



Deliverable Number: D1.1

Deliverable Title: Data Management Plan

Work Package: WP1

Deliverable type: DEC

Dissemination status: PU

Submitted by: Axel Börsch-Supan

Authors: Axel Börsch-Supan

Date Submitted: 30 June 2023

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under project ID 101093849.





[www.mpisoc.mpg.de/max-planck-emeritus-gruppe/bb-future](http://www.mpisoc.mpg.de/max-planck-emeritus-gruppe/bb-future)

In the mid2030s, the health of the baby boomers will have deteriorated and many in these large cohorts will be in need of formal and/or informal long-term care.

This “**care wave**” will transform two generations: the baby boomers in need of care and their children who may supply care. It will have significant implications for labour supply, especially for women, saving behaviour, and therefore for productivity, economic growth and its inclusiveness.

**The overarching objective of BB-Future is to understand the size and the implications of the care wave on economic and social outcomes, to appreciate the quality of this second ageing-related transformation and to develop policy recommendations for advance planning on the EU and Member State levels.**

This deliverable is the Data Management Plan of the BB-Future project.

Please cite this deliverable as: Deliverable 1.1 of the BB-Future project funded under the European Union’s Horizon 2020 research and innovation programme GA No: 101093849.

Available at: [www.mpisoc.mpg.de/max-planck-emeritus-gruppe/bb-future](http://www.mpisoc.mpg.de/max-planck-emeritus-gruppe/bb-future)

## 1. Data Summary

The project will not collect new data but use existing data that is already publicly available. Specifically, the empirical work of the project will use micro data from the **Survey of Health, Ageing and Retirement in Europe (SHARE)** and macro data from the **Social Policy Archive for SHARE (SPLASH)**. The Principal Investigator of this project is also the Managing Director of SHARE-ERIC.

*Step #1: We will use the SHARE data to describe the essential facts* that give the background of our analyses. This includes giving more detail, especially more country-specifics, on the five trends and the stylized facts that we have mentioned in the introduction.

*Step #2: We will fit the micro model to the SHARE data.* The main novelty here is to exploit the richness of the international data, the life histories provided, and the detail by which SHARE measures family relationships. The SHARE data also exhibits the international variation in the five secular trends described above. Moreover, the SPLASH data associated with SHARE documents the different policy institutions such as the pension, healthcare and long-term care systems, labour market regulations, taxation of earnings, asset income and wealth, and income support systems in all EU Member States. All this makes sure that the model has sufficient realism to be useful for policy recommendations. The quasi-experimental international variation also helps towards a clean causal interpretation of our estimates.

*Step #3: We will turn the fitted structural equations into an internally consistent projection model* that will give quantitative answers to the questions posed above and identify optimal policies conditional on assumptions about future trends, taking behavioural reactions to the policies into account, thereby addressing the Lucas critique. The projections will be based on a population weighted set of household types defined by the initial family situation, especially the education of the parents and the number and education of the siblings.

### **Description of the SHARE data:**

SHARE covers 28 European countries including Israel, about 560.000 interviews between 2004 and 2023 from 140.000 respondents. Variables include:

- health variables including self-reported health, physical functioning, physical measurements such as grip strength, walking speed, peak expiratory flow, chair stand, biomarkers extracted from capillary blood, and body mass index (BMI); health behaviours; and use of health care facilities.
- psychological variables including mental health components such as depression, several tests of cognitive functioning, well-being, life satisfaction, and control beliefs;
- socio-economic variables including current work activity, job characteristics, job flexibility, opportunities to work past retirement age, employment history, pension rights, sources and composition of current income, wealth and consumption, housing, and education; in some countries linkage to administrative data on employment, labour income and pension claims;
- social support variables including assistance within and beyond families, transfers of income and assets, social networks including ego-centered network size and intensity, and volunteer activities.

Access to the SHARE data is provided free of charge to all scientists globally. In accordance with the philosophy of sharing the data as soon as possible with the entire scientific community, a release policy has been adopted that gives free, quick and convenient access to all scientific users world-wide. All project participants are registered as SHARE users. SHARE has more than 17,000 registered users all over the world.

## **Description of the Social Policy Archive for SHARE (SPLASH):**

This is a collection of publicly available macro data such as the structure, generosity and financial status of public and private long-term care insurance systems in the EU; demographic data on the EU Member Countries; aggregate data of the health, economic and social condition in each of the EU Member States. These data describe the institutional environment in which the SHARE respondents make their decisions.

## **2. FAIR data**

The project will **not** generate new data sets.

The SHARE data themselves are accessible free of charge to the entire research community, strictly following the FAIR principles.

## **3. Other research outputs**

The project will generate tables, figures and projections based on the SHARE data and the models that have been fitted to the SHARE data. Model projections will be trajectories of outcome variables (such as age and gender-specific labor supply, savings, productivity, caregiving, unmet care needs). These trajectories will be published as tables or figures. All tables and figures will be published first as working papers and then submitted to scientific journals. They will also be presented in a less technical format, e.g., as illustrations top the policy briefs.

We will systematically share the models and results as early and widely as possible in the process. Our open science practices include early and open sharing of research through easily accessible reports on the BB-Future website and pre-prints (discussion and working papers) in the series published by our institutions. The models' equations and the SHARE variables used will be documented to ensure the reproducibility of our research outputs. It is the Max Planck Society's policy to provide open access to all our research outputs (such as publications, data, software, models, algorithms, and workflows). We will welcome participation in open peer-review.

## **4. Allocation of resources**

As stated in section 2, the project will not generate new data sets and thus no resources are allocated to the generation of new data sets.

## **5. Data security**

The SHARE data will be stored according to the conditions under which the SHARE data is distributed to all SHARE users.

## **6. Ethics**

Predictive modelling and using electronic algorithms to forecast can raise ethical questions. For example, possible biases in the data may affect the integrity or reliability of the analysis. Bias can cause distorted results and incorrect conclusions. Hence, we base our work on the SHARE data, which is representative of the population, aged 50 and older. If we restrict analyses to subpopulations, we will construct weights based on probability theory.

Moreover, an Ethics Mentor has been appointed to oversee ethical considerations such as an ethical assessment of modelling; an ethical assessment of project outcomes; performance monitoring; regular

audit of sample predictions or reviewing downstream impacts of errors; ethical data description, exploration, and verification; bias assessment; and societal impact assessment.

The project will make sure that model evidence is assessed objectively to avoid ethical mistakes, which may undermine the success of public policies and regulatory decisions and erode public trust in science and governments.

## **7. Other issues**

None.