WHO's Study on global AGEing and adult health (SAGE)

<table>
<thead>
<tr>
<th>Country</th>
<th>Data collection agency</th>
<th>Principal Investigator</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>China Centre for Disease Control (CCDC)</td>
<td>Dr. Wu Fan</td>
</tr>
<tr>
<td>Ghana</td>
<td>Department of Community Health, Ghana Medical School</td>
<td>Prof R. Biritwum</td>
</tr>
<tr>
<td>India</td>
<td>International Institute of Population Sciences (IIPS)</td>
<td>Prof P. Arokiasamy</td>
</tr>
<tr>
<td>Mexico</td>
<td>Instituto Nacional de Salud Pública (INSP)</td>
<td>Dr. R. Lopez Ridaura</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>Russian Academy of Medical Sciences (RAMS)</td>
<td>Prof T. Maximova.</td>
</tr>
<tr>
<td>South Africa</td>
<td>Human Science Research Council (HSRC)</td>
<td>Prof R. Phaswanamafuya</td>
</tr>
</tbody>
</table>

Response rates

<table>
<thead>
<tr>
<th></th>
<th>HH Response Rate %</th>
<th>HH Cooperation Rate %</th>
<th>IND Response Rate %</th>
<th>IND Cooperation Rate %</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>95</td>
<td>99</td>
<td>93</td>
<td>98</td>
</tr>
<tr>
<td>Ghana</td>
<td>86</td>
<td>98</td>
<td>80</td>
<td>92</td>
</tr>
<tr>
<td>India</td>
<td>88</td>
<td>92</td>
<td>68</td>
<td>92</td>
</tr>
<tr>
<td>Mexico</td>
<td>59</td>
<td>59</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>Russia</td>
<td>79</td>
<td>79</td>
<td>83</td>
<td>88</td>
</tr>
<tr>
<td>South Africa</td>
<td>67</td>
<td>99</td>
<td>77</td>
<td>100</td>
</tr>
</tbody>
</table>

Response rate= % of persons who completed the interviews amongst all eligible persons, including those who were not successfully contacted.
Cooperation rate= % of persons who completed the interviews amongst all eligible persons contacted.

Methodology/Sampling

The unit of analysis for all surveys is the household and individual. The target population is adults, 18 years and older, residing within individual households, institutionalised population excluded. However, since the major focus of SAGE is the health and its determinants in older adults, a much larger sample of respondents who are 50 years and older is interviewed in SAGE with a smaller comparative sample of those 18-49 years in age. The sizes of these individual groups vary from country to country as detailed below. The mode of data collection was face-to-face paper and pencil interviews (PAPI) for Ghana, India, Russian Federation and South Africa, face-to-face computer assisted personal interviews (CAPI) in Mexico and 50% CAPI and 50% PAPI interviews in China. The type of data collected were survey and biomarker data. The data collection period was 2007-2010.
China

Baseline data: SAGE Wave 1

SAGE Wave 1

Sampling
The People’s Republic of China administers 22 provinces, four municipalities, five autonomous regions and two special administrative regions. A province is an administrative division. Provinces were divided into three groups (central, eastern and western) based on their economic development and geographic location. Of the 22 provinces, 10 are classified as central, eight as eastern and four as western. Seven provinces were selected at random: two central, two western and three eastern provinces. Shanghai which is a municipality was selected with certainty.

China implemented a stratified, multistage cluster sampling design. The sample was stratified by provinces, eight in total, and is nationally representative. Guangdong, Shandong, Zhejiang and Shanghai are located in the east and represent the developed provinces of China. Hubei and Jilin are located in the middle region of China. Hubei, though it is located in eastern China, its economic status is similar to that of the middle provinces of China. Shaanxi and Yunnan provinces are located in western China and represent the less developed provinces.

One county in the rural regions and one district/city in the urban regions, was randomly selected from each province. Anji, Gucheng, Nanxiong, Meixian, Ji’an, Luwan, Ruschan and Tonghai were the urban cities selected while Wucheng, Wujiaqiang, Tianhe, Wangyi, Fengman, Qingpu, Lixia and Panglong the rural counties.

Four townships in each selected county and four community blocks in each selected city were selected probability proportional to size and constituted the PSUs. The measure of size being the number of households in the township or community block.

Two villages in each sampled township and two neighbourhood communities in each community block were selected probability proportional to size and constituted the SSUs. The measure of size being the number of households in the village or neighbourhood communities.

From each village and each neighbourhood community, two residential blocks were selected using cluster random sampling.

The CDC maintains an updated list of all households with persons aged 50 years and older (50+). Thirty-five households were randomly selected from each residential block using this list. The CDC also maintains an updated list of all households with persons aged 18-69. Excluding the households already selected seven households with a person aged 18-49 were randomly selected from the second list.

Eighty four households in total were randomly selected from each sampled village or neighbourhood community forming the TSUs. In the final stage, in an 18-49 household, one eligible member was selected using Kish tables. All members aged 50 years and older were eligible from 50+ households (with multiple individual interviews possible).

Stages of selection
Strata: Province = 8
PSU: Township or Community = 64 surveyed
SSU: Village or Neighbourhood Community = 127 surveyed
TSU: Household = 10 752 surveyed
QSU: Individual = 15 050 surveyed

Weights
Household weights for analysis at household level and individual weights for analysis at person level were calculated. These were based on the selection probability at each stage of selection. Household weights are post-stratified by province and locality according to the 2008 household projections provided by China CDC and weight up to the total number of households in the eight provinces. Individual weights are post-stratified by province, sex and age groups (18-49, 50-59, 60-69, 70+) according to the 2008 population projections provided by China CDC and weight up to the total number of persons aged 18+. A second set of household and individual weights are available which are post stratified to weight up to the entire household and 18+ populations respectively in the PRC (nationally representative of these populations). Hong Kong and Macau are excluded from national totals. Weights are not normalized.
Ghana

Baseline data: WHS/SAGE Wave 0, 2003

WHS/SAGE Wave 0

Sampling
Ghana used a stratified, multistage cluster design. The sample was stratified by administrative region and type of locality (urban/rural) resulting in 18 strata and is nationally representative. The sample was selected in three stages. Three hundred enumeration areas (EAs), which constitute PSUs, were selected probability proportional to size. The measure of size being the number of households in the PSU. From each EA, twenty households were selected using systematic sampling. In the final stage, one eligible individual was selected from the household using Kish tables.

Stages of selection
Strata: Region, Locality= 18
PSU: Enumeration Areas=289 surveyed
SSU: Households=5 662
TSU: Individual=5 662

SAGE Wave 1

Sampling
Ghana used a stratified, multistage cluster design. The sample was stratified by administrative region (Ashanti, Brong Ahafo, Central, Eastern, Greater Accra, Northern, Upper East, Upper West, Volta and Western) and type of locality (urban/rural) resulting in 20 strata and is nationally representative. The Census Enumerated Areas (CEA) of the 2000 Population and Housing Census was used as the sampling frame. A sample of 251 EAs was selected as the primary sampling units (PSU). The number of EAs to be selected from each strata was based on proportional allocation (determined by the number of EAs in each strata specified on the census frame). EAs were then selected from each stratum with probability proportional to size; the measure of size being the number of individuals aged 50 years or more in the EA. In each selected EA, a listing of the households was conducted to classify each household into the following mutually exclusive categories:
1) WHS/SAGE Wave 0 follow-up households with one or more members aged 50 years or more;
2) New households with one or more members aged 50 years or more;
3) WHS/SAGE Wave 0 follow-up households which did not include any members aged 50 years or more, but included residents aged 18–49; and,
4) New households which did not include any members aged 50 years or more, but included residents aged 18–49.

Twenty-four households were randomly selected from each selected EA. All WHS/SAGE Wave 0 follow-up 50-plus households were eligible for the household interview (one household respondent was selected). Twenty such households were selected. If this target number was not reached, then the balance was selected using systematic sampling from the new 50-plus households. All 50+ members of the household were eligible for the individual interview (multiple individual interviews possible in these households).

Four WHS/SAGE Wave 0 follow-up 18–49 households were randomly selected, and the WHS/SAGE Wave 0 respondent was eligible for the individual interview. If this target number was not reached, then the balance was selected using systematic sampling from the new 18–49 households. In each of these households one person aged 18–49 was eligible for the individual interview, and the individual to be included was selected using a Kish tables.

Stages of selection
Strata: Region, Locality = 20
PSU: Enumeration Areas = 235 surveyed
SSU: Households = 5 269 surveyed
TSU: Individual = 5 573 surveyed

Weights
Household weights for analysis at household level and individual weights for analysis at person level were calculated. These were based on the selection probability at each stage of selection.

Household weights were post-stratified by region and locality according to the 2010 household projections provided by Stats Ghana. Individual weights were post-stratified by region, locality, sex and age-groups (18–49, 50-59, 60-69, 70+) according to the 2009 projected population estimates provided by Stats Ghana. Weights are not normalized.
India

Baseline data: WHS/SAGE Wave 0, 2003

WHS/SAGE Wave 0

Sampling
India has 28 states and seven union territories. The following were excluded because of the small population and lack of accessibility: the seven union territories, seven northern states (Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura) and Goa. Jammu & Kashmir were excluded as a result of operational difficulties. Of the 28 states, 19 were included in the design which covered 96% of the population.

India used a multistage, stratified clustered sample design. The selection of states was done in accordance with their geographic location and level of development. The 19 states were grouped into six regions: north, central, east, north east, west and south. The level of development was based on four indicators namely: infant mortality rate, female literacy rate, percentage of safe deliveries and per capita income. A composite index of the level of development was computed by giving equal weightage to the four indicators. The states were classified into six levels (in decreasing order) of development based on the composite index as shown in table below.

<table>
<thead>
<tr>
<th>Region</th>
<th>Level of Development</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>North</td>
<td>Punjab</td>
</tr>
<tr>
<td>Central</td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>West Bengal</td>
</tr>
<tr>
<td>North East</td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>Maharashtra</td>
</tr>
<tr>
<td>South</td>
<td>Kerala, Tamil Nadu</td>
</tr>
</tbody>
</table>

Note: States in bold were selected for the survey.

The states were selected randomly such that one state was selected from each region as well as from each level of development category. The sample was stratified by state and locality (urban/rural) resulting in 12 strata and is nationally representative.

Two stage and three stage sampling was adopted in rural and urban areas, respectively.

In the WHS/SAGE Wave 0, in rural areas PSUs (villages) were selected probability proportional to size. The measure of size being the 2001 Census population in the village. SSUs (households) were selected using systematic sampling. TSUs (individuals) were selected using Kish tables.

In urban areas, PSUs (city wards) were selected probability proportional to size. SSUs (census enumeration blocks), two were randomly selected from each PSU. TSU (households) were selected using systematic sampling. QSU (individuals) were selected as in rural areas.

SAGE Wave 1

Sampling
The SAGE sample was pre-determined as all PSUs and households selected for the WHS/SAGE Wave 0 survey were included. Exceptions are three PSUs in Assam which were replaced as they were inaccessible due to flooding. And a further six PSUs were omitted for which the household roster information was not available.

In each selected EA, a listing of the households was conducted to classify each household into the following mutually exclusive categories:
1. Households with a WHS/SAGE Wave 0 respondent aged 50-plus: all members aged 50-plus including the WHS/SAGE Wave 0 respondent were eligible for the individual interview.

2. Households with a WHS/SAGE Wave 0 respondent aged 47-49: all members aged 50-plus including the WHS/SAGE Wave 0 respondent aged 47-49 was eligible for the individual interview.

3. Households with a WHS/SAGE Wave 0 female respondent aged 18-46: all females members aged 18-49 including the WHS/SAGE Wave 0 female respondent aged 18-46 were eligible for the individual interview.

4. Households with a WHS/SAGE Wave 0 male respondent aged 18-46: three households were selected using systematic sampling and one male aged 18-49 was eligible for the individual interview. In the households not selected, all members aged 50-plus were eligible for the individual interview.

**Stages of selection**
- Strata: State, Locality =12
- PSU: Village or CEB =375 surveyed
- SSU: Household =10 424 surveyed
- TSU: Individual =12 198 surveyed

**Weights**
Household weights for analysis at household level and individual weights for analysis at person level were calculated. These were based on the selection probability at each stage of selection.

Household weights were post-stratified by the six states and locality according to the 2006 household projections obtained from Population Projections for India and States 2001-2026: Report of the technical group on population projections constituted by the national commission on population May 2006, Office of the Registrar General and Census Commissioner India.

Individual weights were post-stratified by the six states, locality, sex and age-groups (18-49, 50-59, 60-69, 70+) according to the 2006 projected population estimates. A second set of household and individual weights are available which are post-stratified to weight up to the number of households and 18-plus populations respectively in the country. Weights are not normalized.
Mexico

Baseline data: WHS/SAGE Wave 0, 2003

WHS/SAGE Wave 0

Sampling
Mexico used a stratified multistage cluster design. The sample is stratified by type of federal state and locality (urban/metropolitan/rural) resulting in a total of 96 strata and is nationally representative.

The Basic Geo-Statistical Areas (AGEB) defined by the National Institute of Statistics (INEGI) was used as the sampling frame. An AGEB constitutes a PSU. AGEBs were selected probability proportional to size with the measure of size being the number of inhabitants in the Federal State. Manzanas constitute the SSUs in the urban areas and were selected with probability proportional to size, the measure of size being the number of inhabitants in the manzana. The SSUs in rural areas are households. Fifty households were randomly selected from each SU. In urban areas, households constitute the TSUs. Approximately seven households were randomly selected. In the final stage, one eligible person, 18 years and above, was selected from each household using Kish Tables.

Stages of selection
Strata: Federal State, Locality = 31*3 = 93 (one state Colima omitted)
PSU: AGEB =797
SSU: Manzana =4 433
TSU: Household =40 000
QSU: Individual =40 000

SAGE Wave 1

Sampling
Mexico used a stratified multi-stage cluster sample design. Strata were defined by locality (metropolitan, urban, rural). The Basic Geo-Statistical Areas (AGEB) defined by the National Institute of Statistics (INEGI) was used as the sampling frame. An AGEB constitutes a PSU.

The locality of the PSU was defined as:
   a) Rural - Less than 2 500 inhabitants;
   b) Urban - Less than 100 000 inhabitants and more than 2 499 inhabitants; and,
   c) Metropolitan - More than 99 999 inhabitants and State capitals.

The sample size of SAGE Wave 1 is considerably smaller than that of the WHS/SAGE Wave 0; therefore, in order to obtain a sample for SAGE Wave 1 with less geographical dispersion than that of the WHS/SAGE Wave 0 sample, a sub-sample of 211 PSUs were selected from the 797 WHS PSUs.

PSUs were selected probability proportional to three factors:
   a) (WHS/SAGE Wave 0 50plus): number of WHS/SAGE Wave 0 50-plus interviewed at the PSU,
   b) (State Population): population of the state to which the PSU belongs,
   c) (WHS/SAGE Wave 0 PSU at county): number of PSUs selected from the county to which the PSU belongs for the WHS/SAGE Wave 0;

for instance, if two PSUs in the county “Aguascalientes” were selected for WHS/SAGE Wave 0, then, for such a PSU, the factor (WHS/SAGE Wave 0 PSU at county) would be equal to two. The first and third factors were included to reduce geographic dispersion. Factor two affords states with larger populations a greater chance of selection.

All WHS/SAGE Wave 0 individuals aged 50 years or over in the selected rural or urban PSUs and a random sample 90% of individuals aged 50 years or over in metropolitan PSUs who had been interviewed for the WHS/SAGE Wave 0 were included in the SAGE Wave 1 "primary" sample. The remaining 10% of WHS/SAGE Wave 0 individuals aged 50 or over in metropolitan areas were then allocated as a "replacement" sample for individuals who could not be contacted or did not consent to participate in SAGE Wave 1. A systematic sample of 1000 WHS/SAGE Wave 0 individuals aged 18-49 across all selected PSUs was selected as the "primary" sample and 500 as a "replacement" sample.

This selection process resulted in a sample which had an over-representation of individuals from metropolitan strata; therefore, it was decided to increase the number of individuals aged 50 or over from rural and urban strata. This was achieved by including individuals who had not been part of WHS/SAGE Wave 0 (which became a "supplementary" sample), although the household in which they lived included an individual from WHS/SAGE Wave 0. All individuals aged 50 or over were included from rural and urban "18-49 households" (that is, where an individual aged 18-49 was
included in WHS/SAGE Wave 0) as part of the “primary supplementary” sample. A systematic random sample of individuals aged 50 or over was then obtained from urban and rural households where an individual had already been selected as part of the 50 and over or 18-49 samples. These individuals then formed part of the “primary supplementary” sample and the remainder (that is, those not systematically selected) were allocated to the "replacement supplementary" sample. Thus, all individuals aged 50 or over who lived in households in urban and rural PSUs obtained for SAGE Wave 1 were selected as either a primary or replacement participant. A final "replacement" sample for the 50 and over age group was obtained from a systematic sample of all individuals aged 50 or over from households which included the individuals already selected for either the 50 and over or 18-49. This sampling strategy also provided participants who had not been included in WHS/SAGE Wave 0, but lived in a household where an individual had been part of WHS/SAGE Wave 0 (that is, the "supplementary" sample), in addition to follow-up of individuals who had been included in the WHS/SAGE Wave 0 sample.

<table>
<thead>
<tr>
<th>Population</th>
<th>Stage</th>
<th>Rural</th>
<th>Urban</th>
<th>Metropolitan</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-plus at WHS/SAGE Wave 0</td>
<td>1</td>
<td>498</td>
<td>415</td>
<td>2187</td>
<td>3100</td>
</tr>
<tr>
<td>50-plus Panel Replacement</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>244</td>
<td>244</td>
</tr>
<tr>
<td>18-49 at WHS/SAGE Wave 0</td>
<td>3</td>
<td>132</td>
<td>125</td>
<td>744</td>
<td>1001</td>
</tr>
<tr>
<td>18-49 Panel Replacement</td>
<td>4</td>
<td>61</td>
<td>61</td>
<td>375</td>
<td>497</td>
</tr>
<tr>
<td>50-plus Sample survey</td>
<td>5</td>
<td>59</td>
<td>126</td>
<td>0</td>
<td>185</td>
</tr>
<tr>
<td>50-plus living were 18-49 plus was at WHS/SAGE Wave 0</td>
<td>6</td>
<td>246</td>
<td>198</td>
<td>0</td>
<td>444</td>
</tr>
<tr>
<td>50-plus Sample survey. Replacement</td>
<td>7</td>
<td>334</td>
<td>196</td>
<td>853</td>
<td>1383</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1591</td>
<td>1450</td>
<td>4403</td>
<td>7443</td>
</tr>
</tbody>
</table>

**Stages of selection**

- **Strata:** Locality = 3
- **PSU:** AGEBs = 211
- **SSU:** Households = 4,968 surveyed
- **TSU:** Individual = 5,449 surveyed

**Weights**

Household weights for analysis at household level and individual weights for analysis at person level were calculated. These were based on the selection probability at each stage of selection.

Household weights were post-stratified by locality according to the 2009 population census projections. Individual weights were post-stratified by sex and age-groups (18-34, 35-49, 50-59, 60-105) according to the 2009 population census projections. Weights are divided by 1000 and are not normalized.
Russian Federation

Baseline data: WHS/SAGE Wave 0, 2003

WHS/SAGE Wave 0

Sampling
Russia is divided into seven federal districts: Central, Southern, Northwestern, Far Eastern, Siberian, Urals and Volga. The sample was stratified by federal district. Three of the four federal districts from the European districts were selected: Central, Northwestern and Volga, which cover approximately 60 percent of the population. The Southern Federal district was omitted due to political unrest in Chechnya. The intention was to include the Asian federal districts at a later stage but this idea was abandoned. The sample is therefore a regional sample. Atenum, which are polling districts, constitute the PSUs. One hundred and fifty atenum were selected probability proportional to size, the measure of size being the number of inhabitants aged 18 and above in the Atenum. Of the 150 Atenum included in the sample, 27 have no households. Household which were randomly selected constitute the SSUs. In the final stage one eligible person, 18 years and above, is selected from each household using Kish Tables.

Stages of selection:
Strata: Federal District = 3
PSU: Atenum = 123 surveyed
SSU: Household = 4,427
TSU: Individual = 4,427

SAGE Wave 1

Sampling
All seven federal districts were included in the sample and constitute the strata. The sample is nationally representative. Moscow city and St Petersburg which are federal cities and Moscow oblast were included with absolute certainty. Atenum, which are polling districts, constitute the PSUs. 267 atenum were selected probability proportional to size, the measure of size being the number of inhabitants aged 18 and above in the atenum. Households which were randomly selected constitute the SSUs. In the final stage, all persons aged 50 and over in 50-plus households were eligible for the individual interview, and one person aged 18–49 years was eligible for the individual interview.

Stages of selection
Strata: Federal District = 7
PSU: Atenum = 150 surveyed
SSU: Household = 4,644 surveyed
TSU: Individual = 4,947 surveyed

Weights
Household weights for analysis at household level and individual weights for analysis at person level were calculated. These were based on the selection probability at each stage of selection.

Household weights were post-stratified by federal district and locality according to the 2001 population census figures provided by the Russian Federal State Statistics Service (Rosstat). Individual weights were post-stratified by federal district, locality sex and age-groups (18-49, 50-59, 60-69, 70+) according to the mean predictive population estimates 2011 provided by Rosstat. Weights are not normalized.
South Africa

Baseline data: SAGE Wave 1

SAGE Wave 1

Sampling
South Africa used a stratified multi-stage cluster sample design. Strata were defined by the nine provinces: (Eastern Cape, Free State, Gauteng, Kwa-Zulu Natal, Limpopo, Mpumalanga, North West, Northern Cape and Western Cape), locality (urban or rural), and race (African/Black, White, Coloured and Indian/Asian). As not all combinations of stratification variables were possible, there were 50 strata in total.

The Human Science Research Council's master sample was used as the sampling frame which comprised 1000EAs. A sample of 600 EAs was selected as the primary sampling units (PSU).

The number of EAs to be selected from each strata was based on proportional allocation (determined by the number of EAs in each strata specified on the Master Sample). EAs were then selected from each strata with probability proportional to size; the measure of size being the number of individuals aged 50 years or more in the EA. In each selected EA, 30 households were randomly selected from the Master Sample.

A listing of the 30 selected households was conducted to classify each household into one of two mutually exclusive categories:

1) households with one or more members aged 50 years or more (defined as '50-plus households'); and,
2) households which did not include any members aged 50 years or more, but included residents aged 18-49 (defined as '18-49 households').

All 50-plus households were eligible for the household interview, and all 50-plus members of the household were eligible for the individual interview. Two of the remaining 18-49 households were randomly selected for the household interview. In each of these households, one person aged 18-49 was eligible for the individual interview, and the individual to be included was selected using a Kish table grid.

Stages of selection

Strata: Province, Predominant Population Group, Locality = 50
PSU: EAs = 408 surveyed
SSU: Households = 4,020 surveyed
TSU: Individual = 4,227 surveyed

Weights
Household weights for analysis at household level and individual weights for analysis at person level were calculated. These were based on the selection probability at each stage of selection.

Household weights were post-stratified by province and locality according to the South African Community Survey 2007. Individual weights were post-stratified by province, sex and age-groups (18-49, 50-59, 60-69, 70+) according to the 2009 Medium Mid-Year population estimates from Statistics South Africa. Weights are not normalised.