



German Council for Social
and Economic Data (RatSWD)

www.ratswd.de

RatSWD

Working Paper Series

Working Paper

No. 94

Provision for Old Age

National and International Survey Data
to Support Research and Policy on Aging

Hendrik Jürges

July 2009

Working Paper Series of the Council for Social and Economic Data (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/ 2008 Heike Solga; 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

Provision for Old Age
National and International Survey Data to Support Research and Policy on Aging

Hendrik Jürges

Mannheim Research Institute for the Economics of Aging (juerges[at]mea.uni-mannheim.de)

Abstract

This report reviews recent trends in the collection of multidisciplinary and longitudinal data in the area of aging research, both in Germany and internationally. It also discusses important developments such as linkage with administrative records, the inclusion of health measurements and biomarkers, and the inclusion of populations in institutions, particularly nursing homes.

1. Research Questions

Population aging is one of the megatrends of the twenty-first century. In almost all countries of the developed world, mortality rates are falling, birth rates are below replacement rate, and work rates are falling. World wide, the number of people aged 60 and over is expected to triple until 2050. The aging of the population will shape the world to come and its political agenda. The main policy issues that arise with an aging population concern providing income and health security during old age at affordable budgets. To cope with these particular challenges of rapid population aging, it is important to improve the current scientific understanding of complex linkages between economic, health, and social factors that determine the quality of life in the older population. These interactions primarily take place at the individual level, they are dynamic - aging is a process, not a state - and are related to a country's welfare regime.

To improve our understanding of population aging and its policy implications, researchers need multidisciplinary and longitudinal data. Over the past decade, the international research community has responded to this need by starting to create a worldwide microdata infrastructure that helps researchers to better understand the individual and population aging processes. The aim of this special report is to document the contents and degrees of access to existing national and international datasets containing household- and individual-level information on the economic well-being and health of older populations. In this context, the international perspective is of great importance. Internationally comparative data is necessary in order to exploit the rich variety in policies, institutions, and other factors across different countries. The impact of public policy can be much better understood if one can observe one policy in relation to other policies. Many of the policies that might be considered as solutions to address future public policy challenges resulting from an aging population have already been implemented in some form in some country. For instance, comparisons of different pensions systems and their impact on old-age poverty rates, savings decisions of the working population, or the role of the family and intergenerational relations, can inform policy makers about the likely consequences of pension reform.

In addition to briefly describing the contents of and access to existing survey data sets, the present report also discusses future developments and further needs in the area of aging research. Three important areas in which such developments are likely and needed are (a)

linkage with administrative records, (b) the inclusion of (more) health measurements and biomarkers, and (c) the inclusion of populations in institutions, particularly nursing homes.

2. Status Quo: Data Bases and Access

This section briefly describes a selection of important national and international databases for multidisciplinary research on aging. Summary information and details on data access are given in tabular form in the appendix. Only databases that fulfill several criteria are listed. First, they must of course include - although not exclusively - coverage of the older population (defined as individuals aged 50 and older). Second, they must be based on non-administrative *research-driven* surveys. Third, they must be multidisciplinary; that is, they must include data with some detail from at least two of the following fields: medicine, economics, sociology, and psychology. Fourth, they must have a longitudinal design, because many events associated with aging are dynamic or longitudinal in nature. Obviously, these four criteria together seem to be fairly restrictive. For instance, they exclude the European Community Household Panel (ECHP), many of the health interview surveys carried out on national levels, and also the current European Health Interview Survey (EHIS).

2.1 National data sources for Germany

The Socio-Economic Panel (SOEP): Arguably, the most important data source of longitudinal micro data is the German Socio-Economic Panel (SOEP), which has collected detailed annual data mainly on housing, employment and income since 1984. One advantage of the SOEP is the large sample size. The sample currently includes about 22,000 respondents of which some 10,500 are aged 50 and above. Another advantage is the length of the panel. The SOEP now spans 25 years of annual data on the lives of a substantial number of respondents, following individuals from middle age into old age. The usefulness of the SOEP for certain aspects of aging research, however, has been limited by two facts: first, until recently, it contained little health-related information. For instance, self-rated general health has been included only in 1992, and information on respondents' smoking behavior, height, and weight is available since the late 1990s. However, a more systematic approach to the measurement of health has been taken since 2002, when SOEP included a version of the SF-12 health questionnaire. A more recent development has been the inclusion of physical measures or biomarkers, such as handgrip strength (in 2006). Second, although some broad information on savings and household assets is available annually, the quantitative composition of household assets was

covered only in 1998, 2002, and 2007, making it difficult to track in detail changes in the asset portfolios or in the amount of wealth.

The German Aging Survey (Alterssurvey-DEAS): DEAS is a cross-sectional and longitudinal survey of individuals aged 40 and over. The main topics it covers address circumstances and attitudes in later life, particularly related to housing, employment and retirement, volunteering, family, leisure activities and social participation. Data collection started in 1996 and is repeated every six years. In the first two waves, samples included about 5000 respondents. One advantage of the DEAS is that it collects highly detailed information from the respondents, particularly around psychological variables, whereas panel mortality due to the large gap between waves clearly is a disadvantage. Only 32 percent of the original sample members in 1996 could be reinterviewed in 2002. Also, data on the oldest-old (85+) are only available for reinterviewed panel data members, since the baseline sample in 1996 was restricted to individuals up to age 85.

Sparen und Altersvorsorge in Deutschland (SAVE): SAVE is the most elaborate survey on savings behavior in Germany. It contains detailed factual information on the current financial situation of households (including asset components), savings behavior, psychological determinants of saving, and health. SAVE was started in 2001 and data for several waves are now available. One drawback of the data for longitudinal analyses is the somewhat complex mixture of different samples that comprise it, which limits the exploitable longitudinal dimension.

2.2 *International data sources*

The *Cross-National Equivalent File (CNEF)* is an *ex post* harmonized dataset containing annual panel data from the U.S. Panel Study of Income Dynamics (PSID), the German Socio-Economic Panel (SOEP), the British Household Panel Study (BHPS), the Household Income and Labour Dynamics in Australia (HILDA), and the Canadian Survey of Labour and Income Dynamics (SLID). Future releases of the CNEF will include data from the Swiss Household Panel (SHP). Even for researchers planning research on only one of the contributing countries, CNEF offers the distinct advantage of providing a set of generated and fully imputed variables (in particular income components and equivalence weights) that are not directly available with the original surveys. These generated variables can be merged with the original survey data (and are part of the data distribution, e.g. of the SOEP). While being an excellent data source for comparative analyses of retirement and income security in old age, comparability issues have somewhat limited the use of the data for other fields such as health

(the 2006 CNEF release added a set of variables that measure health behaviors and health outcomes). One disadvantage of CNEF as a multinational data set is that the data cannot be accessed jointly and access is not harmonized across surveys.

The *Survey of Health, Ageing and Retirement in Europe (SHARE)* is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of individuals aged 50 or over. Individual interviews (overall N=33,000, about 3,000 respondents in Germany) are conducted with primary respondents and cohabitating partners. Fifteen European countries including Israel have so far contributed data to the first and/or second waves of SHARE conducted in 2004 and 2006 respectively. The survey's third wave of data collection will collect retrospective life histories in sixteen countries in 2008-09. SHARE is partly harmonized ex ante with the U.S. Health and Retirement Study (HRS) and the English Longitudinal Study on Ageing (ELSA), see below. Data collected include health self-reports, some biomarkers, psychological variables, economic variables (current work activity, sources and composition of wealth and current income), and social support variables (e.g., assistance and transfers within, volunteer activities). One noteworthy feature of the international SHARE database is the inclusion of anchoring vignettes for internationally comparative studies on a wide range of subjective survey data, such as health self-reports, life satisfaction, job satisfaction, satisfaction with health care, or political efficacy. Future waves of SHARE will include the collection of biomarkers and linkage with administrative records.

The *Generations and Gender Survey (GGS)* is a cross-national, multidisciplinary study of the dynamics of family relationships, covering non-institutionalized individuals aged 18-79. The main topics of the survey are related to the respondents' current and past family situation and family-related events, such as partnership formation, childbearing and leaving the parental home. The GGS is designed as a panel survey with at least three waves at an interval of three years. GGS wave 1 data are currently available for Bulgaria, France, Georgia, Germany, Hungary, and Russia.

The *English Longitudinal Study of Ageing (ELSA)* is an interdisciplinary biannual panel survey on health, economic position, and quality of life of older adults in England. ELSA covers the range of topics needed to study the economic, social, psychological, and health elements of individual and societal ageing. The initial ELSA sample (N=12,000), interviewed in 2002, was drawn from respondents (aged 50+) to the Health Survey for England (HSE), contributing baseline data on respondents' health (details of morbidity, lifestyle, diets, and blood samples). Covered topics are similar to SHARE: health and disability; cognitive

functioning; income and wealth; employment and retirement and post-retirement activities; social networks, support and participation. Biomarkers (blood pressure, blood samples, including genetic information, measured height and weight, lung function, grip strength, balance) are collected every four years during a separate visit by trained nurses. A life history interview has been conducted between regular waves in 2007.

The *Italian Survey on Household Income and Wealth (SHIW)* is unusual in the sense that it is conducted by a government agency (the Italian central bank - Banca D'Italia), has a strong research component, and grants free public online data access. All documentation is available in English. SHIW was started in the 1960s collecting data on income and savings. The survey has developed into a general household survey which includes detailed information on employment, wealth, financial decision making, and financial behavior.

The *Irish Longitudinal Study on Ageing (TILDA)* will interview a sample of some 10,000 individuals aged 55 and older living in the Republic of Ireland and collect detailed data on social networks and support, economic circumstances and health, including biomarkers. Three main waves of data collection are planned in 2008, 2013, and 2018, with annual telephone follow-ups and smaller-scale add-on studies. TILDA is specifically designed to deliver data comparable to the U.S. HRS and ELSA. Data access is not yet possible but planned to be free to scientific users.

The *U.S. Health and Retirement Study (HRS)* is the role model of many aging surveys currently conducted throughout the world, in part due to its scientific productivity. Since 1992, the HRS has generated more than 500 papers in peer-reviewed journals. Today, some 60 papers using HRS data appear in a refereed journal every year. The HRS is a biannual panel survey of Americans aged 50 and over, started in 1992. The current sample consists of about 22,000 respondents, providing detailed information on their physical and mental health, insurance coverage, financial status, family support systems, labor market status, and retirement planning. In addition to standard HRS questions, HRS has recently introduced an "enhanced" interview, collecting physical performance measures (grip strength, puff test, timed walk, balance), anthropometry (height, weight, waist), blood pressure, dried blood spots (HBA_{1c}, cholesterol, high-density lipoprotein, C-reactive protein), even salivary DNA (for extraction and storage only).

The *Mexican Health and Aging Study (MHAS)* is a panel study of respondents aged 50 and over in Mexico. At baseline (in 2001), about 15,000 individuals have been interviewed, providing information on health (self-reports and biomarkers), financial and time transfers between generations, sources, amounts of income, types and value of assets, and housing. The

survey design closely follows HRS and MHAS provides comparable data. Respondents have been reinterviewed once in 2003. New interviews are planned with MHAS wave 1 & 2 survivors and a representative sample of cohorts born 1951-1959.

The *Korean Longitudinal Study of Ageing (KLoSA)* is a biannual panel survey of approximately 10,000 South Korean residents (excluding Jeju Island) aged 45 or older, started in 2006. The 2006 main survey includes only respondents living in private households. Collected data include employment status, income, asset, family relations, health, and subjective judgment. KLoSA was designed to allow comparative studies with HRS, ELSA, or SHARE.

Based on SHARE baseline questionnaire, the first wave of data collection for the *Japanese Study on Aging and Retirement (JSTAR)* took place in the first half of 2007. Interviews have been conducted with some 4,300 individuals from five municipalities (Takikawa, Sendai, Adachi, Kanazawa, and Shirakawa). A second wave of data collection (extending the sample to include a sixth municipality) is currently being prepared and will begin in January 2009. Although JSTAR features many design elements of SHARE and HRS, there are also a few major differences. First, the sample was drawn in only five municipalities. One of the reasons for this design is that administrative health records are stored at the municipality level. Data linkage between these records and the survey data is planned. Second, the initial sample is restricted to individuals aged 50-75. Third, JSTAR interviews only one respondent per household. However, some limited information on cohabitating spouses or partners is gathered during the interview.

Aging studies for China, India, Thailand: Planning for initial waves of panel survey comparable to HRS and SHARE is underway in further Asian countries. These are CHARLS, the Chinese Aging and Retirement Longitudinal Study of Aging, HART, The Health and Retirement Study for Thailand, and LASI, the Longitudinal Survey of Aging in India.

3. Recommendations for Future Research

This section describes selected future needs in the area of empirical research on aging. The selection reflects ongoing innovative efforts in connection with some of the surveys described in the previous section (e.g., biomarkers), but also identifies issues that have received little attention so far, such as the systematic inclusion of individuals living in institutions (i.e., nursing homes).

3.1 *Data linkage*

Survey data, as described above, cover a wide range of topics. Information provided by respondents, however, is often incomplete and inaccurate. In contrast, administrative data are (ideally) complete and accurate but contain only very limited information, typically only information that is immediately related to the purpose of the data producer. The advantages of both types of data can be combined by linking administrative records to survey data. Benefits of linkage include (1) validation of respondents' self-reports, in particular if these reports are potentially subject to recall bias; (2) improved measurement of explanatory and dependent variables, reducing bias and increasing precision of model estimates; (3) reduction of respondent burden. For instance, aging surveys can benefit from adding social security records to explain retirement behavior or measurement of economic resources during retirement. Doctors' or health insurance records can be used to improve the measurement of health.

Researchers trying to link administrative data with survey data face several challenges. First, they need to get a unique ID from the respondent (e.g., social security number), which is needed to actually link the data. Asking for this ID (in some countries together with a written consent to link data) can have adverse effects on response or retention rates. Second, the availability of data that could be matched to general population surveys may be limited or require a great deal of cooperation from many agencies (such as getting medical records for German residents from public or private health insurers). Third, due to privacy legislation, data dissemination rules are often not able to conform to the standards set by the survey data to which the administrative data are linked. Restricted access to linked data will make cross-national analyses using linked data a very difficult enterprise. Here, new cross-national solutions of data access that fulfill all legal confidentiality requirements are needed.

3.2 *Biomarkers*

One important recent development in social survey research is the integration of biomarkers. Biomarkers are often associated with genetic information such as DNA samples. The purpose of genetic research in social sciences is not to find *the* gene for some socially relevant trait. Rather, the collaboration of geneticists with social scientist is fruitful because social scientists are experts in measuring social phenomena that may or may not be associated with genes. The two important contributions of social scientists to genetic research are, first, to help to establish the importance of non-genetic factors or interactions of genes and environment, and second, their familiarity with using large scale social, nationally representative samples to

help verify associations found in small-scale medical studies.

The vast majority of biomarkers currently collected and analyzed are non-genetic: anthropometric measurements (height, weight, waist circumference, lung capacity, grip strength, balance), and blood and saliva samples. The scientific value of collecting such biomarkers in large surveys is promising: First, biomarkers improve the measurement of health. Self-reports of health are subject to considerable under-, over-, or misreporting, depending on the circumstances and dimensions at hand. Objective information can be used to validate respondents' reports and to study the amount and determinants of under-, over-, or misreporting in population surveys. However, self-reports of health have their own distinct scientific value. Thus, biomarkers should be seen as complementary measurements rather than substitutes. Second, biomarkers allow studying physiological pathways in the complex relationship between social status and health, providing information on important links that can be used to identify causal relationships. Third, biomarkers provide direct information on pre-disease pathways, in particular by measuring physiological processes that are below the individual's threshold of perception. Combined with longitudinal data on individuals, this information helps to identify the role of the environment in turning health risks into manifest diseases.

Several constraints have been identified for the collection of biomarkers in social surveys. First, collecting biomarkers increases the cost and complexity of data collection. Additional costs, like those associated with visits by trained nurses (practiced by ELSA, for example), may seem prohibitive. Recent developments in minimally invasive methods which allow collection by trained survey interviewers have led to considerable cost decreases. For instance, it is now possible to measure HbA1c (as a measure for diabetes), cholesterol (to measure risk of cardiovascular disease) or C-reactive protein (to measure risk of cardiovascular disease and chronic stress) in dried blood spots. Thus the most important risk factors for chronic disease and work disability in early old age can be measured through the collection of a few drops of blood taken from the finger tips. Together with measuring blood pressure and tobacco consumption, these data allow forecasting for the incidence of cardiovascular disease 10 years hence. Thanks to advances in technology, the costs of collecting blood and analyzing the samples in laboratories are now down to a few Euro per respondent.

Still, while the research potential of collected biomarkers is large, the training of lay interviewers and the logistics of storing and sending specimens should not be underestimated. Second, collecting biomarkers increases respondent burden and may affect the willingness of

survey participants to cooperate in future waves. Third, biomarkers are potentially sensitive information and raise a lot of ethical issues surrounding confidentiality, storage, and respondent information. Survey researchers planning to include biomarkers in their data collection efforts need to be aware of these constraints.

3.3 Coverage of nursing home residents

In the U.S., some 15 percent of individuals aged 85 or older live in nursing homes. Thirty percent of all individuals die in nursing homes. However, the social and health determinants of nursing home admissions and the living conditions and quality of life of nursing home residents is greatly under-researched. The main problem faced by survey researchers probably is that in many countries, including Germany, no sample frames exist that include reliable information on nursing home residents (or other institutionalized populations). The typical approach taken so far is to draw a baseline sample from the non-institutionalized population and to follow respondents who move into nursing homes between waves. In principle, this approach should lead to samples of nursing home residents of reasonable size if panel surveys mature. In practice, however, there is substantial under-coverage due to the problems involved with tracking respondents, gaining access, and also due to a lack of respondents' ability to answer (and an increased need for proxy respondents). The oldest-old are the fastest growing segment of the population, and dementia - already a leading cause of nursing home admissions - is likely to be an increasing concern among the oldest-old. Despite the challenges of collecting data on them, neglecting a significant proportion of the older population in social surveys is hard to justify. Recent experience, for example from the Danish Longitudinal Centenarian Study, shows that many concerns voiced about conducting interviews in nursing homes (e.g., unethical, too costly) are unfounded. The primary impediment in most countries is the lack of a suitable sample frame such as a nationwide person register.

4. Conclusions

The collection of multidisciplinary, longitudinal data on aging is one of today's most active and innovative fields in survey data collection. This report has documented the host of data available to researchers from various disciplines working in the field of individual and population aging. We can observe several exciting developments. First, researchers across different countries (including emerging economies) are trying to collect data that are

comparable internationally. Comparability is sought mainly with U.S. surveys, because the U.S. not only plays a leading role in survey methodology, but also because it is an important reference country, due to its size and due to the particularities of its welfare state. Second, data access for secondary analysis has become easy and quick, thanks to developments in information technology. With few exceptions, data are released after some cleaning often less than one year after they were collected. For many surveys, released data are downloadable from websites literally within minutes. Thus the concepts of primary and secondary analysis become meaningless. Rather, it makes more sense to speak of overlapping groups of data producers and data users. Third, researchers are currently trying hard to bridge boundaries between disciplines, especially between social sciences and medicine. Some biomarkers are already routinely included in a number of ongoing surveys, and the scope of measures that can be collected during normal face-to-face interviews is increasing due to technical progress.

5. Appendix: Summary Information on Current Aging Surveys

Survey	Country	Age range	Sample size	Survey Years
SOEP	Germany	17+	22,000	1985-2007
DEAS	Germany	40-85	6000	1996, 2002, 2008
SAVE	Germany	18+	3,000	2001-2007
CNEF	USA, Germany, UK, Australia, Switzerland, Canada	17+	181,000	1980-2007
SHARE	Denmark, Sweden, Austria, France, Germany, Switzerland, Belgium, Netherlands, Spain, Italy, Greece, Poland, Czechia, Israel, Ireland	50+	35,000	2004-2008 (biannual)
GGG	Bulgaria, France, Georgia, Germany, Hungary, Russia	18-79	60,000	2005, 2008
ELSA	England	50+	12,000	2002-2008 (biannual)
SHIW	Italy	18+	20,000	1977-2006
TILDA	Ireland	55+	10,000	2008
HRS	USA	50+	22,000	1992-2008 (biannual)
MHAS	Mexico	50+	15,000	2001,2003
KLoSA	Korea	45+	10,000	2006,2008
JSTAR	Japan	50-75	4,300	2007,2009

Note: Sample sizes may vary from year to year

Access
SOEP data are available for all academics users from DIW upon signature of a user contract.
Access to the first two waves of DEAS data is provided via the central archive for social science data (ZA) at the University of Cologne (Study Numbers. 3264, 4304), access class C: data access is granted to academic users upon approval of the primary researchers.
SAVE data are available through the central archive for social science data (ZA) in Cologne (Study-Numbers: 4051, 4436, 4437, 4521, 4740), access class C: data access is granted to academic users upon approval of the primary researchers.
The PSID-CNEF file is public use and can be simply downloaded from the CNEF website. Access to BHPS-CNEF, SOEP-CNEF, or HILDA-CNEF requires approval by the BHPS, SOEP, and HILDA primary researchers, respectively. SLID data (a non-research driven survey conducted by Statistics Canada) can only be accessed via remote computing
SHARE data are available online to academic users via the SHARE website (upon signature of a data confidentiality statement) or through the central archive for social science data (Zentralarchiv) in Cologne (Study-Number: 4560), access class C: data access is granted to academic users upon approval of the primary researchers.
Data access is granted only after a research proposal submitted to the data administrators has received a positive review "for relevance to research". Who the reviewers are and what criteria for relevance they use is as yet unclear.
Online access to scientific use files for waves 0 (i.e. HSE data) through 3 (i.e. the 2006 data collection) is available via the UK Economic and Social Data Service (Study-Number 5050).
Microdata are freely available from the Banca D'Italia website
Data are not yet available
Original HRS data are available to researchers after a simple online registration process. Additionally, a user friendly combined and harmonized HRS file is made available by RAND.
Data collected so far are freely available from the MHAS website after a simple registration process.
The 2006 data and English documentation are freely available from the KLoSA website after a simple registration process.
Public release of wave 1 data planned for 2009.