highest percentage of underweight respondents (15.4% and 26.0%, respectively).

The prevalence of current smokers was similar across ELSA, HRS and SHARE whereas it was higher in SAGE (33.8%) and CHARLS (28.9%). CHARLS and LASI had a higher mean intensity of smoking; SAGE showed a lower mean number of cigarettes consumed. Drinking habits also varied substantially across studies. Inn ELSA, 88%were not abstainers, whereas 85.3% in LASI were abstainers. Heavy drinkers were lowest in SAGE, HRS and SHARE.

Compared with SAGE+ Wave 1, there was an increase in the prevalence of frequent heavy drinkers in ELSA (8.2% vs 1.0%). This is likely due to the difference in the questions asked. In 2004 the question was 'number of different types of drink consumed on heaviest day in the last week', whereas in 2010 it was 'number of drinks the respondent had in the last 7 days'.

LASI and SAGE showed the highest prevalence of sedentary responders (49.5% and 27.8%, respectively). Compared with SAGE+ Wave 1, the frequency of participation in vigorous/moderate activity was unchanged for ELSA, HRS and SHARE, whereas SAGE Wave 1 reported a higher prevalence of respondents that performed vigorous activities less than once a week (96.0% vs 60.0% for SAGE+ Wave 2). This difference is because the question in SAGE Wave 0 did not separate activities done at work from those done for sports/fitness/recreational, but in SAGE Wave 1 two different questions were asked, and only the one related to sports/fitness/recreational activities was included in the harmonization.

Differences in physical functioning and performance tests

Table 3 shows the prevalence of ADL and IADL limitations, and of specific aspects of mobility/functioning. Any level of dependency (limitation in one or more activities) varied between 11.6% in SHARE and 39.8% in SAGE for ADLs, and from 14.0% in ELSA to 27.5% in HRS for

Table 3. Physical functioning and performance test characteristics: age-standardized prevalence ratios or mean value for SAGE+ Wave 2 (year 2010), by survey

	SAGE (W1) ($n = 35\,125$)	CHARLS (W1) ($n = 13\,817$)	ELSA (W5) ($n = 8985$)	HRS (W10) ($n = 20\,337$)	LASI (Pilot) ($n = 1109$)	SHARE (W4) ($n = 56\,755$)
Activities of Daily Living (ADLs) limitations[a]						
Independent	60.2	76.8	82.9	66.1	82.6	88.4
Mild	22.9	16.4	12.4	23.4	11.0	8.2
Moderate	8.2	5.2	3.3	7.4	3.6	1.9
Severe/cannot do	8.7	1.6	1.4	3.1	2.8	1.5
Instrumental ADLs limitations[a]	NA				NA	
Independent		82.8	86.0	72.5		84.9
Mild		12.5	11.1	22.4		10.5
Moderate		2.3	1.9	3.6		2.5
Severe/cannot do		2.4	1.0	1.5		2.1
Mobility limitation						
Walking 100 m	34.1	19.8	12.4	46.0	NA	10.5
Sitting for about 2 h	40.6	NA	13.5	18.6	NA	11.3
Getting up from a chair after sitting for long periods	39.3	28.7	24.5	35.4	NA	19.1
Climbing several flights of stairs without resting	NA	34.9	32.6	38.8	NA	28.1
Climbing one flight of stairs without resting	51.3	NA	13.6	34.1	NA	11.6
Stooping, kneeling or crouching	50.9	27.1	35.3	41.5	NA	30.3
Reaching or extending arms above shoulder level	23.5	7.5	10.4	15.3	NA	10.8
Pulling or pushing large objects like a living-room chair	NA	NA	16.6	23.5	NA	14.4
Lifting or carrying weights over 5 kg	51.1	7.9	21.5	20.5	NA	22.0
Picking up a small coin from a table	18.5	3.0	5.2	6.0	NA	4.2
Walking test						NA
Mean time over 2.5 m (s) (SE)[b]	2.7 (0.03)	4.9 (0.04)[c]	3.3 (0.03)[c,d]	3.7 (0.04)[e]	4.6 (0.17)[c]	
Grip strength						
Mean (kg) in dominant hand (SE)	24.3 (0.45)	28.9 (0.12)	30.8 (0.14)	32.5 (0.17)	19.3 (0.56)	32.9 (0.14)

W, wave; NA, not available.

[a]Since limitations on ADL/IADL in CHARLS are based on five items, this is not comparable with the other countries (based on 6 and 7 items, respectively).

[b]Mean values are only weighted and not age-standardized, since the test was administered at different age groups across surveys.

[c]CHARLS, ELSA and LASI did not administer the walking test to participants aged ≤59.

[d]In ELSA have been measured in Wave 4 (2008).

[e]HRS did not administer the walking test to participants aged ≤64.

IADLs. Compared with SAGE+ Wave 1, there was an increase in the level of dependency for both ADLs and IADLs in HRS and SHARE, whereas ELSA showed a decrease. Grip strength measurements varied across studies, with LASI showing the lowest value (19.3) and HRS and SHARE the highest (32.5 and 32.9, respectively).

Differences in social network and life satisfaction

The SNI and life satisfaction variables had many missing data, as indicated in Table 4. About 23% of HRS participants had a high level of ties (SNI = 4) compared with 4.3% in LASI and 2.3% in ELSA. The SNI mean value was higher in HRS (2.7 on scale of 0 to 4) than LASI (2.2) and ELSA (1.7). The most satisfied with their life were individuals in CHARLS and SHARE (85.1%), and the least satisfied were SAGE participants (66.2%).

Strengths and weaknesses

The main strengths of this ex-post harmonization effort of ageing studies are the rigorous and systematic harmonization protocol, and the comparability of nine domains across two waves (Wave 1 (2004) and Wave 2 (2010)) of multiple studies, allowing longitudinal analysis and investigations of trends in health-related issues. The documentation in writing and STATA code provides a transparent and detailed record of the process and the foundation for a reliable dataset. The usefulness of well-harmonized data has considerable potential for population health and public policy. On the other hand, the interpretation of longitudinal findings should always be framed

Table 4. Social Network Index and life satisfaction characteristics: age-standardized prevalence ratios or mean value for SAGE+ Wave 2 (year 2010), by survey

	SAGE (W1) ($n = 35\,125$)	CHARLS (W1) ($n = 13\,817$)	ELSA (W5) ($n = 8985$)	HRS (W10) ($n = 20\,337$)	LASI (Pilot) ($n = 1109$)	SHARE (W4) ($n = 56\,755$)
Social Network Index[a]	NA	NA				NA
0 (no ties)			2.0	0.6	0.5	
1			32.6	11.3	14.1	
2			57.7	30.3	57.4	
3			5.4	35.0	23.7	
4 (high level of ties)			2.3	22.8	4.3	
Mean SNI	NA	NA	1.7 (0.01)	2.7 (0.02)	2.2 (0.03)	NA
Life satisfaction[b]						
Satisfied	66.2	85.1	79.4	73.8	71.3	85.1
Dissatisfied	33.8	14.9	20.6	26.2	28.7	14.9

W, wave; NA, not available.

[a]ELSA has about 42% missing data; LASI has about 56% missing data; HRS has about 67% missing data (items have been asked in the self-completion questionnaire).

[b]HRS has 63% missing data (item has been asked in the self-completion questionnaire).

in view of the limitations of the ex-post harmonization exercise, which mainly lie in the effects of the social and cultural aspects. The variables included in the domains for the harmonized dataset may not be as detailed as those which would be available in the individual datasets; however, these data do facilitate cross-study comparisons. Understanding the specific characteristics of each study is still important when undertaking interpretation of results.

Data resource access

The SAGE+ Wave 2 microdata, along with further details about the harmonization process and all metadata, are available through the WHO data archive at [http://apps.who.int/healthinfo/systems/surveydata/index.php/catalog]. Further information and enquiries can be made to [sagesurvey@who.int] or the corresponding author.

Profile in a nutshell

- The generation of ex-post harmonization datasets across two waves [Wave 1 (2004) and Wave 2 (2010)] allows longitudinal analysis and investigations of trends in health-related issues.
- The ex-post harmonized process was applied to nine health-related thematic domains: sociodemographic and economic characteristics, health states, overall self-report of health and mental state, health examinations, physical and mental performance tests, risk factors, chronic conditions, social network and subjective well-being.
- Differences in health status and functioning domains across countries may be addressed from a public health perspective to develop country-specific intervention programmes.
- The SAGE + Wave 2 microdata are available through the WHO data archive.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

SAGE+ Wave 2 was supported by the US National Institute on Aging through grant R21AG034263, 'Harmonizing health outcomes and determinants across longitudinal studies on aging'. SAGE is supported by WHO and the Division of Behavioral and Social Research (BSR) at the National Institute on Aging (NIA) through Interagency Agreements (OGHA 04034785; YA1323-08-CN-0020; Y1-AG-1005–01) with WHO, and Research Project Grant R01AG034479. In addition, the governments of China, Ghana and South Africa have provided financial or other support for their national studies.

Acknowledgements

SAGE+ gratefully acknowledges that respondents from each of the studies have given considerable time and energy to this research. The study teams have also given time and expertise to ensure that the results are disseminated and data used to inform policy and practice. Mr Yunpeng Huang worked with the harmonization team to generate the required metadata documentation.

Conflict of interest: The authors declare that they have no competing interests.

References

1. Minicuci N, Naidoo N, Chatterji S, Kowal P. Data Resource Profile: Cross-national and cross-study sociodemographic and health-related harmonized domains from SAGE plus ELSA, HRS and SHARE (SAGE+, Wave 1). *Int J Epidemiol* 2016;**45**: 1403.
2. Kowal P, Chatterji S, Naidoo N. Data Resource Profile: The World Health Organization Study on global AGEing and adult health (SAGE). *Int J Epidemiol* 2012;**41**:1639–49.
3. Steptoe A, Breeze E, Banks J, Nazroo J. Cohort Profile: The English Longitudinal Study of Ageing. *Int J Epidemiol* 2013;**42**: 1640–48.
4. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JWR, Weir DR. Cohort Profile: The Health and Retirement Study (HRS). *Int J Epidemiol* 2014;**43**:576–85.
5. Börsch-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.
6. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS). *Int J Epidemiol* 2014;**43**:61–68.
7. Arokiasamy P, Bloom D, Lee J, Feeney K, Ozolins M. Longitudinal Aging Study in India: vision, design, implementation, and some early results. In: Smith JP and Majmundar M (eds). *Aging in Asia: findings From New and Emerging Data Initiatives*. Washington, DC: National Academies Press, 2012.
8. International Household Survey Network. *Metadata Standards – IHSN Technical Note. Godalming*. UK: Metadata Technology, 2012.
9. Ahmad OB, Boschi-Pinto C, Lopez AD, Murray CJL, Lozano R, Inoue M. *Age Standardization of Rates: A New WHO Standard GPE Discussion Paper Series (Vol. No.31)*. Geneva: WHO, 2001.