
SHARE WORKING PAPER SERIES

Participation and Selection Bias in the SHARE Accelerometer Study

Mick P. Couper and Annette Jäckle

Working Paper Series 92-2024

DOI: [10.6103/wp.922024](https://doi.org/10.6103/wp.922024)

SHARE-ERIC | Leopoldstr. 139 | 80804 Munich | Germany | share-eric.eu

About the SHARE Working Paper Series

The series is designed to provide a timely discussion of results based on SHARE data within the SHARE family, i.e., members of the SHARE Country Teams, Area Coordination Teams, and other SHARE bodies. The papers are not peer reviewed; the authors are solely responsible for the scientific content and the graphical layout of their submissions. The respective Country Team Leaders and Area Coordinators are encouraged to look over the submissions by their team members.

The publisher (SHARE ERIC) checks working papers in this series for formal issues such as proper acknowledgements to the funders of SHARE. The publisher takes no responsibility for the scientific content of the paper.

Acknowledgements

The SHARE data collection has been funded by the European Commission, DG RTD through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982, DASISH: GA N°283646) and Horizon 2020 (SHARE-DEV3: GA N°676536, SHARE-COHESION: GA N°870628, SERISS: GA N°654221, SSHOC: GA N°823782, SHARE-COVID19: GA N°101015924) and by DG Employment, Social Affairs & Inclusion through VS 2015/0195, VS 2016/0135, VS 2018/0285, VS 2019/0332, and VS 2020/0313. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C, RAG052527A) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

Participation and Selection Bias in the SHARE Accelerometer Study

Mick P. Couper (University of Michigan) and Annette Jäckle (University of Essex)

Abstract

In Wave 8 of the Survey of Health and Retirement in Europe (SHARE), conducted in early 2020, accelerometers were used in a subset of ten countries to collect physical activity data. Following consent, obtained during the interviewer-administered survey, a subsample of consenters was sent Axivity AX3 accelerometers by post, and asked to wear the device for eight consecutive days attached to their upper thigh. In this secondary analysis, we examine various stages of selection into the accelerometer study, including 1) consent, 2) being shipped a device, 3) using the device at least once, and 4) using the device for at least 8 days (fully adherent). Our focus is on the sample loss at each of these stages, the correlates of (non-)participation at each stage, and the extent to which such loss is non-random (i.e., introducing selection or participation biases). We find evidence of substantial sample loss, especially at the consent stage. However, the resulting selection biases are not large, in terms of either socio-demographic or health-related substantive variables.

Acknowledgements

We thank SHARE for making these data available for secondary analysis. We are especially grateful to Fabio Franzese for answering our queries about the study and the data. This paper is partial replication and extension of Franzese, Schrank, and Bergmann (2023).

Participation and Selection Bias in the SHARE Accelerometer Study

Introduction

As technological changes enable new methods of data collection, survey researchers are trying to find ways to gather additional information through a variety of survey enhancements. The goal of these enhancements is to extend the breadth and depth of measurement without adding unduly to respondent burden. This often involves “passive” methods of data collection, where – after obtaining respondent consent – data are collected through sensors, often using respondents’ smartphones or devices provided by the survey organization. Examples include activity trackers (e.g., McCrorie, Walker, and Ellaway, 2018), GPS trackers (e.g., Elevelt et al., 2021; Olsen et al., 2019), air pollution monitoring devices (e.g., Kapteyn, Saw, and Weerman, 2022), and the like. While the focus is often on increasing the granularity and accuracy of measurement, the addition of these survey enhancements raises questions about potential selection biases, if some types of people are more likely to participate in such tasks than others.

This paper focuses on one such survey enhancement – the use of accelerometers to collect physical activity data. We conduct secondary analyses of an implementation in a large multi-national longitudinal survey. In Wave 8 of the Survey of Health and Retirement in Europe (SHARE), conducted from November 2019 to June 2020, accelerometers were used in a subset of ten countries to collect physical activity data. We examine several key outcomes of the data collection process, including 1) whether eligible SHARE respondents consented to receive and wear an activity tracker, 2) whether the device was shipped to the respondent, 3) whether they wore the device for at least one day (minimally adherent) and 4) whether they wore the device for eight days (fully adherent).

Our interest is in the sample loss at each of these stages, and the extent to which such loss is non-random (i.e., introducing selection biases). We are particularly interested in exploring the “healthy volunteer” hypothesis which posits that people who participate in large-scale health studies that involve activity tracking are often more healthy and more active than the general population.

We examine the following research questions:

RQ1: What are the sample losses at each stage of the process?

This analysis is largely descriptive, documenting sample loss in the SHARE accelerometry study. As noted earlier, there are several stages of participation or potential sample loss. Some of these are the result of sample person decisions (consent, device wear); others are the result of logistical or operational decisions (sampling, shipping). Both types have implications for sample loss, but the consequences for selection bias may be different. We examine the cumulative loss across all of these stages.

RQ2: What are the predictors of participation or loss at each stage of the process?

This research question focuses on the causes or correlates of non-participation or sample loss, and how these may differ across the different stages of selection. An advantage of an accelerometer study being embedded in a large ongoing panel study is the availability of a rich set of covariates on all eligible sample members. In addition to looking at socio-demographic differences, our particular interest is in health and financial wellbeing measures that are the core substantive focus of SHARE. We are also

interested in whether and how these change over the stages of selection. That is, are the factors associated with sample loss consistent across stages, or do they vary? Are the losses offsetting or compounding?

RQ3: What are the selection biases at each stage of the study?

This research question is closely related to RQ2. Whereas RQ2 looks at *causes* or *correlates* of sample loss at each successive stage, conditional on the previous stage, RQ3 explores the *cumulative* effect of sample loss relative to the original sample of eligible SHARE respondents. Here the focus is on the *consequences* of cumulative sample loss on descriptive statistics. Specifically, do we find evidence of “healthy person bias” for a variety of health and activity-related outcomes?

Background

Selection error is a concern at all stages of the survey process, and threatens the inferential value of the survey data in two key ways. First, the reduced sample sizes associated with sample loss decrease the precision of estimates, i.e., increase the variance of estimates. Second, *differential* sample loss – where some subgroups are lost at higher rates than others – may result in selection bias, where estimates are systematically changed due to the loss (see Groves, 2006, for examples of selection bias related to nonresponse). Selection error in surveys occurs in two main ways: 1) non-coverage of segments of the population of interest and 2) nonresponse to the survey (Groves et al., 2009). When survey data are enhanced through additional data collection tasks, selection error can further occur, *conditional* on participation in the survey. These may further erode the inferential value of the combined estimates from the survey and the enhancements.

Depending on the nature of the task, selection errors can take several forms. For example, when the additional task is to link survey data to administrative records, linkage consent and linkage errors can result in selection errors (e.g., Sakshaug et al., 2017; Sakshaug 2021; Yang, Fricker, and Eltinge, 2017). When respondents are asked to use their own devices (e.g., smartphones for GPS tracking), coverage error can occur if not all participants have a compatible smartphone (e.g., Horn et al., 2022; Jäckle et al., 2019; Vine et al., 2021). There may be other eligibility criteria applied before making the request of survey participants. Some survey enhancements elicit an initial indication of willingness, followed by a consent process; at other times, the willingness to participate implies consent. Willing participants are then asked to do something. In the case of accelerometers, this may involve receiving the device, setting it up, initiating the measurement and starting to wear it. For app-based data collection this may involve installing and activating the app and logging in (e.g., Jäckle et al., 2023; McCool et al., 2021). Finally, participants are expected to wear the device or use the app as instructed for the full period of time specified¹.

There is no consistent terminology in the literature to describe the various outcomes of enhancement studies. They are defined differently in each study – and sometimes not well defined. Given the wide variety of protocols (see below), it is unlikely that a common set of outcome definitions is possible. However, clearly specifying the various outcomes and the sample loss at each stage will facilitate

¹ The terms “compliance” and “adherence” are often used interchangeably. We prefer the term “adherence” for the continuing use of the device in accordance with the study protocols. We reserve the use of “compliance” for the initial installation or activation of the device.

comparisons across studies (see Antoun and Wenz, 2021; Couper, 2019). Further, some of these outcomes can be viewed as data losses at the unit (i.e., participant) level or losses at the item level (e.g., periods of non-wear, other temporary data losses). Our focus is on the former. The latter requires analysis of the accelerometer data (see Franzese, Schrank, and Bergmann, 2023), and it is often hard to attribute cause or distinguish between periods of inactivity and non-adherence.

There are thus many opportunities for sample loss in studies like this, whether due to participants themselves (e.g., non-consent, non-compliance, non-adherence) or due to factors outside of participants' control (e.g., study designs, logistical issues, technical limitations, or the like). Put differently, a number of conditions must be met to get complete data from a sensor and loss of data at each step may compromise the inferential value of the resulting data.

Review of fieldwork protocols and outcomes of accelerometry studies

In our review of the relevant literature, we focus on population-based studies using accelerometry or actigraphy. We include large-scale volunteer efforts such as the UK Biobank and the All of Us Research Program in the U.S.A. because of their size and influence in terms of population-based inference. We do not include the many studies involving small samples of volunteers. A summary of these studies is presented in Appendix A.

The advantages of focusing on participants in an existing survey are 1) that they are generally based on probability samples designed to represent the broader population from which they are drawn (in contrast to volunteer studies), and 2) that a rich set of covariates exist about everyone invited to the study, which can be used to quantify the extent of selection biases. The key disadvantage is that sample loss (through non-coverage and nonresponse to the survey) has already occurred. Meaning that our focus is on the *additional* sample loss and potential selection bias arising from participation in the accelerometer-based measurement, *conditional* on participating in the survey.

One of the earlier survey-based uses of accelerometers was the National Health and Nutrition Examination Survey (NHANES). NHANES assesses the health and nutritional status of a representative sample of about 5,000 individuals in the U.S., using both surveys and medical examinations (see <http://www.cdc.gov/nchs/nhanes.htm>). Beginning in 2003, NHANES has also assessed physical activity with participants using hip-worn (2003-6; ActiGraph AM-7164) or wrist-worn (2011-14; Actigraph GT3X+) devices for a period of 7 days. Participants were recruited during the in-person interview in the mobile examination centers and were offered incentives for using and returning the device (\$40 in 2011).

Details on the NHANES physical activity monitor (PAM) component, especially those related to sample loss at each stage of the process, are hard to find. For example, Lee (2013), Liu et al. (2016), Loprinzi et al. (2013) and Troiano et al. (2008) all focus on item- or epoch-level missing data in accelerometry conditional on unit response (providing some data). Similarly, Füzéki, Engeroff, and Banzer's (2017) systematic review of papers using NHANES PAM data in one domain (light physical activity) is largely silent on the issue of selection error or sample loss.

In one recent exception, Antoun and Wenz (2021) conducted secondary analysis of the 2011 and 2013 rounds of NHANES. Of the 7,821 participants who responded to the household interview, were examined in the Medical Examination Center, and were eligible for the accelerometer study in 2011,

6,917 (or 88%) consented and returned the device after use with some data, while 6,467 (or 83% of eligible participants) were deemed fully adherent (defined as wear time of at least 10+ hours on 4+ days of the 7 day study period). Similar rates of 87% and 80% were reported for 2013.

The Health Survey for England (HSE), which surveys a large sample of the UK population on an ongoing basis, included a large accelerometry component in 2008 (see Roth and Mindell, 2013; Hamer, Coombs, and Stamatakis, 2014). Of the 15,102 adults interviewed in 2008, a subsample of 4,507 (30%) were recruited during face-to-face interviews to wear an Actigraph GT1M on their waist during waking hours for seven consecutive days. Participants were offered a gift voucher worth £20 upon return of the device. Of those invited to wear an accelerometer, 890 (20%) declined, 874 (19%) had no accelerometer data and 237 (5%) did not have sufficient wear time (defined as at least 10 hours per day on at least 4 days). This represented a cumulative sample loss of 44% of those sampled for the study.

The SPACES study is a sub-study of the Growing up in Scotland (GUS) cohort study (see McCrorie, 2022) designed to collect physical activity from 10- to 11-year-old children across Scotland. In the eighth wave (2014-5), parents or carers (n=2,402) were provided with brief information about SPACES and asked if their contact details could be passed on to SPACES staff. A total of 2,162 parents and carers agreed and were sent an information packet by mail (with telephone follow up). Participants were asked to wear the ActiGraph GT3X+ on an elasticized waist belt, during waking hours, for eight consecutive days. Of those who agreed, 1,096 (51% of those who agreed to provide contact information; 46% of all parents and carers) completed and returned the registration form and were sent device kits. Of the 1,096 children who agreed to take part, 859 returned data and 776 provided sufficient data (70% of those who agreed to participate, 36% of those who provided contact information, and 32% of eligible GUS cohort members; McCrorie, 2017). No additional incentive was provided. McCrorie, Walker, and Ellaway (2018) described some of the logistical challenges of conducting a study like this primarily by mail.

The Millennium Cohort Study (MCS) used accelerometers in Sweep 4 (2008-9; age 7) and Sweep 6 (2015-6; age 14). In Sweep 4 (MCS4), participants were mailed ActiGraph GT1M accelerometers following consent, and asked to wear it on their waist for seven consecutive days during waking hours (excluding aquatic activities). Participants were offered a £10 voucher for returning the device. According to Griffiths et al. (2013a, 2013b), of the 14,043 persons interviewed, 13,219 consented (94%), 12,625 were sent accelerometers, 10,034 returned them (76% of consented, 72% of interviewed), 8,939 had some accelerometer data (71% of those sent accelerometers, 68% of consented, 64% of interviewed), and 6,675 (53% of those sent accelerometers, 50% of consented, and 48% of interviewed) had reliable data (defined as two or more days of wear lasting 10 or more hours per day). Ahn et al. (2018) provide slightly different numbers and use a different definition of valid wear time (not specified).

In two pilot studies for Sweep 6 of the MCS, two different devices were tested (GENEActiv Original and ActiGraph GT3X+; Gilbert et al., 2017). The GENEActiv device was selected for the main study. Participation packs were delivered in-person by interviewers or, where the device was not charged, by mail. Interviewers were on hand to assist with wear placement. Participants were asked to wear the device on their wrist for one weekday and one weekend day assigned at time of interview. No additional incentives were offered. A total of 10,337 cohort members were eligible for the study. Devices were not placed with 11% of eligible respondents (mostly due to respondent refusal). About a quarter (24%) of eligible respondents were given a device but it was not returned, 8% returned a device that was broken

or no data could be downloaded, 9% returned a device that contained insufficient data (i.e., not worn or worn for less than 10 hours on the selected days). Of eligible respondents, 48% returned a device that contained valid data (41% had valid data for both days; 7% had valid data for one day). Among respondents with whom a device was placed, 54% returned a device with valid data (see Gilbert et al., 2017).

Another large-scale survey including accelerometry data is the Survey of Health, Ageing and Retirement in Europe (SHARE). As this is the focus of our secondary analyses, further details are provided later. However, Franzese, Schrank, and Bergmann (2023) have conducted a detailed analysis of correlates of consent and examined the bias of the sample mean of physical activity due to selective non-consent. Our paper is a partial replication and extension of their work. We examine several steps in the data collection process and look at correlates of participation and potential biases at each successive stage.

A final example is the American Life in RealTime (ALiR), which is part of the Understanding America Study (UAS), a probability-based online panel of U.S. adults (see Chaturvedi et al., 2023). While recruitment and data collection is ongoing, as of April 15th, 2023, of the 2,153 UAS panelists who completed the recruitment survey online, 1,386 (64%) consented to participate in ALiR and were mailed a Fitbit Inspire 2 wrist-worn accelerometer. Points were awarded for participation in various UAS activities; those who wore the Fitbit for a year could earn up to \$126. Of those who consented, 1,038 (75% of consenters, 48% of survey respondents) have enrolled and 348 were lost to follow-up (262 did not confirm their address and could not be sent their device/enrollment materials; 42 received materials but never enrolled; and 44 withdrew) (see Chaturvedi, Angrisani, and Couper, 2023).

In addition to these studies, the English Longitudinal Study on Aging (ELSA) has initiated a large accelerometer-based study (outcomes are not yet available), and the Health and Retirement Study (HRS) in the U.S. has conducted a pilot study with a view to large-scale implementation. Understanding the causes and correlates of sample loss and the potential selection bias implications in large-scale longitudinal surveys is becoming even more imperative. Furthermore, there are a number of smaller or regional population-based studies that have used accelerometers, including Framingham (Spartano et al., 2019; Cho et al., 2022), PARCS (Cato et al., 2020), the Hispanic Community Health Study/Study of Latinos (Evanson et al., 2015), the Whitehall II study (Hassani et al., 2014), and FLASHE (Oh et al., 2021).

Sample loss may be substantial even in volunteer studies. We briefly review two examples of large-scale volunteer databanks (LSVDs; see Brayne and Moffitt, 2022). Doherty et al. (2017) and Leroux et al. (2021) report on the accelerometry study embedded in the UK Biobank. UK Biobank is a large prospective study with 500,000 participants aged 40-69 years when recruited in 2006-2010 (see Sudlow et al., 2015). Between February 2013 and December 2015, participants who had provided a valid email address were sent an email invitation to wear an Axivity AX3 wrist-worn accelerometer. From June 2013, participants were mailed devices in order of acceptance. The devices were set up to activate two working days after postal dispatch and capture data for seven days. Participants were asked to wear the device continuously for seven days then return it by mail. No incentives were offered.

As reported by Doherty et al. (2017), a total of 236,519 participants were approached, of whom 106,053 (44.8%) consented and agreed to wear a device. Data were received from 103,712 devices for analysis (87 participants subsequently withdrew, 1,316 devices were not returned, and 930 were unreadable), of which 6,978 had insufficient wear time. This left 96,600 devices (91% of consented, 41% of invitees) with

“sufficient physical activity data for analysis”. Leroux et al. (2021) report similar numbers, resulting in 96,536 participants with “good” accelerometer data.

The *All of Us* Research Program (AoURP) is an example of a bring-your-own-device (BYOD) activity tracker study (see Cho et al., 2022), in which participants with their own Fitbit devices were asked to upload their data to the AoURP portal. No incentives were offered. Master et al. (2022) reported that of the 329,070 AoURP participants at the time of their analysis, 214,206 (65%) had consented to share electronic health record (EHR) data. Of those sharing EHR data, 6,042 linked their own Fitbit device, had valid Fitbit data over at least six months of total monitoring and were aged at least 18 years. This represented 3% of those with EHR data and 2% of AoURP participants. Perry et al. (2023) reported that of the 362,674 participants with data available, 12,781 provided Fitbit data linked with electronic health records. The number of AoURP participants who owned or used Fitbits at the time is unknown.

There are a few things to note from this brief review of survey-based accelerometer studies. First, there are a large number of studies using accelerometers or activity trackers, and their use appears to be increasing. Second, details on sample dropout are relatively scarce and inconsistently reported. This points to a need for a CONSORT-like statement (see www.consort-statement.org) on sample loss in survey-based accelerometer studies. Third, there is little consideration of differential sample loss or selection bias. For example, NHANES is the most frequently-used source of population-based accelerometry data. While there are investigations of nonresponse bias in the main NHANES study (e.g., Fakhouri et al. 2020), we found no studies examining differential sample loss (with the exception of Antoun and Wenz, 2021) and potential bias in the accelerometry study. Response rates by age and gender are provided for the NHANES interview and examination components (see <https://wwwn.cdc.gov/nchs/nhanes/ResponseRates.aspx>) but not for the accelerometry component. There is clearly need for more work in this area.

Review of selection biases in accelerometry studies

Fry and colleagues (2017) coined the term “healthy volunteer bias” to refer to the tendency for volunteers who participate in large-scale health studies to be more healthy than the general population. Compared with the general population, they found that UK Biobank participants were more likely to be older, female, and live in less socioeconomically deprived areas. Participants were also less likely to be obese, smoke, and drink alcohol on a daily basis and had fewer self-reported health conditions. They concluded that “UK Biobank is not representative of the sampling population; there is evidence of a ‘healthy volunteer’ selection bias” (Fry et al., 2017, p. 1026). Klijs et al. (2015) similarly noted that “participants to cohort studies often have less chronic diseases and have a better level of functioning than those who do not participate.”

The evidence of healthy person bias may be strong for other types of screening subsamples. Blom et al. (2008) noted that “Individuals who might benefit most from [colon cancer] screening are overrepresented among nonparticipants”. Pinsky et al. (2017) similarly concluded: “Volunteers for prevention or screening trials are generally healthier and have lower mortality than the general population.” Lyall et al. (2022) looked at bias in the MRI imaging subsample of the UK Biobank. They concluded: “On a range of cognitive, mental health, cardiometabolic, inflammatory and neurological phenotypes, we found that 47 920 participants who were scanned by January 2021 showed consistent statistically significant ‘healthy’ bias compared with the ~450 000 who were not scanned.”

Healthy person biases are also observed in probability sample surveys that invite their survey respondents to participate in accelerometry studies. In one of the earliest studies to examine selection bias in accelerometry, Inoue et al. (2010) recruited a cohort of adults in Japan by mail. Of the 1,508 respondents to the questionnaire, 786 (52.1%) agreed to wear an accelerometer for seven days. The subsample invited to the accelerometry study was non-random; it included significantly more women and middle-aged and older adults. Participation, conditional on being invited, was greater among nonsmokers and persons who reported a habit of leisure walking.

Loprinzi and colleagues (2013) compared participants with valid accelerometry data (at least 4 days with at least 10 hours per day) with those with invalid accelerometry data (fewer than 4 such days) in the 2003-2004 cycle of NHANES. That is, their analyses were conditional on consent and acceptance of the device. They found that those with valid data were more likely to have a lower BMI, more likely to be non-Hispanic White, more likely to have a high school education, more likely to be married, and were older than those with invalid data. Further, adults with valid data were less likely to smoke and had a smaller number of inactive days within the last 30 days due to poor health compared to those with invalid data.

Roth and Mindell (2013) examined selection bias in the 2008 HSE accelerometry substudy, comparing those who provided sufficient accelerometry data (4-7 valid days, 10+hours per day) with those who provided less than that and those who declined. In this study the subsample offered the accelerometers was again non-random: they were older and more likely to be retired and to report having a longstanding limiting illness than the rest of the HSE participants. Among those who were invited, those providing sufficient data were older, more likely to be employed, and less likely to be a current smoker. Those who declined to wear an accelerometer did not differ significantly from those who had sufficient wear time. That is, they found participation bias in wearing the accelerometers for sufficient time, but refusers did not differ from those providing sufficient data.

Hassani et al. (2014) looked at non-consent bias to accelerometry in the Whitehall II study and found that men and participants reporting more physical activity and more favorable general health were more likely to consent. Similarly, Franzese, Schrank, and Bergmann (2023) looked at consent bias in the SHARE accelerometry study. In multivariate models examining many predictors of consent, they found that younger respondents were significantly more likely to consent. Additionally, respondents with higher self-reported frequency of moderate activities, better overall health, higher memory test score, and better computer skills had higher consent probabilities. In contrast, they found no differences in the probability for consent by gender, socio-economic factors, pain, mobility limitations, weight, personality traits, quality of life, migration status, and living conditions.

Brooks et al. (2021) compared those with accelerometry data in the UK Biobank with the larger cohort. They concluded that those with data tended to be slightly healthier, more likely to be female, younger and with a lower BMI. Leroux et al. (2021), found that individuals who were invited to the UK Biobank accelerometry study were, on average, younger and self-reported better overall health, had a lower incidence of comorbidities, lower rates of cigarette smoking, and a lower body mass index (BMI) than the larger cohort. Those who were invited and participated had lower rates of obesity, being overweight, cigarette smoking, better self-reported health, and lower rates of diabetes than those who declined or did not respond to the accelerometry study.

Weymar et al. (2015) examined accelerometry consent in a German cardiovascular examination program (50% of the 470 participants consented). They found that men were more likely to participate than women. In addition, men were more likely to participate the fewer cardiovascular risk factors they had, but cardiovascular risk factors did not play a similar role for women. Evenson et al. (2015) examined adherence in a seven-day accelerometer study among Hispanic/Latino adults in the U.S. They found that adherence was higher among participants who were male, older, employed or retired, reported higher work activity or lower recreational activity, and with a lower BMI.

This brief summary of healthy person bias suggests that participation biases exist, both in volunteer studies and in probability-based studies that recruit participants. However, the results appear to be mixed, both in terms of the relevant covariates and in terms of the size of the biases. This likely reflects variation in the recruitment context, tasks (which devices, wear time, etc.), outcomes (willingness, consent, adherence, etc.) and covariates. The present study contributes to the growing body of literature by looking at a broader set of covariates and looking at different stages of the recruitment process.

Methods and Data

Data collection methods

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a panel study that collects data on the financial, social and health-related situation of the population aged 50 years and older in 28 European countries and Israel (Börsch-Supan et al., 2013). In the eighth wave of SHARE, conducted in 2019-2020, additional measurements of physical activity using accelerometers were conducted in a sub-sample of respondents in ten countries: Denmark, Sweden, Italy, Spain, Czechia, Poland, Slovenia, Belgium, France and Germany (see Scherpenzeel et al., 2021a, 2021b). This is a secondary analysis of the SHARE accelerometer study (SAS) data (Börsch-Supan 2021, 2022).

The SAS used an Axivity AX3 (Axivity Ltd, Newcastle upon Tyne, United Kingdom). The device is waterproof and suitable for long wear times. It has no on/off switch. As devices were shipped by mail, survey agencies were instructed to start the recording immediately after configuration and prior to shipping. This resulted in large raw data files that extended before and after actual wear time.

A subsample of the panel respondents was selected before the start of fieldwork. The target was a net sample of 200 participants per country. Based on assumptions of a 75% response rate to the Wave 8 CAPI interview, a 50% consent rate to participate in the accelerometer study, and a compliance rate of 75% of consenting participants wearing the device for eight days, gross samples of about 710 persons ($200/(0.75*0.50*0.75)$) per country were needed. The gross sample was a stratified sample selected from each country's longitudinal sample before the W8 fieldwork started. It included only the panel sample, i.e., respondents who participated in SHARE before to allow the use of information from previous interviews, and excluded younger partners aged under 50. Strata were defined by age group and self-reported activity level in previous waves (see Scherpenzeel et al., 2021b for further details). Although the samples were designed to be proportionate to strata sizes, a minimum stratum size of 50 was set, meaning that the final sample was not fully proportionate to the SHARE sample.

During the SHARE face-to-face interviews, the sampled respondents were asked for their consent to participate in the accelerometer study. Next, the survey agencies mailed the devices and related material to consenting participants. With the accelerometer, participating respondents also received an “accelerometer kit”, which contained all the materials needed for wearing the accelerometer, including an information letter and an instruction brochure on how to put on the device, as well as a reply card and a prepaid return envelope to send it back after use. Respondents were instructed to attach the device to their upper thigh using the medical adhesive tape and gauze pads included in the kit. They were instructed to wear the device day and night for eight consecutive days (192 hours) before returning it.

A total of 50 devices were available per country. This meant that each device was used by several respondents. Devices were sent in batches every two weeks, starting four weeks into the field period. A maximum number of 20 accelerometers per country were sent out every two weeks, regardless of the actual number of consents obtained in that period. Given this, not all respondents who agreed to participate in the accelerometer study received a device immediately. Some respondents received the device a few weeks later (median: 31 days), while some did not receive it at all, because of the unexpected early fieldwork termination due to the COVID-19 lockdown. Respondents received an incentive for participating in the accelerometer study. This varied between countries (e.g., €20 in Germany).

Outcome variables

As noted earlier, we are interested in several outcomes: whether the respondent consented to the accelerometry study, whether they were sampled, whether they were shipped a device, whether they provided at least one day of valid data, and whether they provided eight or more days of data (fully adherent). The choice of these outcomes reflects operational exigencies as well as respondent behaviors. We expect some stages to have little effect on sample composition (RQ2 and RQ3); for example, the sampling stage was designed to be random (within country) so should have little impact, but the shipping stage (designed to re-use a limited number of devices) may be subject to logistical factors that may affect composition.

Covariates for bias analyses

The variables used for the bias analyses in RQ2 and RQ3 are the following:

- Survey country
- Socio-demographics: age categories, gender, income quartiles, education, housing tenure, household type, working status, marital status, and number of children
- SHARE response history: household respondent, family respondent, financial respondent, first wave in SHARE, ever did COVID-19 interview
- Physical activity: vigorous activity, moderate activity, difficulty walking 100m, difficulty climbing stairs, physical limitations, activities of daily living (ADL) limitations
- Physical and mental health: body mass index (BMI), any chronic conditions, depressed, happy with life, self-rated health, feel full of energy
- Health behaviors: smoking status, drank alcohol in last week, taking 5 or more prescription drugs, troubled with pain, fruit consumption, meat consumption
- Cognition: memory test score, numeracy score, self-rated reading skill

- Other: able to make ends meet, internet use in past week

These were selected from a large number of variables in SHARE because of their expected relationship with participation in the accelerometer study or healthy person bias, and because they were asked of all SHARE respondents, not just a limited subset.

Analysis methods

We conducted unweighted analyses. That is, our inference is to the sample selected for the accelerometer study in the 10 countries, rather than the broader population age 50 and older in these countries. We analyzed pooled data from the 10 participating countries. Analyses were conducted using SAS 9.4 and Stata 18.

For RQ1, we conducted simple descriptive statistics. For RQ2 we conducted a series of multivariable logistical regression models of sample loss at each stage, conditional on inclusion in the previous stage. We employed a sequential model-fitting strategy. First, we examined bivariate associations of a set of socio-demographic, survey experience and selected substantive measures with each of the outcomes of interest. Next, we employed the following sequential model-fitting strategies to fit a set of reduced multivariable models:

- 1) Remove any variables not significantly ($p < .05$) associated with any of the outcomes in the bivariate analyses. Among socio-demographic variables, only gender is not significantly associated with any of the outcomes, but we retain this a control variable.
- 2) Fit multivariable logistic models of each outcome regressed on the socio-demographic and survey experience covariates. Remove any variables not significant ($p > .05$) in any of the models.
- 3) Fit multivariable logistic models of each outcome regressed on selected substantive variables controlling for the remaining socio-demographic covariates from step 2. Remove any variables not significant ($p > .05$) in any of the models. Also drop variables with high levels of collinearity (e.g., drop vigorous activity because of high correlation with moderate activity).

For RQ3, we examine cumulative selection bias for a variety of descriptive statistics. The bias, or the signed difference in an estimate y between the selected sub-sample (y_s) and the full sample (y_f), is calculated as:

$$bias(y) = y_s - y_f$$

The standard error of the estimated bias is calculated following Lee (2006) as:

$$se(y_s - y_f) = \frac{n_f - n_s}{n_f} \sqrt{var(y_s) + var(y_{ns})}$$

Where n_f is the sample size for the full sample, n_s is the sample size for the selected sub-sample, and y_{ns} is the estimate for the sub-sample that is not selected. We test the significance of the bias estimates using large sample z-tests, dividing each estimated bias by its standard error. We examine the cumulative selection bias, i.e., the bias at each stage of participation relative to the full sample of eligible SHARE participants.

Results

RQ1: What are the sample losses at each stage of the process?

A total of 4,345 SHARE respondents were eligible for the SHARE accelerometer study (SAS) in the ten countries. Table 1 shows the key outcomes at each stage of the process.

Table 1: Key outcomes of the SHARE accelerometer study (SAS)

	N	% of eligible	% of consented	% of sampled	% of shipped	Cumulative sample loss (%)
Eligible for SAS	4,345					0%
Consented	2,363	54.4%				45.6%
Sampled for SAS	1,467	33.8%	62.1%			66.3%
Shipped a device	1,073	24.7%	45.4%	73.1%		75.3%
At least 1 day of valid wear time	856	19.7%	36.2%	58.4%	79.8%	80.3%
Eight or more days of wear time (fully adherent)	521	12.0%	22.0%	35.5%	48.6%	88.0%

Table 1 shows significant sample loss across the stages of the study. The biggest loss (45.6% of the sample) occurs at the consent stage. About two-thirds (62.1%) of those who consented were sampled for the SAS.

As noted earlier, not everyone who was sampled for the SAS was sent a device. There were a variety of reasons for this. Of the 394 sampled for SAS but not sent a device, 355 were classified as “No shipment of device (unknown reason),” while 33 were classified as “No device sent due to Covid-19,” and 6 as “Refusal.” Nonetheless, these sample losses – whether due to respondent unwillingness or inability to follow the protocol or logistical or operational issues – likely reflect what may really occur in complex studies such as this.

If we assume that sampling was random (as designed), if everyone who consented was sampled and sent a device, then the percentage of eligible respondents with at least one day of valid wear time could be as high as 43.4% (0.544×0.798) and the percentage fully adherent could be 26.4% (0.544×0.486). This still represents significant sample loss. We ran a logistic regression (not shown) to test whether the sampling step was random (conditional on consenting). Aside from significant country-level differences (reflecting different sampling rates by country), only education reaches statistical significance. Given this, we drop the sampling step from subsequent analyses.

RQ2: What are the predictors of participation at each stage of the process?

Here we are particularly interested in socio-demographic and health-related covariates of participation. A key question is whether the variables associated with sample loss at one stage (e.g., consent) are the same or different than those associated with loss at other stages (e.g., receiving a device, wearing a device). That is, do there appear to be similar or different mechanisms operating at each stage?

Bivariate associations of a set of socio-demographic, survey experience and selected substantive measures with each of the outcomes of interest are presented in Appendix Table B1 and B2. Remember that our objective is not to find the best predictors of each stage of participation, but rather to explore similarities and differences between the outcomes in terms of covariates. Table 2 contains a summary of

the significance tests from the reduced multivariable models described earlier. Appendix Table B3 has the average marginal effects (AMEs) and standard errors (s.e.'s) from these reduced models.

Table 2: Multivariable Models of Participation at Selected Stages: P-values from Tests of Statistical Significance from Reduced Models

Variable	Consent	Shipped, given consent and sampled	Wear 1+ days, given shipped	Wear 8+ days, given shipped
<i>Socio-demographic and survey experience variables</i>				
Country	<0.001	<0.001	<0.001	<0.001
Age	0.16	0.80	0.002	0.21
Gender	0.087	0.42	0.42	0.67
Income	0.10	0.55	0.52	0.041
Education	0.19	0.53	0.23	0.99
Working status	0.15	0.014	0.94	0.42
Number of children	0.004	0.83	0.25	0.18
Financial R	0.042	0.90	0.06	0.025
Ever did COVID survey	<0.001	0.002	0.97	0.11
<i>Selected substantive variables</i>				
Moderate activity	<0.001	0.81	0.58	0.004
Self-rated health	0.10	0.09	0.074	0.13
Feel full of energy	<0.001	0.003	0.34	0.98
Smoking status	0.012	0.62	0.040	0.70
Alcohol	<0.001	0.95	0.93	0.41
5+ drugs per day	0.001	0.10	0.32	0.15
Memory score	0.005	0.008	0.78	0.53
Numeracy score	0.007	0.24	0.28	0.020
Reading skill	0.052	0.080	0.70	0.48
Internet use	<0.001	0.065	0.13	0.043

The only variable that is significant across all stages of the accelerometry data collection is country. There is considerable variation in consent rates across countries, ranging from a low of 33.7% in Czechia to a high of 70.2% in Poland. For later stages, this variation is partly an artifact of the sample design for the SAS (samples of roughly equal size were selected in each country, rather than proportionate to size of the countries). While several other socio-demographic variables are significantly associated with participation at various stages in the bivariate analyses (see Appendix B3), few of these remain in the multivariable models. However, several substantive variables are significantly associated with consent. Fewer variables are significantly associated with later stages of participation, but this may be due to sample loss across stages (see Table 1). Age remains a significant predictor of wearing the device at least once, conditional on being shipped a device. Income is significantly associated with being fully adherent. Engaging in moderate activity is associated with consent and with full adherence, conditional on receiving a device.

Focusing on specific examples from the detailed results in Appendix B, we can see that age is associated with participation at all stages in the bivariate tables (Appendix Table B1), with older people (age 80+) being less likely to consent and less likely to wear the activity tracker than younger people (age 50-59). However, the younger age cohort is also less likely to be fully compliant than the middle cohorts (ages 60-79). In the multivariable models, the effect for wear 1+ days remains significant, with the older cohorts being more likely to wear the device (+15.5 percentage points [p.p.] for the 60-69 group and +7.2 p.p. for the 80+ group). Those who report engaging in moderate physical activity more than once a week are more likely to consent (+15.6 p.p. relative to those who hardly ever or never engage in physical activity). The moderate activity group is also more likely to be fully adherent to the protocol, conditional on being shipped a device (+17.4 p.p. relative to the low/no activity group). Similarly, those who report often feeling full of energy consent at a higher rate (+9/3 p.p.) than those who never feel full of energy; however, there are no significant differences in adherence.

In summary, a number of socio-demographic variables, survey experience and substantive variables are associated with the various stages of participation. When controlling for other variables in a multivariable model, a number of these associations remain significant. However, we see little evidence of these effects increasing over each successive stage of selection. This supports the conclusion from RQ1, that the biggest sample loss is at the consent stage, and the strongest associations with participation occur at that stage. This suggests that fieldwork efforts targeted at reducing disparities in consent are likely to be most productive.

RQ3: What are the selection biases at each stage of the study?

In an activity-based study such as this, our particular interest is in health-related outcomes. Again, we are interested in whether these selection biases compound over successive stages of the study, or whether the effects are largely offsetting.

Appendix Table C1 presents the selection bias estimates for selected socio-demographic variables. These are all relative to the full eligible sample estimate in the second column. To illustrate, 18.5% of the eligible SHARE respondents in these 10 countries are age 80 and over, but only 11.7% (18.5 – 6.8) of those who provided full accelerometer data (8+ days) are in this age group, representing a significant ($p < .05$) underrepresentation of this group. We see that older respondents (80+) are under-represented at all stages of the study, but this bias does not appear to increase substantially (from -4.2 at consent to -6.8 at 8+ days wear). Those in the lowest income quartile are less likely to consent, but more likely to participate in later stages of the study (i.e., the sign reverses), whereas those in the highest income quartile and those in the lowest education level are less likely to participate at all stages of the process. The last row of Table C1 shows the average absolute bias across all demographic variables. The average bias tends to increase slightly, from 1.9 percentage points at consent to 3.3 percentage points at full adherence. The average standard errors also tend to increase, as expected with the cumulative sample loss across the stages of participation.

Appendix Table C2 similarly shows the selection bias estimates for selected substantive variables. Again, these are all relative to the full eligible sample estimate in column 2. The pattern of biases is not always consistent or expected across all stages. For example, those whose scores on the memory test were classified as “excellent or very good” are under-represented among full adherents, but those whose

score was “good” are over-represented, and there is no significant bias in full adherence among those with “fair or poor” memory scores.

The Internet use variable is an interesting one. Those who reported recently using the Internet are over-represented among participants at all stages of the process; however, the bias does not increase much over the course of the study, suggesting that this may be associated with consent, but less so with the later stages of participation. Computer or Internet skills are not necessary for participation in the accelerometry study, but perhaps this variable serves as a proxy for technical ability.

As our focus in RQ3 is on the healthy person bias hypothesis, Figure 1 presents the biases at each stage for selected indicators of health or physical activity. The numbering on the horizontal axis denotes the variable examined, as documented in the notes for Figure 1. The vertical axis denotes the bias, that is, the percentage point difference between the sub-sample selected at that stage and the full sample. The four panels in Figure 1 represent the four stages of participation: consent, shipment, 1-day wear, and 8-day wear.

Figure 1: Selection Biases for Selected Health Variables

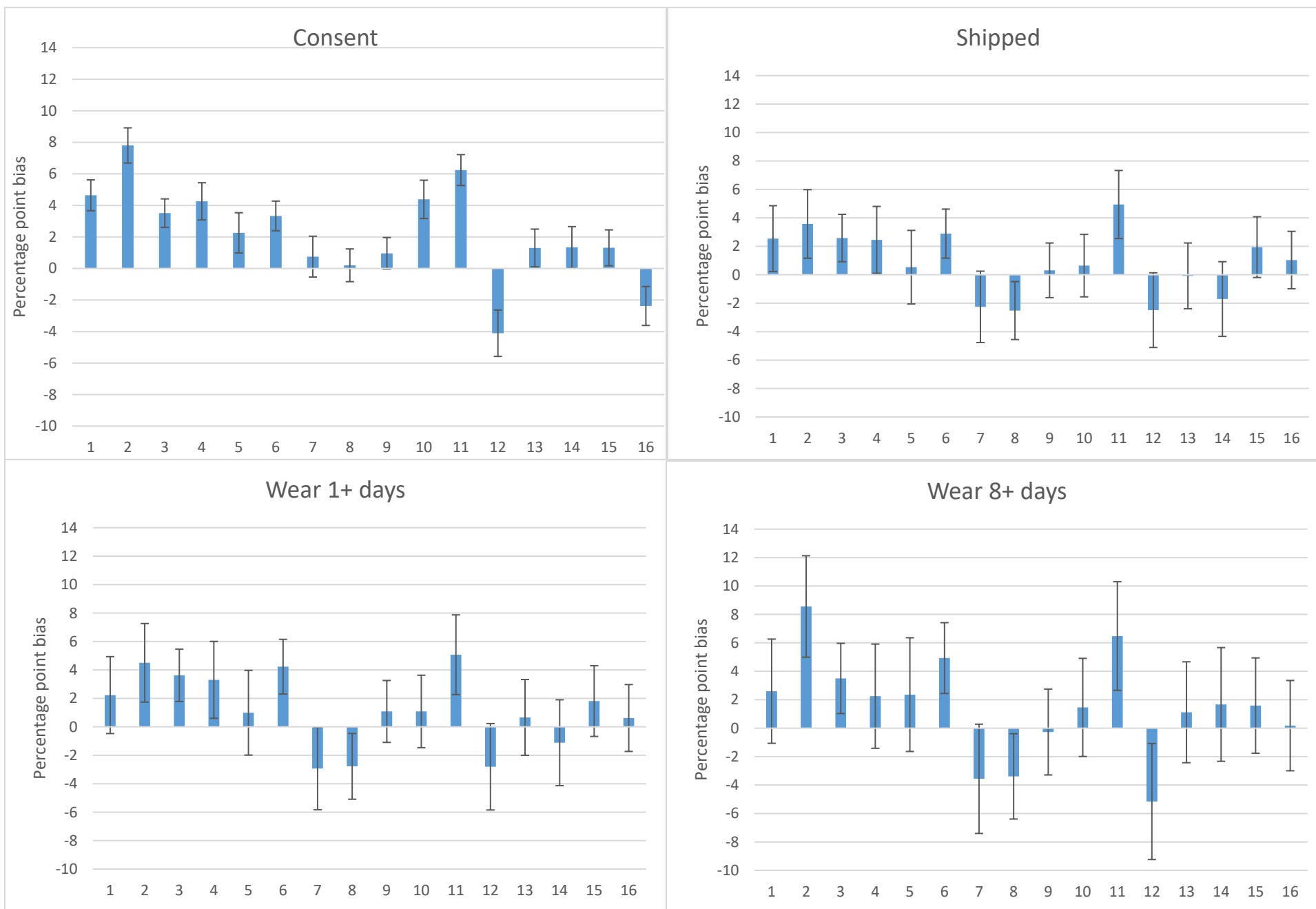


Figure 1 Legend: Variables and Categories

1=Vigorous activity more than once a week; 2=Moderate activity more than once a week; 3=No difficulty walking 100 m; 4=No difficulty climbing stairs; 5=No physical limitations (GALI); 6=No ADL (activities of daily living) limitations; 7=Normal BMI; 8=No chronic diseases; 9=Not hospitalized in last 12 months; 10=Self-rated health excellent or very good; 11=Often feel full of energy; 12=Never smoked; 13=Taking fewer than 5 drugs per day; 14=Currently experiencing no pain; 15=Eat fruit every day; 16=Eat meat rarely or never

From Figure 1, we can see that, for consent, most of the biases are positive, indicating a more healthy or active lifestyle among those who consented, relative to the full population of eligible SHARE respondents. For example, those who report engaging in moderate activity more than once a week (variable 2) are over-represented in the group that consented by 7.8 percentage points, relative to the full sample. However, not all of the biases are statistically significant as indicated by the 95% confidence interval error bars. Further, there are some notable exceptions to the “healthy volunteer bias” expectation: those who report having never smoked (variable 12) and those who report never or rarely eating meat (variable 16) are significantly under-represented among consenters.

Looking across the four charts, we see little evidence that the selection biases increase (or compound) over the different stages of participation. For example, those who report engaging in moderate activity more than once a week (variable 2 again) are overrepresented in the group that was full adherent by 8.6 percentage points, relative to the full sample. Those who report having never smoked are under-represented by 4.1 percentage points among consenters and by 5.2 percentage points among fully adherent participants. As expected, given the sample size reductions, the error bars become larger and more of them span zero (indicating non-significant bias) across the stages of participation.

Looking at the average absolute bias across all substantive variables (see the last row of Table C2), we see that, as with the socio-demographic variables, the average standard errors increase over each stage. However, we see little evidence of the average absolute bias increasing: it is 2.4 percentage points at consent, declines slightly for the next two stages and is 2.5 percentage points for full adherence.

Thus, the evidence for healthy person bias can be described as mixed. Nonetheless, there are substantial biases (as high as 9.2 percentage points) for some variables, suggesting caution in assuming that the set of fully-adherent participants is representative of the SHARE respondent population in terms of substantive related to health and physical activity.

Discussion

We examined the extent and effect of sample loss over several stages of participation in an accelerometer-based activity tracker study conducted as part of an ongoing longitudinal survey of aging in ten European countries. We observed considerable sample loss over the stages of participation, from consent to being shipped a device, to wearing the device for at least one day to full adherence, or wearing the device for eight days. The biggest losses occur at the consent stage, administered during an in-person interview. This points to potential opportunities to reduce this stage of loss, whether through enhanced training of interviewers or improved consent materials and processes. Franzese, Schrank, and Bergmann (2023) explored reasons for non-consent in the SAS, and this is a good starting point for identifying barriers to consent and potential interventions. Having received a device, four out of five participants wear it for at least one day, but fewer than half (48.6%) do so for all eight days requested.

Some of the sample losses are potentially avoidable. For example, a subset of those who consented were selected for the accelerometry study and sent a device. This was done to save money. However, even if everyone who consented was sent a device, this would increase the percentage of fully-adherent participants among all eligible participants from 12% to 26.4%, still representing substantial loss. Finding ways to increase the number of consenting participants – while still adhering to the principle of

informed consent would both increase the number of potential participants and facilitate targeted sampling in later stages to compensate for differential sample loss. As the cost of wearable devices targeted at consumers comes down, the possibility of providing everyone with a device – and having participants wear the devices for longer – becomes more feasible for large-scale population studies. A hybrid solution, where devices are only provided to those who do not already have one – could also further reduce costs and increase numbers of participants.

In terms of differential sample loss, we see that the biggest effects are found at the consent stage. We do not see much evidence of differential sample loss increasing further over the later stages of participation. Relatively few variables are significantly associated with all four stage of participation we examined in bivariate analyses, and even fewer remain significant across all stages when controlling for other covariates in a multivariable model.

We find little evidence that bias is increasing systematically across the different stages of participation. The average absolute bias (see Table C2, last row) for all substantive variables is 2.4, decreasing to 1.7 at the sampling and shipping and initial adherence stages, and then increasing slightly to 2.5 percentage points for full adherence. There are some notable exceptions, but the general extent of the bias is relatively small. Focusing more narrowly on selected health variables, we see some evidence supporting the “healthy volunteer” hypothesis, but this is by no means consistent or particularly strong. We see little evidence of bias accumulating over the stages of participation. That is, the biases observed at the consent stage seem largely to persist throughout the process.

There are several practical implications of our findings. First, given that socio-demographic data are already available on panel members, demographic differences in selection – especially at the consent stage – can be reduced, either by oversampling those who are less likely to participate, or by developing strategies to increase consent and participation rates among such groups. Second, weighting adjustments can be applied to minimize the impact of such disparities in consent. Third, making more devices available for participants and reducing the wait time between consent and shipping could reduce sample loss at this stage. If sampling is still required, disproportionate selection may help compensate for differential losses at the consent stage. Fourth, given the substantial drop off after initial wear – 80% of those sent a device wear it for at least one day but fewer than half do so for all eight days – understanding reasons for non-adherence among those sent a device would be fruitful. One possibility is the added inconvenience of attaching a research-grade accelerometer to the thigh, relative to a wrist-worn device. Similarly, finding ways to detect non-wear in real-time and provide reminders to participants may help reduce non-adherence.

In summary then, the process of having existing survey panel members wear an accelerometer for at least a week results in substantial sample loss. In volunteer studies, such losses occur before any measurement, making it hard to ensure (or measure) representation in such accelerometer-based activity studies. Despite the substantial sample loss, the resulting selection biases are not large and are potentially manageable.

References

- Ahn, J.V., Sera, F., Cummins, S., and Flouri, E. (2018), "Associations between Objectively Measured Physical Activity and Later Mental Health Outcomes in Children: Findings from the UK Millennium Cohort Study." *Journal of Epidemiology and Community Health*, 72 (2): 94-100, <http://dx.doi.org/10.1136/jech-2017-209455>
- Antoun, C., Conrad, F.G., Couper, M.P., and West, B.T. (2019), "Simultaneous Estimation of Multiple Sources of Error in a Smartphone-Based Survey." *Journal of Survey Statistics and Methodology*, 7 (1): 93–117, <https://doi.org/10.1093/jssam/smy002>.
- Antoun, C., and Wenz, A. (2021), "Participation Metrics for Accelerometer-based Research." Paper presented at the Mobile Apps and Sensors in Surveys workshop, virtual, March, <https://massworkshop.org/2021-workshop/mass-2021-antoun-wenz/>.
- Bergmann, M., Franzese, F., and Schrank, F. (2022), "Determinants of Consent in the SHARE Accelerometer Study." SHARE Working Paper Series 78-2022, <https://www.doi.org/10.17617/2.3372782>.
- Blom, J., Yin, L., Lidén, A., Dolk, A., Jeppsson, B., Pålman, L., Holmberg, L., and Nyrén, O. (2008). "A 9-Year Follow-up Study of Participants and Nonparticipants in Sigmoidoscopy Screening: Importance of Self-Selection." *Cancer Epidemiology, Biomarkers and Prevention*, 17 (5): 1163-1168, <https://doi.org/10.1158/1055-9965.EPI-07-2764>.
- Börsch-Supan, A. (2021). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 8. Release Version: 1.0.0. SHARE-ERIC. Data Set. <https://www.doi.org/10.6103/SHARE.w8.100>.
- Börsch-Supan, A. (2022). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 8. Accelerometer sampling version 2. Release version: 8.0.0. SHARE-ERIC. Internal data set.
- Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., and Zuber, S. (2013), "Data Resource Profile: the Survey of Health, Ageing and Retirement in Europe (SHARE)." *International Journal of Epidemiology*, 42 (4): 992-1001, <https://www.doi.org/10.1093/ije/dyt088>.
- Brayne, C., and Moffitt, T.E. (2022). "The Limitations of Large-Scale Volunteer Databases to Address Inequalities and Global Challenges in Health and Aging." *Nature Aging*, 2 (September): 775-783, <https://doi.org/10.1038/s43587-022-00277-x>
- Brooks, T.G., Lahens, N.F., Grant, G.R., Sheline, Y.I., FitzGerald, G.A., and Skarke, C. (2021). "Phenome-Wide Association Study of Actigraphy in the UK Biobank." medRxiv preprint, <https://doi.org/10.1101/2021.12.09.21267558>
- Cato, M.S., Wyka, K., Ferris, E.B., Evenson, K.R., Wen, M., Dorn, J.M., Thorpe, L., and Huang, T. T.-K. (2020), "Correlates of Accelerometry Non-Adherence in an Economically Disadvantaged Minority Urban Adult Population." *Journal of Science and Medicine in Sport*, 23 (8): 746-752, <https://doi.org/10.1016/j.jsams.2020.01.013>

Chaturvedi, R.R., Angrisani, M., and Couper, M.P. (2023), “American Life in Realtime – Benchmark Person-generated Health Data for Equitable Digital Health.” Paper presented at the CIPHER workshop, Washington DC, March.

Chaturvedi, R.R., Angrisani, M., Troxel, W.M., Gutsche, T., Ortega, E., Jain, M., Boch, A., and Kapteyn, A. (2023), “American Life in Realtime: A Benchmark Registry of Health Data for Equitable Precision Health.” *Nature Medicine* 29: 283-286, <https://doi.org/10.1038/s41591-022-02171-w>.

Cho, P.J., Yi, J., Ho, E., Shandhi, M.M.H., Dinh, Y., Patil, A., Martin, L., Singh, G., Bent, B., Ginsburg, G., Smuck, M., Woods, C., Shaw, R., and Dunn, J. (2022), “Demographic Imbalances Resulting From the Bring-Your-Own-Device Study Design.” *JMIR Mhealth Uhealth*, 10 (4): e29510, <https://doi.org/10.2196/29510>.

Couper, M.P. (2019), “Mobile Data Collection: A Survey Researcher’s Perspective.” Paper presented at the Mobile Apps and Sensors in Surveys workshop, Mannheim, March, <https://massworkshop.org/2019-workshop/>

Doherty, A., Jackson, D., Hammerla, N., Ploëtz, T., Olivier, P., Granat, M.H., White, T., van Hees, V.T., Trenell, M.I., Owen, C.G., Preece, S.J., Gillions, R., Sheard, S., Peakman, T., Brage, A., and Wareham, N.J. (2017), “Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study.” *PLoS ONE*, 12 (2): e0169649, <https://doi.org/10.1371/journal.pone.0169649>

Elevelt, A., Bernasco, W., Lugtig, P., Ruiter, S., and Toepoel, V. (2021), “Where You At? Using GPS Locations in an Electronic Time Use Diary Study to Derive Functional Locations.” *Social Science Computer Review*, 39 (4): 509-526, <https://doi.org/10.1177/0894439319877872>.

Evenson, K.R., Sotres-Alvarez, D., Deng, Y.U., Marshall, S.J., Isasi, C.R., Esliger, D.W., and Davis, S. (2015), “Accelerometer Adherence and Performance in a Cohort Study of US Hispanic Adults.” *Medicine and Science in Sports and Exercise*, 47 (4): 725-34, <https://doi.org/10.1249/MSS.0000000000000478>.

Fakhouri, T.H.I., Martin, C.B., Chen, T.C., Akinbami, L.J., Ogden, C.L., Paulose-Ram, R., et al. (2020), “An Investigation of Nonresponse Bias and Survey Location Variability in the 2017–2018 National Health and Nutrition Examination Survey.” National Center for Health Statistics, Vital and Health Statistics Series 2, No. 185, <https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx#estimation-and-weighting-procedures>.

Franzese, F., Schrank, F., and Bergmann, M. (2023), “Determinants of Consent in the SHARE Accelerometer Study.” *Journal for the Measurement of Physical Behaviour*, 6 (4): 271–288, <https://doi.org/10.1123/jmpb.2022-0046>.

Fry, A., Littlejohns, T.J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., and Allen, N.E. (2017). “Comparison of Sociodemographic and Health-related Characteristics of UK Biobank Participants with Those of the General Population.” *American Journal of Epidemiology*, 186 (9): 1026-1034, <https://doi.org/10.1093/aje/kwx246>

Füzéki, E., Engeroff, T., and Banzer, W. (2017), “Health Benefits of Light-Intensity Physical Activity: A Systematic Review of Accelerometer Data of the National Health and Nutrition Examination Survey (NHANES).” *Sports Medicine*, 47: 1769-1793, <https://doi.org/10.1007/s40279-017-0724-0>.

- Gilbert, E., Conolly, A., Tietz, S., Calderwood, L., and Rose, N. (2017), "Measuring Young People's Physical Activity Using Accelerometers in the UK Millennium Cohort Study." London: Centre for Longitudinal Studies, Working paper 2017/15, https://cls.ucl.ac.uk/working_papers/
- Griffiths, L.J., Cortina-Borja, M., Sera, F., Poulou, T., Geraci, M., Rich, C., Cole, T.J., Law, C., Joshi, H., Ness, A.R., Jebb, S.A., and Dezaux, C. (2013a), "How Active Are Our Children? Findings from the Millennium Cohort Study." *BMJ Open*, 3: e002893, <https://doi.org/10.1136/bmjopen-2013-002893>.
- Griffiths, L.J., Rich, C., Geraci, M., Sera, F., Cortina-Borja, M., Poulou, T., Platt, L., Johnson, J., and Dezaux, C. (2013b), Technical Report on the Enhancement of Millennium Cohort Study Data with Accelerometer-Derived Measures of Physical Activity and Sedentary Behaviour in Seven Year Olds. London: Centre for Longitudinal Studies, technical report, http://www.esdsacuk/doc/7238/mrdoc/pdf/mcs4_pa_technical_report.pdf.
- Groves, R.M. (2006), "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly*, 70 (5): 646-675.
- Groves, R.M., Fowler, F.J. Jr., Couper, M.P., Lepkowski, J.M., Singer, E., and Tourangeau, R. (2009), *Survey Methodology, 2nd Edition*. New York: Wiley.
- Hamer, M., Coombs, N., and Stamatakis, E. (2014), "Associations between Objectively Assessed and Self-reported Sedentary Time with Mental Health in Adults: an Analysis of Data from the Health Survey for England." *BMJ Open*, 4: e004580, <https://doi.org/10.1136/bmjopen-2013-004580>
- Hassani, M., Kivimaki, M., Elbaz, A., Shipley, M., Singh-Manoux, A., and Sabia, S. (2014), "Non-Consent to a Wrist-Worn Accelerometer in Older Adults: The Role of Socio-Demographic, Behavioural and Health Factors." *PLoS ONE*, 9 (10): e110816, <https://doi.org/10.1371/journal.pone.0110816>
- Horn, D., Burton, J., Couper, M.P., Jäckle, A., and Vine, J. (2022), "Selectivity of Digital Trace Data: the Case of the UK COVID-19 Contact Tracing Apps." Paper presented at the Understanding Society Survey Methods Conference, virtual, September 21st.
- Inoue, S., Ohya, Y., Odagiri, Y., Takamiya, T., Kamada, M., Okada, S., Tudor-Locke, C., and Shimomitsu, T. (2010), "Characteristics of Accelerometry Respondents to a Mail-based Surveillance Study." *Journal of Epidemiology*, 20 (6): 446-52, <https://doi.org/10.2188/jea.je20100062>.
- Jäckle, A., Burton, J., and Couper, M.P. (2023), "Understanding Society: Minimizing Selection Biases in Data Collection Using Mobile Apps", *Fiscal Studies*, 44:361–376. <https://doi.org/10.1111/1475-5890.12351>.
- Jäckle, A., Burton, J., Couper, M.P., and Lessof, C. (2019), "Participation in a Mobile App Survey to Collect Expenditure Data as Part of a Large-Scale Probability Household Panel: Response Rates and Response Biases." *Survey Research Methods*, 13 (1): 23-44, <https://doi.org/10.18148/srm/2019.v1i1.7297>.
- Lee, P.H. (2013), "Data Imputation for Accelerometer-Measured Physical Activity: The Combined Approach." *American Journal of Clinical Nutrition*, 97 (5): 965-71, <https://doi.org/10.3945/ajcn.112.052738>.

- Lee, S. (2006), "An Evaluation of Nonresponse and Coverage Errors in a Prerecruited Probability Web Panel Survey." *Social Science Computer Review*, 24 (4): 460-475, <https://doi.org/10.1177/0894439306288085>.
- Klijs, B., Scholtens, S., Mandemakers, J.J., Snieder, H., Stolk, R.P., and Smidt, N. (2015), "Representativeness of the LifeLines Cohort Study." *PLoS ONE*, 10 (9): e0137203. <https://doi.org/10.1371/journal.pone.0137203>.
- Leroux, A., Xu, S., Kundu, P., Muschelli, J., Smirnova, E., Chatterjee, N., and Crainiceanu, C. (2021). "Quantifying the Predictive Performance of Objectively Measured Physical Activity on Mortality in the UK Biobank." *Journals of Gerontology: Medical Sciences*, 76 (2): 1486-1494, <https://doi.org/10.1093/gerona/glaa250>
- Liu, B., Yu, M., Graubard, B.I., Troiano, R.P., and Schenker, N. (2016), "Multiple Imputation of Completely Missing Repeated Measures Data Within Person from a Complex Sample: Application to Accelerometer Data in the National Health and Nutrition Examination Survey." *Statistics in Medicine*, 35 (28): 5170-5188, <https://doi.org/10.1002/sim.7049>.
- Loprinzi, P.D., Cardinal, B.J., Crespo, C.J., Brodowicz, G.R., Andersen, R.E., and Smit, E. (2013), "Differences in Demographic, Behavioral, and Biological Variables between Those with Valid and Invalid Accelerometry Data: Implications for Generalizability." *Journal of Physical Activity and Health*, 10: 79-84, <https://doi.org/10.1123/jpah.10.1.79>.
- Lyall, D.M., Quinn, T., Lyall, L.M., Ward, J., Anderson, J.J., Smith, D.J., Stewart, W., Strawbridge, R.J., Bailey, M.E.S., and Cullen, B. (2022), "Quantifying Bias in Psychological and Physical Health in the UK Biobank Imaging Sub-sample." *Brain Communications*, 4 (3): fcac119, <https://doi.org/10.1093/braincomms/fcac119>.
- Kapteyn, A., Saw, H.-W., and Weerman, B. (2022), "Measuring Air Quality with Wearable Devices." Paper presented at the Mobile Apps and Sensors in Surveys workshop, Utrecht, June, <https://massworkshop.org/programme-of-the-2022-workshop/kapteyn-et-al/>
- Master, H., Annis, J., Huang, S., Beckman, J.A., Ratsimbazafy, F., Marginean, K., Carroll, R., Natarajan, K., Harrell, F.E., Roden, D.M., Harris, P. and Brittain, E.L. (2022), "Association of Step Counts Over Time with the Risk of Chronic Disease in the *All of Us* Research Program." *Nature Medicine*, 28: 2301-2308, <https://doi.org/10.1038/s41591-022-02012-w>.
- McCool, D., Schouten, B., Mussmann, O., and Lugtig, P. (2021), "An App-Assisted Travel Survey in Official Statistics. Possibilities and challenges." *Journal of Official Statistics*, 37 (1): 149-170, <https://doi.org/10.2478/jos-2021-0007>.
- McCrorie, P. (2017), "Integrating GPS Technology into Large Scale, Population Level, Data Collections: Practical Utility for Science, and Concerns and Considerations Regarding its Application in 10-11 Year Old Children." Paper presented at the CLOSER workshop, London, May.
- McCrorie, P. (2022), "The SPACES study: Harnessing the Value of Wearable Technology and GUS Data Linkage to Explore the Impact of The Built, Natural, and Social Environment on Young People's Health and Wellbeing." Paper presented at the virtual Growing Up in Scotland conference, June.

- McCrorie, P., Walker, D., and Ellaway, A. (2018), "The Unanticipated Challenges Associated with Implementing an Observational Study Protocol in a Large Scale Physical Activity and GPS Data Collection." *JMIR Research Protocols*, 7 (4): e110, <https://doi.org/10.2196/resprot.9537>
- Oh, A.Y., Caporaso, A., Davis, T., Dwyer, L.A., Nebeling, L.C., Liu, B., and Hennessy, E. (2021), "Effect of Incentive Amount on US Adolescents' Participation in an Accelerometer Data Collection Component of a National Survey." *Field Methods*, 33 (3): 219-235, <https://doi.org/10.1177/1525822X21989841>.
- Olsen, J.R., Mitchell, R., McCrorie, P., and Ellaway, A. (2019), "Children's Mobility and Environmental Exposures in Urban Landscapes: A Cross-sectional Study of 10–11 Year Old Scottish Children." *Social Science and Medicine*, 22: 11-22, <https://doi.org/10.1016/j.socscimed.2019.01.047>
- Perry, A.S. Annis, J.S. Master, H., Naylor, M., Hughes, A., Kouame, A., Natarajan, K., Marginean, K., Murthy, V., Roden, D.M., Harris, P.A., Shah, R., and Brittain, E.L. (2023), "Association of Longitudinal Activity Measures and Diabetes Risk: An Analysis From the National Institutes of Health *All of Us* Research Program." *Journal of Clinical Endocrinology and Metabolism*, 108 (5): 1101-1109, <https://doi.org/10.1210/clinem/dgac695>.
- Pinsky, P.F., Miller, A., Kramer, B.S., Church, T., Reding, D., Prorok, P., Gelmann, E., Schoen, R.E., Buys, S., Hayes, R.B., and Berg, C.D. (2007), "Evidence of a Healthy Volunteer Effect in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial." *American Journal of Epidemiology*, 165 (8): 874-881, <https://doi.org/10.1093/aje/kwk075>.
- Roth, M.A., and Mindell, J.S. (2013), "Who Provides Accelerometry Data? Correlates of Adherence to Wearing an Accelerometry Motion Sensor: The 2008 Health Survey for England." *Journal of Physical Activity and Health*, 10 (1): 70-78, <https://journals.humankinetics.com/view/journals/jpah/10/1/article-p70.xml>.
- Sakshaug, J.W. (2021), "Measuring and Controlling for Non-Consent Bias in Linked Survey and Administrative Data." Pp 155-178 in A.Y. Chun, M.D. Larsen, G. Durrant, and J.P. Reiter (eds.), *Administrative Records for Survey Methodology*. New York: Wiley.
- Sakshaug, J.W., Hülle, S., Schmucker, A., and Liebig, S. (2017), "Exploring the Effects of Interviewer- and Self-Administered Survey Modes on Record Linkage Consent Rates and Bias." *Survey Research Methods*, 11 (2): 171–188, <https://doi.org/10.18148/srm/2017.v11i2.7158>
- Scherpenzeel, A., Angleys, N., Franzese, F., and Weiss, L. (2021a), "Asking for Participation in the SHARE Accelerometer Study." Pp. 56–57, 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.
- Scherpenzeel, A., Angleys, N., Franzese, F., and Weiss, L. (2021b), "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.
- Spartano, N.L., Lin, H., Sun, F., Lunetta, K.L., Trinquart, L., Valentino, M., Manders, E.S., Pletcher, M.J., Marcus, G.M., McManus, D.D., Benjamin, E.J., Fox, C.S., Olgin, J.E., and Murabito, J.M. (2019), "Comparison of On-Site Versus Remote Mobile Device Support in the Framingham Heart Study Using the

Health eHeart Study for Digital Follow-up: Randomized Pilot Study Set Within an Observational Study Design." *JMIR Mhealth Uhealth*, 7 (9): e13238, <https://doi.org/10.2196/13238>

Sudlow, C., Gallacher, J., Allen, N., et al. (2015), "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age." *PLoS Med*, 12 (3): e1001779, <https://doi.org/10.1371/journal.pmed.1001779>

Tourangeau, R. (2020), "How Errors Cumulate: Two Examples." *Journal of Survey Statistics and Methodology*, 8 (3): 413–432, <https://doi.org/10.1093/jssam/smz019>.

Troiano, R.P., Berrigan, D., Dodd, K.W., Mâsse, L.C., Tilert, T., and McDowell, M. (2008), "Physical Activity in the United States Measured by Accelerometer." *Medicine & Science in Sports & Exercise*, 40 (1): 181-188, <https://doi.org/10.1249/mss.0b013e31815a51b3>.

Vine, J., Burton, J., Couper, M.P. and Jäckle, A. (2021), "Measuring Smartphone Operating System Versions in Surveys: How to Identify Who Has Devices Compatible with Survey Apps." Paper presented at the virtual General Online Research Conference, September.

Weymar, F., Braatz, J., Guertler, D., van den Berg, N., Meyer, C., John, U., Felix, S.B., Dörr, M., and Ulbricht, S. (2015), "Characteristics associated with non-participation in 7-day accelerometry." *Preventive Medicine Reports*, 2: 413-418, <https://doi.org/10.1016/j.pmedr.2015.05.003>

Yang, D., Fricker, S., and Eltinge, J. (2019), "Methods for Exploratory Assessment of Consent-To-Link in a Household Survey," *Journal of Survey Statistics and Methodology*, 7 (1): 118-155, <https://doi.org/10.1093/jssam/smx031>

Appendix Table A: Summary of Large-scale Population-based Accelerometry Studies

	Population	Eligible sample size	Device	Wear time and location	Placement mode and incentive	Participation rates (among eligible cases)
<i>Probability-based studies</i>						
NHANES 2003-6	Adults in U.S.	---	ActiGraph AM-7164	Hip; 7 days	FTF; \$40 cash	---
NHANES 2011-14	Adults in U.S.	7,821 in 2011	Actigraph GT3X+	Wrist; 7 days	FTF; \$40 cash	88% consented and returned devices; 83% fully adherent
HSE	Adults in England	4,507	Actigraph GT1M	Wrist; 7 days	FTF; £20 voucher	80% agreed to participate; 56% with sufficient data
SPACES	Children age 10-11 in Scotland	2,162	ActiGraph GT3X+	Waist; 7 days	Mail; none	51% returned registration form and sent devices; 36% provided sufficient data
MCS 2008-9	Children age 7 in UK	14,043	ActiGraph GT1M	Waist; 7 days	Mail; £10 voucher	94% consented; 72% returned devices; 64% some accelerometer data; 48% sufficient data (2+ days, 10+ hours)
MCS 2015-6	Children age 14 in UK	10,337	GENEActiv	Wrist, 1 weekday, 1 weekend day	FTF; none	89% given devices; 48% returned devices with valid data (41% for both days; 7% for one day)
SHARE	Adults age 50+ in 10 European countries	4,345	Axivity AX3	Upper thigh; 8 days and nights	Mail; varying (€20 in Germany)	54% consented; 25% sent a device; 12% fully adherent
ALiR	Adults in U.S.	2,153	Fitbit Inspire 2	Wrist; ongoing	Mail; up to \$126	64% consented; 48% enrolled
<i>Volunteer studies</i>						
UK Biobank	Adults age 40-69 in UK at time of recruitment	~236,000	Axivity AX3	Wrist; 7 days	Mail; none	45% consented; 41% with sufficient data
All of Us	Adults in U.S.	~350,000	Fitbit (own device)	Wrist; ongoing	None; none	---

Appendix Table B1: Bivariate Associations with Participation at Selected Stages: Socio-demographic and Survey Experience Variables

Variable	Consent		Shipped, given consent and sampled		Wear 1+ days, given shipped		Wear 8+ days, given shipped	
	%	Test statistic	%	Test statistic	%	Test statistic	%	Test statistic
Overall rates	54.4		73.1		79.8		48.6	
Country		$\chi^2=313.3$, d.f.=9, p<.0001		$\chi^2=170.9$, d.f.=9, p<.0001		$\chi^2=78.57$, d.f.=9, p<.0001		$\chi^2=126.21$, d.f.=9, p<.0001
Belgium	61.9		75.6		81.8		42.4	
Czechia	33.7		97.7		80.8		56.9	
Denmark	68.6		46.9		53.7		26.9	
France	55.1		65.4		66.4		35.3	
Germany	50.7		88.3		81.1		42.0	
Italy	35.8		75.0		77.0		28.7	
Poland	70.0		80.1		100.0		83.7	
Slovenia	61.8		86.6		86.2		66.4	
Spain	35.9		69.6		76.6		37.2	
Sweden	68.4		52.4		79.8		44.9	
Age		$\chi^2=70.4$, d.f.=3, p<.0001		$\chi^2=2.76$, d.f.=3, p=0.43		$\chi^2=18.00$, d.f.=3, p=0.0004		$\chi^2=15.16$, d.f.=3, p=0.0017
50-59	58.9		76.8		74.2		42.7	
60-69	59.5		72.9		86.1		55.6	
70-69	53.7		72.1		79.9		49.1	
80+	42.1		71.0		72.7		39.6	
Gender		$\chi^2=0.076$, d.f.=1, p=0.78		$\chi^2=1.67$, d.f.=1, p=0.20		$\chi^2=1.10$, d.f.=1, p=0.29		$\chi^2=2.96$, d.f.=1, p=0.086
Male	54.6		71.4		81.3		51.7	
Female	54.2		74.4		78.7		46.4	
Income quartile		$\chi^2=105.1$, d.f.=3, p<.0001		$\chi^2=28.96$, d.f.=3, p<.0001		$\chi^2=4.93$, d.f.=3, p=0.18		$\chi^2=5.86$, d.f.=3, P=0.12
Q1 (lowest)	49.3		79.7		83.9		54.0	
Q2	45.6		78.7		77.5		48.9	
Q3	56.9		70.4		79.7		45.4	
Q4 (highest)	65.8		64.5		77.3		45.0	
Education		$\chi^2=138.8$, d.f.=3, p<.0001		$\chi^2=8.55$, d.f.=3, p=0.036		$\chi^2=11.46$, d.f.=3, p=0.0095		$\chi^2=12.54$, d.f.=3, p=0.0057
Primary	40.3		70.5		72.5		39.0	
Lower secondary	47.0		73.1		80.5		44.8	
Upper secondary	57.8		77.0		83.9		53.8	

Tertiary	64.7		69.1		77.5		48.7	
Tenure		$\chi^2=1.38,$		$\chi^2=0.016,$		$\chi^2=3.00,$		$\chi^2=5.97,$
Owned	54.9	d.f.=1,	73.2	d.f.=1,	81.0	d.f.=1,	50.8	d.f.=1,
Rented, other	51.9	p=0.24	72.9	p=0.90	76.2	p=0.083	42.2	p=0.015
Household type		$\chi^2=9.73,$		$\chi^2=8.42,$		$\chi^2=19.82,$		$\chi^2=13.39,$
Single	50.8	d.f.=2,	75.2	d.f.=2,	71.6	d.f.=2,	40.1	d.f.=2,
Couple	56.1	p=0.0077	73.7	p=0.015	83.7	p<.0001	52.5	p=0.0012
Other	54.6		62.5		80.0		50.0	
Working status		$\chi^2=24.66,$		$\chi^2=7.40,$		$\chi^2=2.67,$		$\chi^2=9.96,$
Retired	52.9	d.f.=2,	73.3	d.f.=2,	81.0	d.f.=1,	51.8	d.f.=2,
Working	62.0	p<.0001	68.1	p=0.025	79.4	p=0.26	44.8	p=0.0069
Other	51.0		78.8		75.4		39.4	
Marital status		$\chi^2=18.48,$		$\chi^2=1.51,$		$\chi^2=20.89,$		$\chi^2=15.25,$
Married, civil union	55.6	d.f.=3,	72.2	d.f.=3,	83.3	d.f.=3,	52.2	d.f.=3,
Single	54.8	p=0.0003	76.2	p=0.68	65.6	p=0.0001	39.1	p=0.0016
Separated, divorced	57.6		75.2		70.5		34.8	
Widowed	47.2		74.9		76.0		45.5	
Number of children		$\chi^2=11.08,$		$\chi^2=4.39,$		$\chi^2=9.88,$		$\chi^2=8.35,$
None	48.6	d.f.=3,	74.3	d.f.=3,	66.7	d.f.=3,	34.6	d.f.=3,
One	55.6	p=0.011	74.3	p=0.22	83.0	p=0.020	53.6	p=0.039
Two	52.9		75.3		80.9		49.5	
Three or more	57.1		69.9		79.6		47.6	
Household respondent		$\chi^2=1.20,$		$\chi^2=0.082,$		$\chi^2=7.90,$		$\chi^2=10.48,$
Yes	54.9	d.f.=1,	72.9	d.f.=1,	77.6	d.f.=1,	45.4	d.f.=1,
No	53.1	p=0.27	73.7	p=0.77	85.2	p=0.0049	56.4	p=0.0012
Family respondent		$\chi^2=0.97,$		$\chi^2=0.14,$		$\chi^2=7.38,$		$\chi^2=9.23,$
Yes	54.9	d.f.=1,	72.9	d.f.=1,	77.7	d.f.=1,	45.7	d.f.=1,
No	53.2	p=0.32	73.8	p=0.71	85.1	p=0.0066	56.0	p=0.0024
Financial respondent		$\chi^2=4.38,$		$\chi^2=0.058,$		$\chi^2=5.46,$		$\chi^2=3.81,$
Yes	55.4	d.f.=1,	73.0	d.f.=1,	78.0	d.f.=1,	46.7	d.f.=1,
No	51.9	p=0.036	73.6	p=0.81	84.5	p=0.020	53.4	p=0.051
First wave in SHARE		$\chi^2=22.07,$		$\chi^2=14.78,$		$\chi^2=20.91,$		$\chi^2=42.83,$
W1	51.9	d.f.=4,	68.1	d.f.=4,	73.9	d.f.=4,	39.6	d.f.=4,
W2	56.5	p=0.0002	75.2	p=0.0052	90.2	p=0.0003	68.0	p<.0001

W4	50.9		78.4		77.4		45.5	
W5	55.0		68.7		77.2		42.0	
W6	63.2		77.2		84.0		53.6	
Ever did a COVID-19 interview		$\chi^2=13.43,$		$\chi^2=36.38,$		$\chi^2=4.05,$		$\chi^2=0.21,$
Yes	55.8	d.f.=1,	76.2	d.f.=1,	80.7	d.f.=1,	48.8	d.f.=1,
No	48.9	p=0.0002	57.4	p<.0001	73.4	p=0.044	46.8	p=0.65
n	4,345		1,467		1,073		1,073	

Appendix Table B2: Bivariate Associations with Participation at Selected Stages: Selected Substantive Variables

Variable	Consent		Shipped, given consent and sampled		Wear 1+ days, given shipped		Wear 8+ days, given shipped	
	%	Test statistic	%	Test statistic	%	Test statistic	%	Test statistic
Overall rates	54.4		73.1		79.8		48.6	
Vigorous activity		$\chi^2(3)=84.33,$ p<.0001		$\chi^2(3)=0.53,$ p=0.91		$\chi^2(3)=4.87,$ p=0.18		$\chi^2(3)=6.78,$ p=0.079
Hardly ever or never	47.6		74.0		78.2		45.5	
1-3 times a month	56.0		71.6		87.5		59.4	
Once a week	59.3		72.3		81.8		51.3	
More than once a week	63.2		72.7		79.0		48.6	
Moderate activity		$\chi^2(3)=180.29,$ p<.0001		$\chi^2(3)=1.70,$ p=0.64		$\chi^2(3)=5.71,$ p=0.13		$\chi^2(3)=21.46,$ p<.0001
Hardly ever or never	33.1		72.8		72.2		29.6	
1-3 times a month	50.1		74.7		76.1		42.3	
Once a week	49.7		76.6		81.9		48.7	
More than once a week	61.0		72.3		80.9		52.1	
Difficulty walking 100m		$\chi^2(1)=54.21,$ p<.0001		$\chi^2(1)=0.34,$ p=0.56		$\chi^2(1)=4.68,$ p=0.031		$\chi^2(1)=0.86,$ p=0.35
Yes	40.4		71.3		72.3		44.5	
No	56.6		73.4		80.7		49.1	
Difficulty climbing stairs		$\chi^2(1)=43.28,$ p<.0001		$\chi^2(1)=0.064,$ p=0.80		$\chi^2(1)=1.48,$ p=0.22		$\chi^2(1)=0.020,$ p=0.89
Yes	47.1		73.6		77.5		48.9	
No	57.8		73.0		80.7		48.4	
Physical limitations		$\chi^2(1)=10.62,$ p=0.0011		$\chi^2(1)=2.85,$ p=0.091		$\chi^2(1)=0.36,$ p=0.55		$\chi^2(1)=1.37,$ p=0.24
Not limited	56.9		71.2		80.5		50.4	
Limited	52.0		75.1		79.1		46.8	
Activities of daily living limitations		$\chi^2(1)=44.90,$ p<.0001		$\chi^2(1)=1.04,$ p=0.31		$\chi^2(1)=7.14,$ p=0.0076		v3.92, p=0.048
None	56.5		73.6		81.0		49.7	
One or more	42.4		70.1		71.0		40.5	
BMI		$\chi^2(3)=8.18,$ p=0.042		$\chi^2(3)=9.66,$ p=0.022		$\chi^2(3)=0.87,$ p=0.83		$\chi^2(3)=3.55,$ p=0.31
Underweight	37.5		58.8		80.0		40.0	
Normal	54.1		70.3		78.2		46.7	
Overweight	55.4		72.7		80.7		52.0	
Obese	54.1		78.8		80.5		46.0	

Any chronic conditions		$\chi^2(1)=0.13,$ $p=0.71$		$\chi^2(1)=7.86,$ $p=0.0051$		$\chi^2(1)=0.21,$ $p=0.65$		$\chi^2(1)=0.57,$ $p=0.45$
Yes	55.0		74.6		80.0		49.1	
No	54.3		66.2		78.5		45.9	
Depressed		$\chi^2(1)=2.79,$ $p=0.095$		$\chi^2(1)=1.65,$ $p=0.20$		$\chi^2(1)=2.23,$ $p=0.14$		$\chi^2(1)=4.38,$ $p=0.036$
Yes	52.4		70.8		76.8		43.3	
No	55.2		74.1		80.9		50.5	
Happy with life		$\chi^2(2)=27.62,$ $p<.0001$		$\chi^2(2)=0.37,$ $p=0.83$		$\chi^2(2)=1.27,$ $p=0.53$		$\chi^2(2)=0.81,$ $p=0.67$
Often	57.9		73.0		78.7		48.6	
Sometimes	50.9		72.8		81.7		49.7	
Rarely or never	47.6		75.2		80.0		44.9	
Self-rated health		$\chi^2(1)=54.86,$ $p<.0001$		$\chi^2(1)=2.88,$ $p=0.090$		$\chi^2(1)=0.45,$ $p=0.50$		$\chi^2(1)=0.37,$ $p=0.54$
Very good/excellent	64.4		69.8		81.2		50.2	
Good/fair/poor	51.3		74.3		79.3		48.0	
Feel full of energy		$\chi^2(3)=104.27,$ $p<.0001$		$\chi^2(3)=12.19,$ $p=0.0067$		$\chi^2(3)=0.032,$ $p=0.99$		$\chi^2(3)=2.55,$ $p=0.47$
Often	64.4		74.4		80.0		50.5	
Sometimes	50.9		70.5		79.6		49.1	
Rarely	48.4		79.4		79.6		44.8	
Never	37.1		60.7		79.4		41.2	
Smoking status		$\chi^2(2)=40.10,$ $p<.0001$		$\chi^2(2)=6.83,$ $p=0.033$		$\chi^2(2)=4.86,$ $p=0.088$		$\chi^2(2)=3.51,$ $p=0.17$
Current	55.8		79.3		75.0		48.9	
Former	61.0		69.9		83.1		52.5	
Never	50.3		73.4		79.3		46.1	
Drank alcohol in last 7 days		$\chi^2(1)=55.31,$ $p<.0001$		$\chi^2(1)=7.45,$ $p=0.0063$		$\chi^2(1)=1.04,$ $p=0.31$		$\chi^2(1)=0.033,$ $p=0.86$
Yes	59.6		70.4		78.6		48.3	
No	48.3		76.8		81.2		48.9	
Taking 5+ drugs a day		$\chi^2(1)=4.40,$ $p=0.036$		$\chi^2(1)=2.55,$ $p=0.11$		$\chi^2(1)=1.16,$ $p=0.28$		$\chi^2(1)=0.73,$ $p=0.39$
Yes	51.8		76.2		77.6		46.4	
No	55.4		72.0		80.6		49.4	
Troubled with pain		$\chi^2(1)=3.73,$ $p=0.053$		$\chi^2(1)=0.019,$ $p=0.89$		$\chi^2(1)=0.57,$ $p=0.45$		$\chi^2(1)=4.61,$ $p=0.032$
Yes	52.9		73.3		78.9		45.4	
No	55.8		73.0		80.7		51.9	
Fruit consumption		$\chi^2(2)=7.27,$ $p=0.026$		$\chi^2(2)=2.53,$ $p=0.28$		$\chi^2(2)=5.97,$ $p=0.051$		$\chi^2(2)=1.46,$ $p=0.48$
Every day	55.3		72.5		79.6		48.3	
2-6 times a week	52.8		77.2		83.6		51.4	

Twice a week or less	47.5		70.4		68.0		42.0	
Meat consumption		$\chi^2(2)=28.40,$ p<.0001		$\chi^2(2)=24.37,$ p<.0001		$\chi^2(2)=1.93,$ p=0.38		$\chi^2(2)=2.08,$ p=0.35
Every day	59.0		67.1		78.2		46.5	
2-6 times a week	53.8		73.8		81.6		50.9	
Twice a week or less	47.6		83.7		78.1		46.5	
Memory test score		$\chi^2(2)=8.18,$ p=0.017		$\chi^2(2)=18.97,$ p<.0001		$\chi^2(2)=1.04,$ p=0.59		$\chi^2(2)=0.73,$ p=0.69
Excellent or very good	55.7		66.0		77.7		46.0	
Good	55.2		78.0		80.9		49.5	
Fair or poor	51.1		70.3		79.3		48.6	
Numeracy score		$\chi^2(4)=68.25,$ p<.0001		$\chi^2(4)=10.93,$ p=0.027		$\chi^2(4)=10.33,$ p=0.035		$\chi^2(4)=13.73,$ p=0.0082
1 Bad	39.0		73.3		54.5		31.8	
2	42.1		73.1		78.9		37.4	
3	55.7		75.5		80.8		49.9	
4	58.1		75.1		79.0		49.3	
5 Good	58.7		65.2		83.1		55.8	
Self-rated reading skill		$\chi^2(3)=97.16,$ p<.0001		$\chi^2(3)=5.50,$ p=0.14		$\chi^2(3)=0.62,$ p=0.89		$\chi^2(3)=0.47,$ p=0.93
Excellent	63.2		70.6		78.8		49.0	
Very good	53.7		77.5		81.1		46.9	
Good	49.9		73.6		80.1		49.5	
Fair or poor	40.6		73.3		79.1		49.1	
Able to make ends meet		$\chi^2(3)=20.17,$ p=0.0002		$\chi^2(3)=13.74,$ p=0.0033		$\chi^2(3)=8.51,$ p=0.037		$\chi^2(3)=7.34,$ p=0.062
Great difficulty	49.5		77.4		76.4		38.2	
Difficulty	51.3		75.7		84.6		53.3	
Fairly easily	52.6		76.9		81.6		50.4	
Easily	58.6		67.8		75.8		46.2	
Internet use in past week		$\chi^2(1)=183.59,$ p<.0001		$\chi^2(1)=0.62,$ p=0.43		$\chi^2(1)=0.13,$ p=0.72		$\chi^2(1)=0.41,$ p=0.52
Yes	62.4		72.5		80.1		49.2	
No	41.3		74.5		79.1		47.1	
n	4,345		1,467		1,073		1,073	

Appendix Table B3: Multivariable Models of Participation at Selected Stages: Reduced Models

Variable	Consent		Shipped, given consent and sampled		Wear 1+ days, given shipped		Wear 8+ days, given shipped	
	AME	s.e.	AME	s.e.	AME	s.e.	AME	s.e.
Country								
Belgium	10.4	3.5	-15.7	4.9	2.1	5.2	0.9	6.0
Czechia	-11.3	3.3	9.6	3.0	3.4	5.1	25.7	6.1
Denmark	14.8	3.1	-41.0	5.2	-34.3	7.1	-17.6	5.7
France	5.7	3.0	-23.8	4.5	-14.1	5.6	-5.2	5.5
Germany	---	---	---	---	---	---	---	---
Italy	-3.8	3.6	-15.4	5.3	-0.4	5.6	-3.5	6.6
Poland	26.5	3.3	-6.4	4.6	*	*	52.5	4.8
Slovenia	18.5	3.1	0.1	3.9	7.5	4.8	34.5	5.8
Spain	-2.6	3.6	-18.7	5.2	-0.2	5.8	3.7	6.8
Sweden	17.6	3.1	-33.1	5.0	-5.3	6.3	4.1	6.3
Test statistic	$\chi^2(9)=244.2, p<0.001$		$\chi^2(9)=253.2, p<0.001$		$\chi^2(8)=41.7, p<0.001$		$\chi^2(9)=318.9, p<0.001$	
Age								
50-59	---	---	---	---	---	---	---	---
60-69	1.2	0.9	-3.5	3.5	15.5	5.0	8.2	4.8
70-69	-0.9	0.3	-3.3	4.3	9.0	6.1	3.6	5.7
80+	-4.3	3.3	-4.0	5.0	7.2	6.9	2.0	6.7
Test statistic	$\chi^2(3)=5.17, p=0.16$		$\chi^2(3)=1.0, p=0.80$		$\chi^2(3)=14.81, p=0.002$		$\chi^2(3)=4.47, p=0.21$	
Gender								
Male	---	---	---	---	---	---	---	---
Female	2.6	1.5	-1.9	2.4	-2.4	3.0	-1.3	3.1
Test statistic	$Z(1)=1.71, p=0.087$		$Z(1)=0.81, p=0.42$		$Z(1)=0.81, p=0.42$		$Z(1)=0.43, p=0.67$	
Income quartile								
Q1 (lowest)	---	---	---	---	---	---	---	---
Q2	-4.0	2.2	5.5	3.8	0.2	4.3	6.5	4.1
Q3	0.04	2.5	4.1	4.1	5.4	5.6	12.5	4.6
Q4 (highest)	0.8	2.8	4.4	4.5	3.1	5.3	13.5	5.3
Test statistic	$\chi^2(3)=6.21, p=0.10$		$\chi^2(3)=2.1, p=0.55$		$\chi^2(3)=2.25, p=0.52$		$\chi^2(3)=8.24, p=0.041$	
Education								

Primary	---	---	---	---	---	---	---	---
Lower secondary	0.1	2.6	-5.7	3.9	7.5	5.1	1.0	5.3
Upper secondary	2.9	2.4	-3.0	3.5	7.1	5.0	-0.007	5.0
Tertiary	5.1	2.8	-3.6	3.9	2.2	5.6	-0.26	5.5
	$\chi^2(3)=4.76, p=0.19$		$\chi^2(3)=2.2, p=0.53$		$\chi^2(3)=4.34, p=0.23$		$\chi^2(3)=0.08, p=0.99$	
Working status								
Retired	---	---	---	---	---	---	---	---
Working	0.9	2.6	-5.6	4.2	1.6	5.0	-6.4	5.2
Other	4.7	2.5	6.2	3.5	1.4	4.7	-0.69	5.1
	$\chi^2(2)=3.78, p=0.15$		$\chi^2(2)=8.53, p=0.014$		$\chi^2(2)=0.13, p=0.94$		$\chi^2(2)=1.72, p=0.42$	
Number of children								
None	---	---	---	---	---	---	---	---
One	4.7	3.0	1.0	4.8	11.8	5.9	11.0	6.2
Two	1.2	2.8	-0.9	4.4	8.2	5.7	3.7	5.8
Three or more	6.6	2.8	-1.8	4.5	7.8	5.8	3.8	5.9
	$\chi^2(3)=13.34, p=0.004$		$\chi^2(3)=0.87, p=0.83$		$\chi^2(3)=4.08, p=0.25$		$\chi^2(3)=4.89, p=0.18$	
Financial respondent (yes)	3.2	1.6	-0.3	2.5	-5.9	2.9	-7.3	3.2
	$Z(1)=2.04, p=0.042$		$Z(1)=0.11, p=0.9$		$Z(1)=1.92, p=0.06$		$Z(1)=2.24, p=0.025$	
Did a COVID-19 interview (yes)	10.3	1.9	9.2	3.2	-0.1	3.9	-6.9	4.4
	$Z(1)=5.49, p<0.001$		$Z(1)=3.04, p=0.002$		$Z(1)=0.04, p=0.97$		$Z(1)=1.58, p=0.11$	
Moderate activity								
Hardly ever or never	---	---	---	---	---	---	---	---
1-3 times a month	9.9	3.4	2.4	5.4	1.5	6.9	8.6	7.0
Once a week	8.6	2.9	3.1	4.5	6.1	5.7	14.4	5.8
More than once a week	15.6	2.4	0.4	3.9	6.0	4.9	17.4	4.9
	$\chi^2(3)=46.73, p<0.001$		$\chi^2(3)=0.97, p=0.81$		$\chi^2(3)=1.95, p=0.58$		$\chi^2(3)=13.32, p=0.004$	
Self-rated health								
Very good/excellent	3.1	1.9	4.7	2.7	6.3	3.3	5.6	3.8
Good/fair/poor	---	---	---	---	---	---	---	---
	$Z(1)=1.63, p=0.10$		$Z(1)=1.69, p=0.09$		$Z(1)=1.79, p=0.074$		$Z(1)=1.50, p=0.13$	
Feel full of energy								
Often	9.3	4.0	16.9	6.9	-6.5	7.0	-4.2	9.1
Sometimes	2.0	3.8	10.8	6.7	-6.7	6.7	-3.7	8.9
Rarely	3.3	3.9	18.6	6.8	-1.2	6.9	-3.8	9.0

Never	---	---	---	---	---	---	---	---
	$\chi^2(3)=19.61, p<0.001$		$\chi^2(3)=14.02, p=.003$		$\chi^2(3)=3.33, p=0.34$		$\chi^2(3)=0.21, p=0.98$	
Smoking status								
Current	---	---	---	---	---	---	---	---
Former	1.7	2.3	-2.8	3.6	12.0	4.7	3.5	4.6
Never	-3.1	2.1	-3.3	3.4	8.9	4.5	1.4	4.2
	$\chi^2(2)=8.88, p=0.012$		$\chi^2(2)=0.95, p=0.62$		$\chi^2(2)=6.44, p=0.040$		$\chi^2(2)=0.72, p=0.70$	
Drank alcohol in last 7 days (yes)	6.0	1.6	-0.2	2.5	0.3	3.0	2.6	3.2
	$Z(1)=3.86, p<0.001$		$Z(1)=0.07, p=0.95$		$Z(1)=0.09, p=0.93$		$Z(1)=0.83, p=0.41$	
Taking 5+ drugs a day (yes)	5.7	1.7	4.4	2.6	-3.2	3.3	-5.0	3.5
	$Z(1)=3.28, p=0.001$		$Z(1)=1.65, p=0.10$		$Z(1)=0.99, p=0.32$		$Z(1)=1.4, p=0.15$	
Memory test score								
Excellent or very good	-6.9	2.1	-6.5	3.5	-2.4	4.3	-4.7	4.5
Good	-2.1	1.7	2.9	2.6	0.2	3.2	-3.2	3.5
Fair or poor	---	---	---	---	---	---	---	---
	$\chi^2(2)=10.78, p=0.005$		$\chi^2(2)=9.67, p=.008$		$\chi^2(2)=0.51, p=0.78$		$\chi^2(2)=1.25, p=0.53$	
Numeracy score								
1 Bad	---	---	---	---	---	---	---	---
2	-4.7	4.7	0.9	8.3	18.6	11.4	2.1	10.6
3	2.5	4.6	3.3	8.1	18.5	11.3	13.8	10.3
4	0.7	4.7	2.2	8.2	16.6	11.5	13.0	10.4
5 Good	-3.0	4.9	-4.6	8.6	22.5	11.7	19.6	10.8
	$\chi^2(4)=14.03, p=0.007$		$\chi^2(4)=5.54, p=0.24$		$\chi^2(4)=5.10, p=0.28$		$\chi^2(4)=11.68, p=0.020$	
Self-rated reading skill								
Excellent	5.1	2.8	0.7	4.4	-3.3	5.1	-7.9	5.7
Very good	0.5	2.7	3.4	4.3	0.3	4.9	-7.6	5.6
Good	0.8	2.6	-4.6	4.1	-3.7	4.8	-8.3	5.4
Fair or poor	---	---	---	---	---	---	---	---
	$\chi^2(3)=7.61, p=0.052$		$\chi^2(3)=6.62, p=0.08$		$\chi^2(3)=1.42, p=0.70$		$\chi^2(3)=2.48, p=0.48$	
Internet use in past week (yes)	10.4	1.9	5.8	3.2	5.5	3.7	7.5	3.7
	$Z(1)=5.44, p<0.001$		$Z(1)=1.85, p=0.065$		$Z(1)=1.53, p=0.13$		$Z(1)=2.03, p=0.043$	

Appendix Table C1: Selection Bias for Selected Demographic Variables

Variable	SHARE W8 respondents (%)	Consented Bias (s.e.)	Device shipped Bias (s.e.)	Wear 1+ days Bias (s.e.)	Wear 8+ days Bias (s.e.)
Country					
Belgium	6.70	0.92 (0.29)	2.53 (0.64)	2.77 (0.78)	1.36 (1.08)
Czechia	10.66	-4.05 (0.64)	1.46 (0.81)	1.61 (0.95)	3.55 (1.35)
Denmark	11.08	2.90 (0.29)	-4.83 (0.81)	-6.86 (0.82)	-7.62 (0.92)
France	12.04	0.15 (0.00)	-0.95 (0.86)	-2.81 (0.97)	-3.98 (1.20)
Germany	12.93	-0.87 (0.51)	0.39 (0.89)	0.62 (1.03)	-1.42 (0.34)
Italy	8.10	-2.77 (0.55)	0.00 (0.72)	-0.27 (0.83)	-3.30 (0.97)
Poland	8.12	2.33 (0.24)	3.90 (0.69)	6.95 (0.82)	12.61 (1.26)
Slovenia	10.49	1.44 (0.35)	0.32 (0.81)	1.19 (0.95)	4.28 (1.35)
Spain	8.65	-2.94 (0.56)	0.11 (0.74)	-0.24 (0.86)	-1.94 (1.08)
Sweden	11.23	2.90 (0.30)	-2.94 (0.83)	-2.94 (0.93)	-3.55 (1.17)
Age					
50-59	19.49	1.62 (0.49)	1.48 (1.04)	0.02 (1.21)	-1.07 (1.61)
60-69	31.35	2.93 (0.56)	2.20 (1.21)	4.87 (1.40)	7.04 (1.93)
70-69	30.68	-0.38 (0.65)	0.45 (1.22)	0.51 (1.41)	0.80 (1.90)
80+	18.48	-4.18 (0.69)	-4.13 (1.03)	-5.40 (1.15)	-6.77 (1.43)
Gender (Female)	58.25	-0.19 (0.68)	0.84 (1.30)	0.04 (1.51)	-1.82 (2.05)
Income quartile					
Q1 (lowest)	24.99	-2.35 (0.67)	2.78 (1.13)	4.21 (1.32)	5.91 (1.82)
Q2	24.99	-4.04 (0.72)	-0.56 (1.15)	1.28 (1.32)	-0.43 (1.77)
Q3	24.99	1.16 (0.57)	2.62 (1.15)	0.24 (1.33)	-1.39 (1.76)
Q4 (highest)	25.02	-5.24 (0.44)	-2.47 (1.16)	-3.17 (1.32)	-4.10 (1.72)
Education					
Primary	19.77	-5.13 (0.73)	-2.81 (1.06)	-4.35 (1.20)	-6.14 (1.50)
Lower secondary	15.63	-2.13 (0.59)	0.59 (0.96)	0.73 (1.11)	0.66 (1.48)
Upper secondary	39.79	2.53 (0.63)	1.77 (1.28)	3.90 (1.45)	6.27 (2.00)
Tertiary	24.81	4.73 (0.45)	0.45 (1.14)	0.28 (1.32)	0.53 (1.78)
Tenure (Owned)	73.76	0.72 (0.61)	0.42 (1.16)	1.59 (1.32)	3.78 (1.72)
Household type					

Single	29.67	-1.95 (0.69)	0.53 (1.21)	-2.56 (1.44)	-4.71 (1.83)
Couple	60.39	1.90 (0.65)	1.96 (1.27)	5.03 (1.40)	6.98 (1.91)
Other	9.94	0.04 (0.41)	-2.49 (0.79)	-2.47 (0.88)	-2.26 (1.15)
Working status					
Retired	67.13	-1.84 (0.67)	-1.52 (1.26)	-0.55 (1.45)	2.92 (1.88)
Working	19.24	2.68 (0.45)	-1.16 (1.05)	-1.25 (1.20)	-2.54 (1.57)
Other	13.62	-0.84 (0.51)	2.68 (0.89)	1.80 (1.06)	-0.38 (1.40)
Marital status					
Married, civil union	68.47	1.57 (0.63)	-0.44 (1.24)	2.56 (1.38)	4.66 (1.82)
Single	5.71	0.05 (0.32)	0.26 (0.61)	-0.8 (0.70)	-0.91 (0.91)
Separated, divorced	9.34	0.56 (0.37)	1.09 (0.77)	-0.12 (0.89)	-1.86 (1.13)
Widowed	16.48	-2.17 (0.60)	-0.91 (0.99)	-1.64 (1.13)	-1.89 (1.48)
Number of children					
None	7.92	-0.85 (0.43)	-0.65 (0.72)	-1.84 (0.80)	-2.73 (0.98)
One	17.31	0.38 (0.51)	0.77 (1.00)	1.50 (1.16)	2.65 (1.60)
Two	41.59	-1.17 (0.71)	-0.21 (1.31)	0.35 (1.51)	0.64 (2.03)
Three or more	33.19	1.64 (0.61)	0.08 (1.25)	-0.01 (1.44)	-0.56 (1.93)
Average (absolute) bias and average s.e.		1.93 (0.54)	1.37 (1.02)	2.02 (1.17)	3.29 (1.54)

Appendix Table C2: Selection Bias for Selected Substantive Variables

Variable	SHARE W8 respondents (%)	Consented Bias (s.e.)	Device shipped Bias (s.e.)	Wear 1+ days Bias (s.e.)	Wear 8+ days Bias (s.e.)
Vigorous activity					
Hardly ever or never	49.14	-6.10 (0.79)	-3.47 (1.35)	-4.39 (1.56)	-6.33 (2.07)
1-3 times a month	9.37	0.28 (0.39)	0.42 (0.77)	0.45 (0.90)	1.57 (1.24)
Once a week	13.03	1.17 (0.41)	1.35 (0.88)	1.72 (1.04)	2.16 (1.43)
More than once a week	28.49	4.64 (0.50)	2.54 (1.18)	2.23 (1.38)	2.60 (1.87)
Moderate activity					
Hardly ever or never	15.70	-6.13 (0.75)	-4.98 (0.96)	-6.00 (1.05)	-9.17 (1.18)
1-3 times a month	6.65	-0.52 (0.38)	0.03 (0.66)	-0.34 (0.76)	0.89 (0.98)
Once a week	13.46	-1.15 (0.53)	1.45 (0.90)	1.84 (1.05)	1.51 (1.44)
More than once a week	64.19	7.80 (0.57)	3.57 (1.23)	4.50 (1.41)	8.56 (1.82)
Difficulty walking 100m (yes)	13.67	-3.51 (0.63)	-2.29 (0.76)	-2.58 (0.91)	-3.62 (1.02)
Difficulty climbing stairs (yes)	31.81	-4.26 (0.75)	-2.63 (1.02)	-2.50 (1.25)	-3.30 (1.43)
Any physical limitations (GALI) (yes)	51.69	-2.26 (0.73)	-0.53 (1.32)	-1.00 (1.53)	-2.36 (2.07)
ADL limitations (ADL2_X) (yes)	15.10	-3.33 (0.48)	-2.89 (0.95)	-4.23 (1.06)	-4.93 (1.33)
BMI (BMI2_X)					
Underweight	1.47	-0.46 (0.24)	0.54 (0.31)	-0.54 (0.34)	-0.71 (0.40)
Normal	36.00	-0.15 (0.67)	2.26 (1.28)	-2.93 (1.48)	-3.56 (1.96)
Overweight	38.99	0.75 (0.66)	0.99 (1.29)	1.43 (1.49)	3.81 (2.00)
Obese	23.54	-1.42 (0.59)	1.81 (1.11)	2.04 (1.30)	0.45 (1.75)
Any chronic conditions CHRONIC1PLUS) (yes)	81.45	-0.20 (0.54)	2.52 (0.98)	2.78 (1.13)	3.39 (1.48)
Depressed (EURODCAT_X) (yes)	27.87	-1.04 (0.65)	-0.56 (1.19)	-1.59 (1.37)	-3.49 (1.81)
Happy with life (yes)					

Less than very good health (SPHUS2_X) (yes)	76.32	-4.38 (0.62)	-0.64 (1.14)	-1.08 (1.32)	-1.47 (1.78)
Feel full of energy (ACO23_X)					
Often	33.83	6.24 (0.50)	4.94 (1.22)	5.07 (1.43)	6.48 (1.95)
Sometimes	43.75	-2.83 (0.74)	-4.42 (1.34)	-4.50 (1.54)	-4.02 (2.04)
Rarely	17.88	-1.97 (0.61)	0.85 (1.11)	0.81 (1.18)	-0.61 (1.56)
Never	4.53	-1.44 (0.41)	-1.37 (0.54)	-1.38 (0.60)	-1.85 (0.73)
Smoking status (SMOKER)					
Current	14.84	0.39 (0.48)	1.56 (0.93)	0.58 (1.09)	1.66 (1.49)
Former	30.47	3.72 (0.54)	0.94 (1.21)	2.24 (1.40)	3.50 (1.91)
Never	54.68	-4.11 (0.75)	-2.49 (1.34)	-2.81 (1.55)	-5.16 (2.08)
Drank alcohol in last 7 days (BR039_X) (yes)	53.67	5.15 (0.61)	1.32 (1.31)	0.53 (1.52)	1.03 (2.04)
Taking at least 5 different drugs a typical day (PH082_X) (yes)	27.41	-1.23 (0.66)	0.08 (1.18)	-0.66 (1.34)	-1.12 (1.82)
Troubled with pain (PH084_X) (yes)	49.46	-1.34 (0.72)	1.71 (1.31)	1.12 (1.52)	-1.67 (2.06)
Fruit consumption					
Every day	76.34	1.31 (0.58)	1.94 (1.09)	1.81 (1.27)	1.59 (1.71)
2-6 times a week	17.26	-0.50 (0.54)	-0.21 (1.02)	0.61 (1.16)	0.78 (1.57)
Twice a week or less	6.40	-0.81 (0.40)	-1.74 (0.64)	-2.43 (0.70)	-2.37 (0.88)
Meat consumption					
Every day	33.65	2.87 (0.59)	-0.75 (1.26)	-1.4 (1.45)	-2.17 (1.93)
2-6 times a week	47.34	-0.49 (0.70)	-0.28 (1.32)	0.79 (1.52)	1.99 (2.04)
Twice a week or less	19.01	-2.38 (0.63)	1.03 (1.02)	0.62 (1.20)	0.18 (1.62)
Memory test score					
Excellent or very good	23.77	1.02 (0.56)	-3.74 (1.14)	-4.27 (1.29)	-4.77 (1.68)
Good	47.85	0.69 (0.68)	4.71 (1.28)	5.42 (1.49)	5.7 (2.02)
Fair or poor	28.38	-1.72 (0.67)	-0.98 (1.2)	-1.16 (1.38)	-0.93 (1.84)
Numeracy score					
1 Bad	2.72	-0.77 (0.31)	-0.67 (0.43)	-1.31 (0.45)	-1.37 (0.54)
2	16.02	-3.62 (0.65)	-2.32 (0.98)	-2.47 (1.11)	-5.46 (1.36)

3	30.49	0.74 (0.62)	2.59 (1.20)	3.03 (1.40)	3.48 (1.91)
4	33.26	2.25 (0.60)	1.88 (1.24)	1.56 (1.44)	2.44 (1.95)
5 Good	17.51	1.4 (0.47)	-1.48 (1.01)	-0.81 (1.16)	0.91 (1.58)
Self-rated reading skill					
Excellent	34.78	5.64 (0.53)	1.20 (1.25)	0.74 (1.46)	1.50 (1.96)
Very good	25.55	-0.32 (0.62)	1.11 (1.15)	1.56 (1.33)	0.17 (1.79)
Good	27.66	-2.27 (0.68)	-0.54 (1.19)	-0.44 (1.37)	-0.02 (1.84)
Fair or poor	12.01	-3.04 (0.60)	-1.76 (0.86)	-1.85 (0.98)	-1.65 (1.29)
Able to make ends meet					
Great difficulty	7.34	-0.66 (0.40)	0.95 (0.69)	0.60 (0.80)	-0.82 (1.04)
Difficulty	21.47	-1.20 (0.61)	0.89 (1.08)	2.24 (1.26)	3.10 (1.72)
Fairly easily	33.16	-1.09 (0.68)	0.29 (1.25)	1.06 (1.44)	1.58 (1.94)
Easily	38.02	2.94 (0.61)	-2.14 (1.30)	-3.91 (1.50)	-3.86 (1.98)
Internet use in past week (yes)	61.96	9.14 (0.55)	5.42 (1.22)	5.68 (1.41)	6.37 (1.90)
Average (absolute) bias and average s.e.		2.43 (0.58)	1.70 (1.08)	1.99 (1.25)	2.52 (1.55)