

Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782

Start Date of Project: 01/01/2019

Duration: 40 months

Deliverable 5.4 User friendly test data release

Of one biomedical data set linked to survey data and including metadata (Access to biomedical data), including user guide

Dissemination Level	PU
Due Date of Deliverable	28/02/22, M38
Actual Submission Date	05/04/22, M40
Work Package	WP 5 - Innovations in Data Access
Task	Task 5.1 Legal, ethical and technological issues of access to biomedical data
Type	Report
Approval Status	Approved by EC - 27 April 2022
Version	V1.0
Number of Pages	p.1 – p.12

Abstract: To facilitate research with the biomedical data that was collected, SHARE aims to release user friendly datasets including a comprehensive documentation. In the case of the accelerometer study, all necessary steps towards this aim (data processing, structure and content of the data, and documentation) were tested before the first release and further improved with release updates.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability.



SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020); H2020-INFRAEOSC-04-2018, under the agreement No. 823782

History

Version	Date	Reason	Revised by
0.0	20/11/2021	First draft	Fabio Franzese
0.1	10/01/2022	Revised draft	Fabio Franzese
0.2	02/02/2022	Revision	Kalina Georgieva
0.3	03/02/2022	Revision by CentERdata (Partner in Task 5.1)	Maurice Martens
1.0	16/02/2022	Final edits	Fabio Franzese

Author List

Organisation	Name	Contact Information
SHARE ERIC	Fabio Franzese	franzese@mea.mpisoc.mpg.de

Executive Summary

Collecting biomedical data in a social science survey is an enormous challenge due to ethical and legal issues that need to be considered. The ultimate goal is the release of the data to enable the research community to fully exploit the potential of these data. SSHOC Task 5.1 is dedicated to the release of biomedical data collected in SHARE wave 6 (DBSS) and wave 8 (accelerometer data).

In the accelerometer study conducted in SHARE wave 8, a sub-sample of respondents were asked to wear an accelerometer – i.e. a sensor that captures acceleration – on their thigh for eight days. This procedure is a way to obtain measures of physical activity that are more reliable and comparable than self-reports of physical behaviour. However, it comes with higher complexity in terms of data preparation.

This report describes the releases of the accelerometer data that were generated and provided to researchers since the termination of the data collection in spring 2020. First, as described in MS23, an internal release – called “Release 0” – was generated and distributed within SHARE team members in October 2020. In June 2021 the first scientific release (version 1.0.0) of “generated modules” were made available to the research community. An update of this data (version 8.0.0) – including additional files that provide more detailed information on the accelerometer measurements – is online since February 2022.

The processing and release of the data are based on the data access protocol and the ethics guidelines that were elaborated in Task 5.1 at earlier stage (D5.2 & D5.3). Each release version was revised, improved, and extended based on feedback from users as well as our own experience in generating and using the data. A major extension are the additional datasets included in release 8.0.0 that provide acceleration measurements for 5-second intervals. The SHARE Wave 8 Methodology (Bergmann and Börsch-Supan 2021) that was released together with release 8.0.0 includes detailed documentation on the implementation of the study. For the future, it is planned to further extend the existing data with additional metrics that can be generated by algorithms (existing ones and those that may be developed in the future) based on the existing raw accelerometer data.

Abbreviations and Acronyms

DBSS	Dried blood spot sample(s)
EOSC	European Open Science Cloud
SHARE	Survey of Health, Ageing and Retirement in Europe
SSH	Social Sciences and Humanities

Table of Contents

1. Introduction.....	6
2. SHARE accelerometer study.....	6
3. Data preparation.....	7
4. Data releases.....	7
4.1 Internal test release: Release 0	8
4.2 First scientific release: Release 1.0.0.....	8
4.3 Release update: Release 8.0.0.....	9
4.4 Access to data.....	9
5. Conclusion.....	9
6. References	9

1. Introduction

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a longitudinal survey of the population aged 50+ in 28 European countries and Israel (Börsch-Supan et al. 2013). The SHARE questionnaire is designed to enable interdisciplinary studies and therefore comprises manifold questions on the social, financial and health related situation of respondents' life. From its beginning, the questionnaire was accompanied by different objective measures of health such as grip strength, peak expiratory flow, and walking speed. According to this strategy of collecting objective health data, SHARE has collected dried blood spot samples (DBSS) in 2015 in twelve countries (Börsch-Supan et al. 2020). In 2019, ten countries conducted a study to measure physical activity by means of thigh-worn accelerometers.

Aim of SSHOC Task 5.1 is to elaborate procedures to collect, process, and release biomedical data according to the FAIR principles. In SHARE these procedures were established for both, DBSS and accelerometry data, and resulted in D5.1 "Guidelines for ethics considerations in making biomedical survey data FAIR" (Börsch-Supan et al. 2021) as well as D5.2 and D5.3 (data access plans). This report presents the final result of all these efforts for one of the data types collected in SHARE: the release data and associated documentation of the SHARE accelerometer study.

2. SHARE accelerometer study

Physical (in)activity is a crucial health behaviour as acknowledged in the physical activity guidelines by the World Health Organization (Bull et al. 2020). Little activity is related to the onset of mental (Cocker et al. 2021) and physical diseases (Liu et al. 2017) as well to higher mortality (Ekelund et al. 2020). This is why it is important to measure physical activity in surveys on health. However, comparison of self-reported – by questionnaire or diary – and device-measured data revealed that people tend to overestimate their own activity and underestimate their inactivity (Hagstromer et al. 2010; Prince et al. 2020).

In SHARE wave 8 accelerometers were used to collect physical activity data in a subset of 10 countries: Denmark, Sweden, Germany, France, Belgium, Spain, Italy, Czech Republic, Poland, and Slovenia. Preselected respondents were asked for participation during the regular SHARE face-to-face interview. Consenting respondents received the accelerometer together with all necessary materials and instructions via mail. Participants attached the devices themselves on their upper thigh and were asked to wear the device for eight consecutive days, if possible without breaks. The used Axivity AX3 accelerometer (Axivity Ltd, Newcastle upon Tyne, United Kingdom) which is a small and lightweight device. It is waterproof, meaning respondents could wear it while swimming or showering, but had to take it off for having a sauna or scuba diving. After wear-time participants sent the device back to the survey agency who downloaded the data and provided it to SHARE Central.

The fieldwork of SHARE wave 8 started in October 2019; the first respondents received the accelerometers in November 2019. Fieldwork was suspended in March 2020 due to the Covid-19 pandemic. By that point, 856 respondents had provided valid accelerometer measurements. More information on the implementation of the SHARE accelerometer study can be found in the SHARE Wave 8 Methodology (Bergmann and Börsch-Supan 2021).

3. Data preparation

Raw accelerometer sensor data were recorded with 50Hz, i.e. 50 measurements of three axis per second, which result in large data files that lead to some challenges. First, the data transfer during fieldwork between the survey agencies and SHARE Central had to be organised in an efficient, easy, and secure way. Therefore the DeviceCTRL online tool was provided by CentERdata as described in MS22 (Franzese 2020). DeviceCTRL was used as a database for the raw data files as well as for other information of the measurements, such as dates of shipping, start and end of wear time, breaks in wear time, and problems noted by the participants. Second, as raw accelerometer data are not easy to handle and their processing is not common knowledge in SSH, the large amount of data should be aggregated to provide easy to understand measures for SHARE users.

(Meta) data from DeviceCTRL, e.g. start, end and breaks of wear time, is matched with actual accelerometer measurements to adequately process the sensor data. Processing of the raw data is performed with GGIR (Migueles et al. 2019), an open source package for the statistical software R (R Core Team 2021). GGIR allows to carry out non-wear detection and imputation of non-wear time as well as auto-calibration to adjust for differences in local gravity and device specific deviations (van Hees et al. 2014). High frequency data of three axes are combined to the ENMO (Euclidean norm minus one), measure of total acceleration (van Hees et al. 2013), for periods of five seconds (so-called epochs). Based on the ENMO on epoch level, several metrics can be generated, such as the average ENMO over different periods (e.g. hour, day, and week) and intensity distribution. Accelerometer measurements are considered as valid with at least 16 hours of wear time per day (determined by GGIR non-wear detection). Further processing, necessary to implement the generated metrics into the general SHARE procedures, is performed using Stata (StataCorp LP, College Station, USA).

4. Data releases

Additional to the survey data, SHARE release data include “generated modules” that provide processed, easy to use variables based on the raw data (e.g. indices on health, coded education information, etc.). The metrics based on accelerometer measurements describing the physical activity of respondents are provided in such generated modules, too. The generated modules for the accelerometer data can be extended in the future with additional metrics.

Internal test release: Release 0

Before SHARE data are released publicly, an internal release is provided to all members of the SHARE team as documented in MS23. It is intended to have “test users” having checked the data and documentation to resolve possible bugs. The so-called “Release 0” was circulated internally in October 2020. It comprised one generated module based on GGIR processed accelerometer data. The dataset contained one line per observation day, i.e. several lines per respondent. The data file was accompanied by a document with a description of the accelerometer study, the metrics, and the structure of the dataset.

First scientific release: Release 1.0.0

The first scientific release of the generated modules based on the accelerometry measurements was part of Release 1.0.0 of SHARE wave 8 (Börsch-Supan 2021) that is available since June 2021.

For the accelerometer data three generated modules are available in Release 1.0.0 (compared to one in Release 0). For the convenience of users, data were split into “hour”, “day”, and “total” datasets. The “hour” data provide average ENMO and time spent in ENMO levels on an hourly basis. This allows users to analyse the level of physical activity over the whole course of the day. The “day” dataset gives information on a daily level – as it was in Release 0. The “total” dataset provides information on the whole observation period per respondent – for most respondents one week, some respondents have less days with sufficient valid wear time.

Together with the metrics describing physical activity, the generated modules provide additional (meta) data that is of interest when analysing physical activity. The month of the measurement might be relevant due to seasonal effects (Turrisi et al. 2021). The weekday might be of importance as people often have different routines on weekends compared to working days, especially for the working population. As some of the measurements were collected at the onset of the Covid-19 pandemic, the datasets also include indicators for pandemic related restrictions that might influence physical behaviour, such as stay-at-home orders.

The data processing and the metrics provided in the generated module are described in section 15.17 of the “Release Guide 1.0.0 of Wave8”, available on the SHARE webpage¹.

¹ SHARE Release Guide Wave8: http://www.share-project.org/fileadmin/pdf_documentation/SHARE_release_guide_1-0-0-w8.pdf; accessed 07.01.2022

Release update: Release 8.0.0

Since February 2022 release 8.0.0 (Börsch-Supan 2022) is available via the *SHARE Research Data Center*². Release 8.0.0 comprises an update of the generated accelerometer modules. Minor revisions include minor changes to metrics, as an updated version of GGIR was used as well as some new variables in the dataset on day-level, namely indicators for most and least active 8 and 16 hours of the day.

The major advancement is the detailed information that is now provided in the “epoch” datasets. For each respondent a dataset indicating physical activity in 5-second intervals (so-called “epochs”) over the whole observation time is available. Acceleration data on epoch level enable to look at activity profiles in detail, e.g. for diurnal activity patterns (cf. Xiao et al. 2015). Like the other information in the generated modules, the epoch-level data is generated with GGIR. The documentation in the release guide is extended accordingly with a description of epoch data. The SHARE Wave 8 Methodology that was released together with release 8.0.0 adds detailed documentation on the implementation of the study (Scherpenzeel et al. 2021).

Access to data

All scientific releases of SHARE data are available through the *SHARE Research Data Center* for registered SHARE users. Users are required to register individually³ and agree with the conditions of use⁴; SHARE data is available for scientific purpose only. The data access protocol and its compliance to the FAIR principles is described in SSHOC D5.3 (Franzese 2021).

5. Conclusion

The measurement of physical activity with accelerometers is more reliable and comparable than self-reports of physical behaviour, which is why this method is used in more and more studies. Since the data processing is elaborate, it is crucial that data providers offer ready-to-use variables that can be easily picked up by researches of different disciplines.

The data collection in wave 8 was the first accelerometer measurement in SHARE. Therefore, it was crucial to prepare the scientific release as comprehensively as possible by having a test round as describes in

² SHARE data access: <http://www.share-project.org/data-access.html>; accessed 07.01.2022

³ All details are available at User Registration: <http://www.share-project.org/data-access/user-registration.html>; accessed 07.01.2022

⁴ SHARE Conditions of Use: <http://www.share-project.org/data-access/share-conditions-of-use.html>; accessed 07.01.2022

MS23. The first scientific release was successfully accomplished only one year after completion of the collection of the accelerometer data. As the raw sensor data can be processed in different ways – different software that aim at different aspects of physical activity – and this field is emerging fast, the idea is to constantly improve and extend the generated modules in the SHARE releases with new metrics. Using the existing raw data with up-to-date analysis software – that will be available sometime in the future – enables SHARE to provide users with new possibilities for their research. The first advancement of the SHARE accelerometer data, epoch-level information, is already available in release 8.0.0.

References

- Bergmann, Michael, and Axel Börsch-Supan, editors. 2021. *SHARE Wave 8 Methodology: Collecting Cross-National Survey Data in Times of COVID-19*. Munich: MEA, Munich Center for the Economics of Aging.
- Börsch-Supan, Axel, Martina Brandt, Christian Hunkler, Thorsten Kneip, Julie Korbmacher, Frederic Malter, Barbara Schaan, Stephanie Stuck, and Sabrina Zuber. 2013. "Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)." *International journal of epidemiology* 42(4):992–1001. doi:10.1093/ije/dyt088.
- Börsch-Supan, Axel, Brändle, Carolina, Bristle, Johanna, Fabio Franzese, Daniel Schmidutz, and Luzia Weiss. 2021. *D5.1 Guidelines for ethics considerations in making biomedical survey data FAIR (Access to biomedical data)*: Zenodo. doi:10.5281/zenodo.4785198.
- Börsch-Supan, Axel. 2021. *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 8. Release version: 1.0.0. SHARE-ERIC. Data set*. doi:10.6103/SHARE.w8.100.
- Börsch-Supan, Axel. 2022. *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 8. Release version: 8.0.0. SHARE-ERIC. Data set*. doi:10.6103/SHARE.w8.800.
- Börsch-Supan, Martina, Luzia Weiss, Karen Andersen-Ranberg, and Axel Börsch-Supan. 2020. "Collection of Dried Blood Spots in the Survey of Health, Ageing and Retirement in Europe (SHARE): From implementation to blood-marker analyses." *SHARE Working Paper Series* 47-2020. doi:10.17617/2.3245285.
- Bull, Fiona C., Salih S. Al-Ansari, Stuart Biddle, Katja Borodulin, Matthew P. Buman, Greet Cardon, Catherine Carty, Jean-Philippe Chaput, Sebastien Chastin, Roger Chou, Paddy C. Dempsey, Loretta DiPietro, Ulf Ekelund, Joseph Firth, Christine M. Friedenreich, Leandro Garcia, Muthoni Gichu, Russell Jago, Peter T. Katzmarzyk, Estelle Lambert, Michael Leitzmann, Karen Milton, Francisco B. Ortega, Chathuranga Ranasinghe, Emmanouel Stamatakis, Anne Tiedemann, Richard P. Troiano, Hidde P. van der Ploeg, Vicky Wari, and Juana F. Willumsen. 2020. "World Health Organization 2020 Guidelines on Physical Activity and Sedentary Behaviour." *British journal of sports medicine* 54(24):1451–62. doi:10.1136/bjsports-2020-102955.

- Cocker, Katrien de, Stuart J. H. Biddle, Megan J. Teychenne, and Jason A. Bennie. 2021. "Is All Activity Equal? Associations Between Different Domains of Physical Activity and Depressive Symptom Severity Among 261,121 European Adults." *Depression and anxiety* 38(9):950–60. doi:10.1002/da.23157.
- Ekelund, Ulf, Jakob Tarp, Morten W. Fagerland, Jostein S. Johannessen, Bjørge H. Hansen, Barbara J. Jefferis, Peter H. Whincup, Keith M. Diaz, Steven Hooker, Virginia J. Howard, Ariel Chernofsky, Martin G. Larson, Nicole Spartano, Ramachandran S. Vasan, Ing-Mari Dohrn, Maria Hagströmer, Charlotte Edwardson, Thomas Yates, Eric J. Shiroma, Paddy Dempsey, Katrien Wijndaele, Sigmund A. Anderssen, and I-Min Lee. 2020. "Joint Associations of Accelerometer-Measured Physical Activity and Sedentary Time with All-Cause Mortality: A Harmonised Meta-Analysis in More Than 44 000 Middle-Aged and Older Individuals." *British journal of sports medicine* 54(24):1499–506. doi:10.1136/bjsports-2020-103270.
- Franzese, Fabio. 2020. *MS22 Inventory of computing space needed for processing and analysing accelerometer data (v1.1)*: Zenodo. doi:10.5281/zenodo.4696220.
- Franzese, Fabio. 2021. *D5.3 Data access protocol for accelerometer data, linked to survey data, conforming FAIR principles (Access to biomedical data)*: Zenodo. doi:10.5281/zenodo.5608517.
- Hagstromer, Maria, Barbara E. Ainsworth, Pekka Oja, and Michael Sjostrom. 2010. "Comparison of a Subjective and an Objective Measure of Physical Activity in a Population Sample." *Journal of physical activity & health* 7(4):541–50. doi:10.1123/jpah.7.4.541.
- Liu, Xuejiao, Dongdong Zhang, Yu Liu, Xizhuo Sun, Chengyi Han, Bingyuan Wang, Yongcheng Ren, Junmei Zhou, Yang Zhao, Yuanyuan Shi, Dongsheng Hu, and Ming Zhang. 2017. "Dose-Response Association Between Physical Activity and Incident Hypertension: A Systematic Review and Meta-Analysis of Cohort Studies." *Hypertension (Dallas, Tex. : 1979)* 69(5):813–20. doi:10.1161/HYPERTENSIONAHA.116.08994.
- Migueles, Jairo H., Alex V. Rowlands, Florian Huber, Séverine Sabia, and Vincent T. van Hees. 2019. "GGIR: A Research Community-Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data." *Journal for the Measurement of Physical Behaviour* 2(3):188–96. doi:10.1123/jmpb.2018-0063.
- Prince, Stephanie A., Luca Cardilli, Jennifer L. Reed, Travis J. Saunders, Chris Kite, Kevin Douillette, Karine Fournier, and John P. Buckley. 2020. "A Comparison of Self-Reported and Device Measured Sedentary Behaviour in Adults: A Systematic Review and Meta-Analysis." *The international journal of behavioral nutrition and physical activity* 17(1):31. doi:10.1186/s12966-020-00938-3.
- R Core Team. 2021. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (<https://www.R-project.org/>).
- Scherpenzeel, Annette, Nora Angleys, Fabio Franzese, and Luzia Weiss. 2021. "Measuring Physical Activity in SHARE: The SHARE Accelerometer Study." Pp. 183–93, in *SHARE Wave 8 Methodology: Collecting Cross-National Survey Data in Times of COVID-19*, edited by M. Bergmann and A. Börsch-Supan. Munich: MEA, Munich Center for the Economics of Aging.

- Turrisi, Taylor B., Kelsey M. Bittel, Ashley B. West, Sarah Hojjatinia, Sahar Hojjatinia, Scherezade K. Mama, Constantino M. Lagoa, and David E. Conroy. 2021. "Seasons, Weather, and Device-Measured Movement Behaviors: A Scoping Review from 2006 to 2020." *The international journal of behavioral nutrition and physical activity* 18(1):24. doi:10.1186/s12966-021-01091-1.
- van Hees, Vincent T., Lukas Gorzelniak, Emmanuel C. Dean León, Martin Eder, Marcelo Pias, Salman Taherian, Ulf Ekelund, Frida Renström, Paul W. Franks, Alexander Horsch, and Søren Brage. 2013. "Separating Movement and Gravity Components in an Acceleration Signal and Implications for the Assessment of Human Daily Physical Activity." *PloS one* 8(4)e61691. doi:10.1371/journal.pone.0061691.
- van Hees, Vincent T., Zhou Fang, Joss Langford, Felix Assah, Anwar Mohammad, Inacio C. M. da Silva, Michael I. Trenell, Tom White, Nicholas J. Wareham, and Søren Brage. 2014. "Autocalibration of Accelerometer Data for Free-Living Physical Activity Assessment Using Local Gravity and Temperature: An Evaluation on Four Continents." *Journal of applied physiology (Bethesda, Md. : 1985)* 117(7):738–44. doi:10.1152/jappphysiol.00421.2014.
- Xiao, Luo, Lei Huang, Jennifer A. Schrack, Luigi Ferrucci, Vadim Zipunnikov, and Ciprian M. Crainiceanu. 2015. "Quantifying the Lifetime Circadian Rhythm of Physical Activity: A Covariate-Dependent Functional Approach." *Biostatistics (Oxford, England)* 16(2):352–67. doi:10.1093/biostatistics/kxu045.