



---

## **The contribution of paradata to panel cooperation in SHARE**

Johanna Bristle, Martina Celidoni, Chiara Dal Bianco, Guglielmo Weber

***Working Paper Series 19-2014***

---

SHARE • Survey of Health, Ageing and Retirement in Europe • [www.share-project.org](http://www.share-project.org)



# The contribution of paradata to panel cooperation in SHARE

Johanna Bristle

*Munich Center for the Economics of Aging (MEA)  
Max Planck Institute for Social Law and Social Policy*

Martina Celidoni

*University of Padua – Department of Economics and Management*

Chiara Dal Bianco

*Ca' Foscari University of Venice – Department of Economics*

Guglielmo Weber

*University of Padua – Department of Economics and Management*

## Abstract

This paper deals with panel cooperation in a cross-national face-to-face survey, the Survey of Health, Ageing and Retirement in Europe (SHARE). Our outcome of interest is panel cooperation in the fourth wave, conditional on participating in the previous wave. We focus on the contribution of *paradata*, related to fieldwork strategies, features of the (current) interviewer and variables describing respondents' prior interview experience.

Using a multilevel approach, we find that factors at all three levels (survey agency, interviewer and respondent) influence cooperation. At the highest level, we highlight the importance of everyday communication between survey agency coordinators and interviewers to gain cooperation. At the interviewer level we find that interviewers' quality of work and experience significantly affect cooperation propensity. Respondent-level characteristics represented in paradata, such as prior interview experience, have a large influence on the cooperation decision overall.

**Keywords:** panel data, attrition, paradata, field practices, interviewer effects

*Acknowledgement: We would like to thank Thorsten Kneip, Julie M. Korbmacher, Oliver Lipps, Omar Paccagnella, Lisa Riedel and Annette Scherpenzeel for valuable comments and suggestions. We received helpful comments at the Panel Survey Methods Workshop, Ann Arbor - MI (May 2014), the 4th SHARE User conference, Liege - BE (November 2013), the 5<sup>th</sup> European Survey Research Association (ESRA) conference, Ljubljana - SI (July 2013) and the weekly seminar at the Munich Center for the Economics of Aging. All remaining errors are our own.*

*This paper uses data from SHARE wave 4 release 1.1.1, as of March 28th 2013 (DOI: 10.6103/SHARE.w4.111) or SHARE wave 1 and 2 release 2.6.0, as of November 29 2013 (DOI: 10.6103/SHARE.w1.260 and 10.6103/SHARE.w2.260) or SHARELIFE release 1, as of November 24th 2010 (DOI: 10.6103/SHARE.w3.100). The SHARE data collection has been primarily funded by the European Commission through the 5th Framework Programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life), through the 6th Framework Programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5- CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Programme (SHARE-PREP, N° 211909, SHARE-LEAP, N° 227822 and SHARE M4, N° 261982). Additional funding from 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 and OGHA 04-064) and the German Ministry of Education and Research as well as from various national sources is gratefully acknowledged (see [www.share-project.org](http://www.share-project.org) for a full list of funding institutions)*

## 1. Introduction

Declining response rates are a well-known challenge in all survey-related fields. For longitudinal studies, nonresponse or attrition accumulates over waves and particularly harm the panel dimension of the data. Researchers use both ex-post approaches, e.g. imputation schemes and weighting, and ex-ante strategies, i.e. preventing missing information during fieldwork, to improve data quality, obtain more efficient estimates (De Leeuw et al., 2003) and account for the possibility that there are systematic differences between responding and non-responding individuals (Rubin, 1976). The ex-ante approach in particular investigates the determinants of survey response to reduce attrition by improving fieldwork.

Computerized survey tools allow researchers to exploit additional information about the survey process as determinants in the empirical analysis of nonresponse behaviour. This additional information is usually called *paradata* and can be defined as a “by-product of the data collection process capturing information *about that process*” (Durrant and Kreuter 2013, p.1). While paradata at the individual or interviewer level have been used in this strand of literature, information at the survey agency level, within a multi-country survey, has rarely been taken into account to explain participation. One possible motivation for this gap could be that, in cross-national research, information at the survey agency level may not be available or not harmonised across countries (Blom et al, 2008) so that comparability is limited. The Survey of Health, Ageing and Retirement in Europe (SHARE), which provides harmonised information on elderly individuals at the European level, collects such data in wave 4. This additional source of information gives us the opportunity to investigate the nature of nonresponse at the survey agency level, possibly due to differences between survey organisations regarding survey fieldwork strategies, because all those data are standardized and can be compared across countries.

Our approach is theoretically based on the framework of survey participation by Groves and Couper (1998), where factors expected to influence survey participation are divided into the two major areas: “out of researcher control” and “under researcher control”. In this paper we are particularly interested in factors that can be influenced by the researcher, namely survey agencies’ fieldwork strategies, features of the interviewer and the respondent-interviewer interaction.

Our outcome of interest is cooperation in the fourth wave of SHARE and we are interested in understanding which paradata contribute to explain it among: (1) previous experience with SHARE interviews, (2) current interviewer characteristics, and (3) agency-level fieldwork indicators.

The structure of the paper is as follows: section 2 reviews the literature, section 3 presents the features of the available data with a special focus on the variables derived. The heterogeneity in cooperation rates is described in section 4, section 5 presents the empirical strategy, section 6 comments the empirical results and section 7 concludes.

## 2. Previous findings

Panel studies are affected by attrition that can bias parameter estimates, due to potential differences between those who stay in the panel and those dropping out. Researchers have sometimes used ex-ante analysis to understand how to prevent missing information during fieldwork. The literature interested in the determinants of survey participation recently proposed also the use of paradata to better understand response behaviour (e.g. Kreuter et al., 2010 and Kreuter, 2013). However, even if paradata represent a rich source of new information, little attention has been paid for instance to indicators such as keystroke or item level paradata (Couper and Kreuter, 2013) as well as additional information at higher levels.

Regarding higher levels, e.g. country or survey agency, heterogeneity might be explained by differences in survey characteristics, in population characteristics or in data collection practices. This was highlighted by Blom (2012) by looking at country differences in contact rates in the European Social Survey (ESS) - a survey similar to SHARE in its attempt at ex-ante harmonization across several European countries, but unlike SHARE because it lacks the longitudinal dimension. By using counterfactual analysis the author attributes differences in contact rates to differential survey characteristics (mostly related to interviewers' contact strategies), population characteristics and coefficients. Like Blom (2012), we investigate what drives variability also at the country level, but we are interested in cooperation – rather than contact – and use multilevel analysis as empirical strategy. Multilevel analysis has become standard in this strand of literature, but is often used to explain the heterogeneity at the item, respondent or interviewer level. The exclusion of a country level is due to the restricted number of countries involved in other cross national surveys as well as the unavailability of harmonized information to compare. An exception is Lipps and Benson (2005), who analysed contact strategies in the first wave of SHARE using a multilevel model taking into account also the country level, but did not find significant determinants for country differences.

An advantage of using the fourth wave of SHARE, as we do, is that we can exploit additional information collected at the survey agency level to better understand if different fieldwork practices can explain the related variance. In SHARE, countries and survey agencies mostly overlap – however, since in two countries (Belgium and France) more than one survey agency collected the data we will use the term survey agency, instead of country, for the third (highest) level.

Taking the role of the interviewer into account is crucial for attrition analyses in face-to-face surveys. Following Lynn et al. (2011) we decided to focus on the current interviewer<sup>1</sup>. Results on interviewers' experience are consistent over different studies, leading to the conclusion that experience is positively associated with gaining cooperation (e.g. Lipps and Pollien, 2011, Jäckle et al., 2013, Hox and de Leeuw, 2002, Groves and Couper 1998), although it is still unclear what drives the effect, i.e. whether this is a selection effect (bad interviewers quit) or a learning effect (interviewers improve their skills over time). Durrant et al. (2010) showed that experience in terms of the skill level acquired matters more than time spent in the job.

---

<sup>1</sup> While Pickery et al. (2001) state that the previous interviewer is more relevant, a more recent study of Lynn et al. (2011) shows the opposite.

A more recent discussion is concerned with interviewers' work quality. As in Korbmacher and Schroeder (2013), paradata at the interviewer level can refer also to data quality indicators such as rounding behaviour in measurements. Based on the literature on "satisficing behaviour" in surveys (Krosnick, 1991), the underlying hypothesis is that interviewers who are diligent at specific tasks during the interview are more engaged and more successful in gaining cooperation than interviewers who show deviant interviewing behaviour. Diligent interviewers are those who fulfil their task thoroughly to optimize the quality of their interviews, whereas deviant interviewers use "satisficing strategies", such as skipping introductions or rounding measurements, to minimize effort.

Paradata at the respondent level are more common and can provide measures for the interviewer-respondent interaction and the overall experience during the previous interview. As indicators for past interview experience, we will focus on interviewer assessments on respondents' attitude, item nonresponse and interview length.

In general, a "pleasant" interview experience in a prior wave is assumed to be beneficial for future cooperation, but often this is not measured directly. Lugtig (2014) highlights four mechanisms of attrition at the respondent level, namely shocks (e.g. moving, health decline), habit (consistent participation pattern), absence of commitment, and panel fatigue. Paradata can especially help in capturing commitment and panel fatigue to single out respondents at risk of future attrition due to non-cooperation. This can be based on interviewer assessments, e.g. willingness to answer or if the respondent asked for clarifications, or directly derived from the interview data, e.g. item nonresponse. The latter in particular has been shown to be a good predictor for participation in later waves. According to the theory of a *latent cooperation continuum* (Burton et al., 1999), in fact, item nonresponse – not providing valid answers to some questions - can be considered a precursor of unit nonresponse – not providing any answers - in the following wave. In other terms, a steady progress of attrition is assumed and empirically supported (Loosveldt et al., 2002).

Interview length also contributes to shape past interview experience and can be computed using keystroke data. In longitudinal surveys, interview length might affect the decision to participate in later waves. On the one hand a longer interview can be seen as a burden and affect cooperation negatively; on the other hand length might also measure the respondent's motivation and commitment to the survey and therefore can have a positive influence on cooperation. The results in the literature on interview length are mixed with some showing a positive association (Fricker et al., 2012) with cooperation and some not finding any effects (Lynn, 2013, Sharp and Frankel, 1983). Brendan et al. (1995) bridge the mixed results of the other studies by taking the longitudinal perspective into account. They find that long interviews are positively correlated with cooperation during the first waves of a panel, but the association vanishes in later waves.

In this paper we provide an exploratory analysis to assess the role of respondent, interviewer and survey agency paradata in explaining cooperation in SHARE wave 4.

## 3. Data

### 3.1 SHARE and sample selection

SHARE is a multidisciplinary harmonised European survey, targeting individuals aged 50+ and their partners, and represents a principal source of data to describe and investigate causes and consequences of the aging process for the European population (see Börsch-Supan et al., 2013). SHARE has been conducted for the first time in 2004/2005 (wave 1) in eleven European countries (Austria, Belgium, Denmark, France, Germany, Greece, Italy, Netherlands, Spain, Sweden and Switzerland) and Israel. In the second wave Poland, Czech Republic and Ireland joined SHARE and additional refreshment samples were added in order to ensure representativeness of the targeted population. Wave 3, called SHARELIFE, conducted between 2008 and 2009, differs from the standard waves, since it collects life histories of individuals who participated in wave 1 or wave 2. The fourth wave of SHARE, which started in 2011, is a regular wave.

The regular main questionnaire is composed by about twenty modules, each focuses on a specific topic, e.g. demographics, mental and physical health, cognitive functions, employment and pensions. The questionnaire of SHARELIFE differs from the standard waves, since it has very few questions on the current condition<sup>2</sup> but focuses on gathering information regarding life histories of individuals who participated in wave 1 or wave 2. We exploit mainly the third and the fourth wave of SHARE, by looking at cooperation in wave 4 given participation in SHARELIFE. The two waves are not completely comparable given the rather special content of the third wave but the choice is driven mainly by the availability of paradata. The particular sample definition we refer to, implies that we have to be cautious when extending our results, as Lepkowski and Couper (2002) noticed in fact the response process in later waves of a panel differs in important ways compared to the initial wave (wave 1 in our case).

Both standard and retrospective SHARE interviews are conducted via face-to-face, computer-assisted personal interviews (CAPI). Not every eligible household member is asked to answer every module of the standard CAPI questionnaire: selected household members served as family, financial or household respondents. They answered questions about children and social support, financial issues or household features, on behalf of the couple or the household, respectively. This means that the length of the questionnaire varies among respondents by design, and this has to be taken into account when analysing participation. An advantage in using SHARELIFE is that differences among types of respondents are limited since there is a distinction only between first and second respondent on the basis of very few questions on household current economic situation (e.g. Household Income). In all SHARE waves there is also the possibility to have a shorter proxy interview for cognitively impaired respondents. The proxy can answer on behalf of the eligible individual for most of the modules.

---

<sup>2</sup> Variables related to the current condition are household income, health status, economic status and current income from employment, self-employment and pensions.

We describe more precisely our sample definition in Table 1. The number of individuals interviewed in SHARELIFE are 23,885<sup>3</sup>; we then deleted 282 cases that were not preloaded or not eligible for fieldwork wave 4, e.g. due to legal restrictions. We do not consider individuals from the longitudinal sample whose households were not contacted in wave 4 (579 cases) to avoid confusion between contact and cooperation since they are theoretically two separate stages (Lynn, 2008). When linking the different data sources, interviewer information was not linkable for France (2,451 cases), about 10% of Sweden and Czech Republic (393 cases), and some idiosyncratic cases (e.g. unresolvable typographical errors). Survey agency information was not available for Poland, which accounts for 1,839 cases. We do not have complete information on interviewers in wave 4 for 26 cases; wave 3 missing data concern 901 individuals, distributed among all countries<sup>4</sup>.

**Table 1. Sample definition**

Number of observations released in SHARELIFE	23,885
<i>Sample restrictions</i>	
not preloaded/not eligible in w4	282
household not contacted in w4	579
<i>Linkage restrictions</i>	
not linked with interviewer information	3,146
no survey agency information provided	1,839
<i>Incomplete data restrictions</i>	
Missing data at interviewer level	26
Missing data at respondent level	901
<b>Final sample size</b>	<b>17,112</b>

### 3.2 Collection and preparation of paradata in SHARE

The collection of paradata is greatly facilitated by using computer-assisted sample management tools and interview instruments. In the following section we describe the data sources in SHARE and the preparation of the variables we derive from them.

For sample management SHARE uses a tailor-made sample management system (SMS). This program is installed on each interviewer's laptop and enables interviewers to manage their assigned subsample. The success of a cross-national study such as SHARE heavily depends on the way the data is collected in the various countries. Therefore using a harmonised tool for collecting interview data as well as contact data is crucial to ensure the comparability of the results. The SMS tool enables interviewers to easily register every contact with a household or individual respondent and enter result codes for every contact attempt (e.g. no contact, contact-try again, or refusal). These

<sup>3</sup> We do not consider Greece because it did not participate in wave 4.

<sup>4</sup> As missing information is especially related to questions of the IV module regarding area and type of building, we ran our analysis including those observations by adding binary indicators for missing information. Results, available upon request, do not change.

data were also used by Lipps and Benson (2005) to analyse contact strategies in the first wave of SHARE. Among the information collected through the SMS tool, we used the number of contacts that interviewers registered before gaining household cooperation or a final refusal. Furthermore, the sample definition was partly constructed based on contact information (see Table 1).

While the interview is conducted, additional paradata are collected by tracking keystroke data. Here, every time a key is pressed on the keyboard of the laptop, this is registered and stored by the software in a text file. From these text files, time stamps at the item level can be computed. Additionally, the SMS records the number of times an item was accessed, back-ups, if a remark was set and the remark itself. We computed the interview length of wave 3 based on those files. In order to control for the potential effect of interview length on cooperation propensity, we include it and its square term to account also for possible non-linear effects. Controlling for interview length helps in taking into account that SHARE interviews vary by design due to the complex structure of the questionnaire. Also the fact that proxy interviews are allowed is supposed to be captured in part by the length of the interview and also by controlling for the health status of the respondent.

Furthermore, as paradata at the respondent level, we include information derived from the CAPI interviews in wave 3, and in particular the percentage of item nonresponse to monetary items. The questions considered to construct this variable are household income (HH017), value of the property (AC019), first monthly wage for employed individuals (RE021) or the first monthly work income (RE023) for self-employed individuals, the current wage if the respondent is still in employment (RE027), the current income if the respondent is still self-employed (RE029), pension benefit (RE036), the wage at the end of the main job if retired (RE041) and income at the end of the main job if retired and worked as self-employed (RE043). The public release of SHARE contains also a section where interviewers are asked to evaluate the reluctance of respondents (IV module). Related to this, we included also a dummy variable indicating whether the interviewer reports a high willingness to answer and whether she does not ask for clarifications. Furthermore, information on the area (urban vs. rural) is derived from the iv module.

Additionally, interviewer information and survey agency fieldwork strategies were gathered and delivered by the survey agencies for wave 4. Interviewer information includes demographics (year of birth, education, gender) and their previous experiences in conducting SHARE interviews. Interviewers' education level is not available for all countries. For those survey agencies that provided this information, we applied the International Standard Classification of Education (ISCED-97) for harmonising the country-specific answers<sup>5</sup>.

Among interviewers' controls we add also a measure of *work quality*, following Korbmacher and Schroeder (2013). We try to capture interviewers' quality based on the grip strength test that SHARE proposes in every wave. The test consists of measuring respondents' grip strength twice for each hand. In the CAPI, interviewers are explicitly told to record a value between 0 and 100, without rounding numbers to multiples of 5 and 10. "Previous waves showed that multiples of 5 and

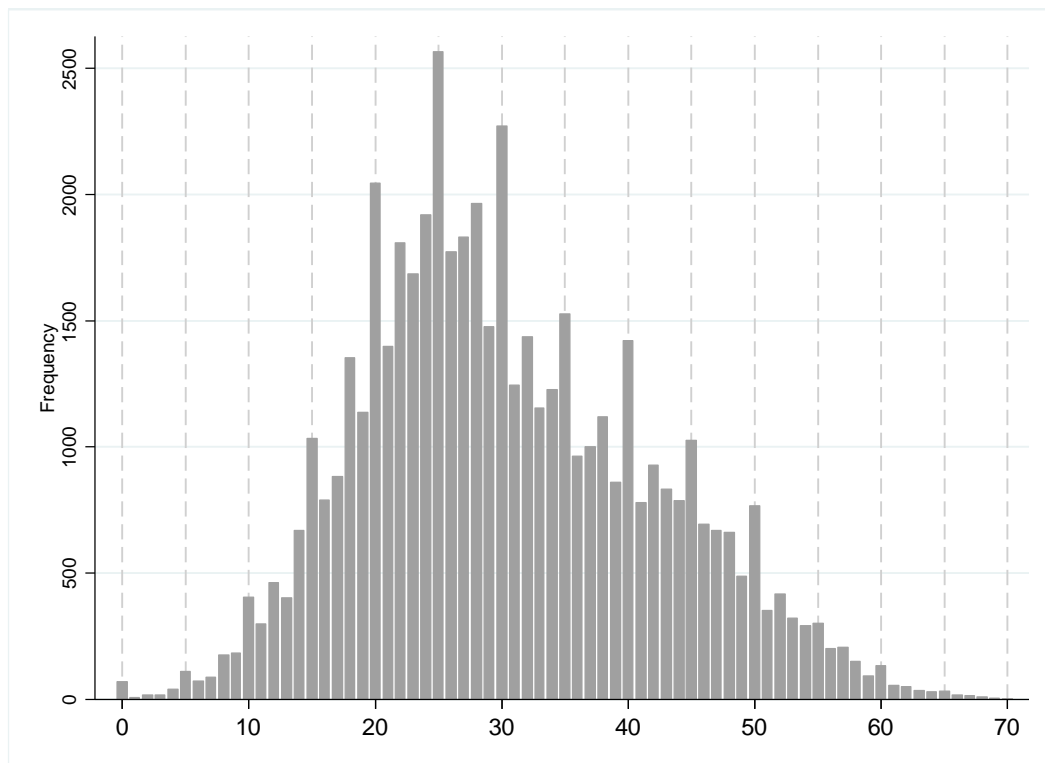
---

<sup>5</sup> We exploit this information to run robustness analysis with the subsample of agencies who provided the education information.



10 were recorded more than statistically expected” (Korbmacher and Schroeder, 2013); in Figure 1 we report the wave 4 pattern of grip strength measurement. If interviewers have percentages of multiples of 5 and 10 that lie outside a 90% confidence interval centred on the statistically expected value of 20.8%, then the interviewer is not measuring grip strength properly. We identify interviewers that do not overdo rounding with a dummy that takes value one if the percentage is below the lower cut-off of the confidence interval; another dummy captures individuals that round too often, taking value 1 if the percentage exceeds the upper cut-off of the confidence interval.

**Figure 1. Frequency of grip strength values**



Additional information is gathered at the survey agency level about fieldwork strategies, i.e. topics covered are recruitment, training, contacting respondents, translation, technical support, interview content, sampling process, management of interviewers and duration of fieldwork. Those data are collected mostly by means of open-ended questions, but some questions have a dropdown list. Open questions are difficult to handle within a multi-country framework. For this reason we focused on questions with standard answering options which show some variability. We will consider especially the following questions: ‘Who decides what project is prioritized assuming interviewers work on several projects simultaneously’ with ‘interviewer, agency or both’ as possible answers and ‘How often are you in contact with your interviewers about the SHARE study’ with the following answering options ‘Less than once a month, once a month, several times a month, once a week, several times a week or everyday’. We define two variables: *priority\_agency* which takes value one if the survey agency decides the priority of projects (four out of eleven survey agencies decide the priority of projects) and *everyday* which equals one if the survey agency has contact with the

interviewers on a daily basis (two out of eleven survey agencies have daily contacts with interviewers). An overview table of all variables used for the analysis with descriptive statistics can be found in the Appendix (Table A1-A3).

#### 4. Attrition in SHARE Wave 4

After describing the features of SHARE and the paradata used, we present more in detail response behaviour patterns in wave 4 for those who participated in SHARELIFE<sup>6</sup>, the sample we are interested in.

The standard distinction in the survey participation process is in terms of location, contact and cooperation (Lepkowski and Couper, 2002):

- a) location of the sample unit means to find geographically eligible individuals at a given address,
- b) contact means to reach an eligible sample unit by telephone or face-to-face visits, and
- c) cooperation is the completion of the interview.

Given that step a) usually is less problematic in a panel (Lepkowski and Couper, 2002) and we cannot test it, the final response rate will be the product of contact and cooperation rates, at least in simplified terms.

Kneip (2013) reports household contact rates for the panel sample of SHARE wave 4 that are consistently above 90 percent with an average of about 95 percent across all countries, while household cooperation, varying between about 60 percent and about 90 percent, shows greater variation across countries. Hence, retention rates, which combine contact and cooperation, varied between 56 percent and about 90 percent<sup>7</sup>.

This highlights that establishing contact was not an issue in the panel sample for most countries and non-contact seems to be a very limited phenomenon compared to other surveys such as ECHP, where Nicoletti and Peracchi (2005) analysed participation, modelling contact and cooperation as sequential events. In our case, the very limited number of individuals in non-contacted households (2.4%) does not induce us to adopt the same strategy of distinguishing the two sequential events. We therefore model the decision of cooperating, i.e. completing the interview, given that almost all households were contacted successfully.

Figure 2 presents the percentage, among contacted individuals, who do not cooperate in wave 4. The figure highlights some heterogeneity among survey agencies with rather low non-cooperation rates (about 12%) in Italy, Denmark and Switzerland and higher non-cooperation rates, around 20%, in Austria, Germany, Sweden and Spain<sup>8</sup>.

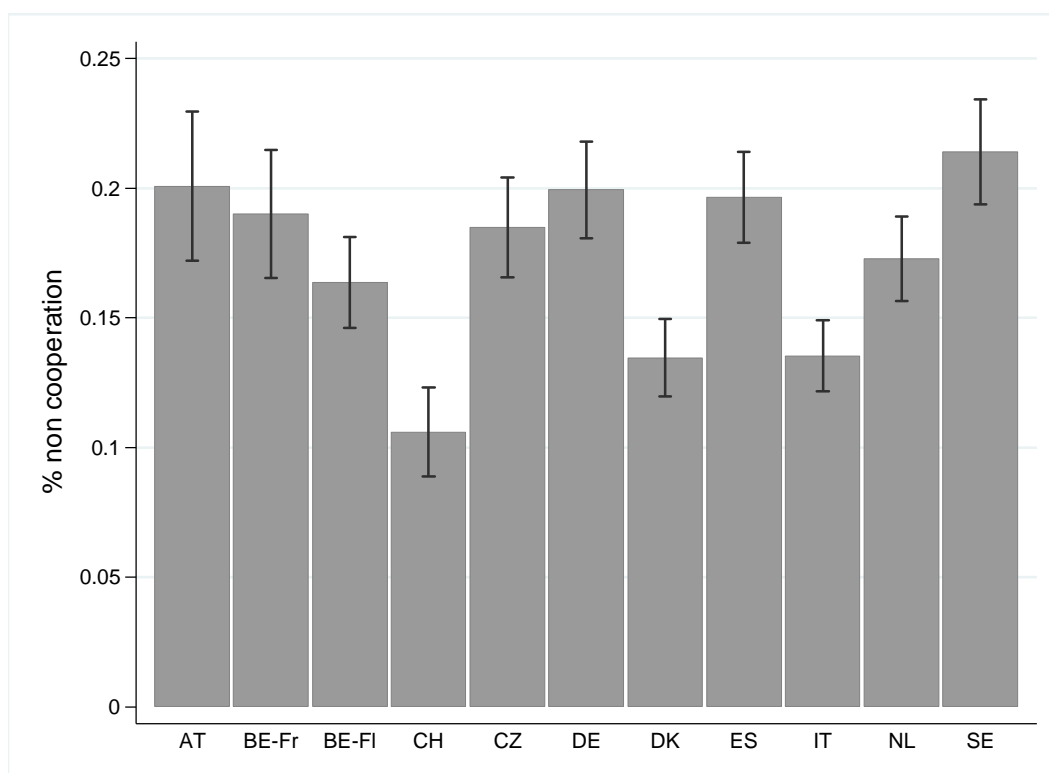
---

<sup>6</sup> Our sample of analysis slightly differs from the panel sample because we do not consider those interviewed in wave 1 or 2 who did not participate in wave 3 (SHARELIFE).

<sup>7</sup> All rates in Kneip (2013) are constructed according to AAPOR standards.

<sup>8</sup> These numbers are own calculations based on our sample restrictions. For official rates, please refer to Kneip (2013).

**Figure 2. Percentage of non-cooperation in W4 by survey agency**

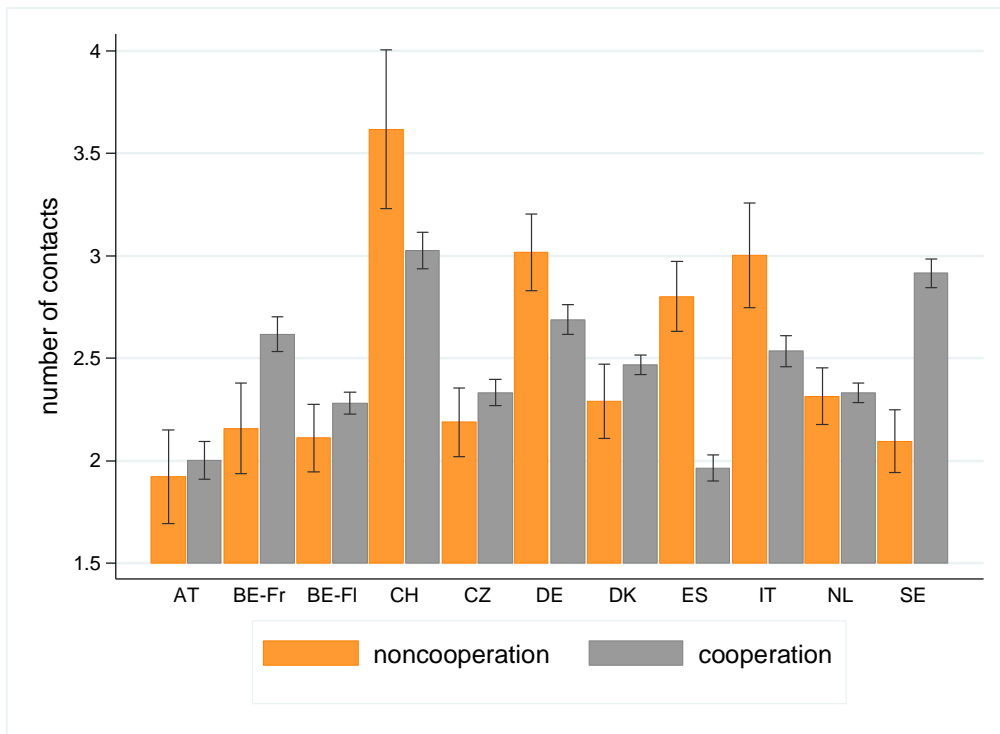


#### **4.1 Cooperation and selected respondent-level covariates**

In this subsection we describe the associations between our outcome of interest, *cooperation*, and selected covariates at the respondent level: the number of contacts, the length of the interview, and item nonresponse to monetary questions.

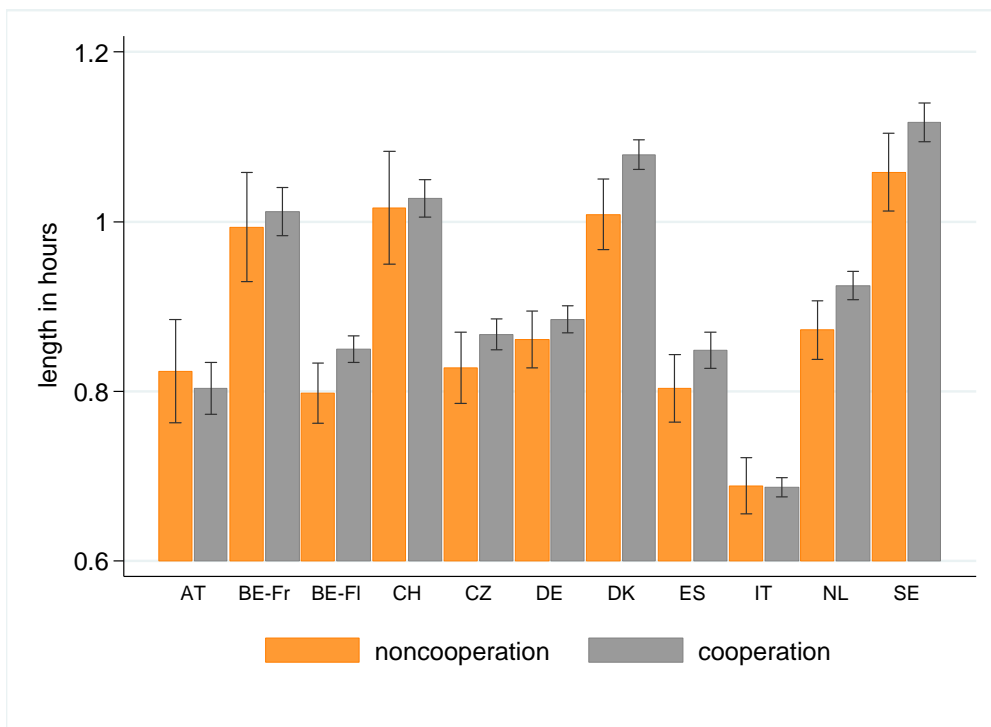
In Figure 3 we show the average number of contacts with households, meaning successful contact attempts that interviewers registered, by individual cooperation. We can see that in some cases, i.e. Austria, partly Belgium, Czech Republic, Denmark and the Netherlands, on average cooperating and non-cooperating individuals do not show statistically significant differences in terms of number of contacts; the same two groups instead show differences in Belgium - BE-Fr - , Switzerland, Germany, Spain, Italy and Sweden. But among the latter we have two opposite situations: in Belgium (French part) and Sweden for instance on average the number of registered contacts are higher for the cooperating individuals, whereas for Switzerland, Germany, Spain and Italy we can observe the opposite. The role of contacts in gaining cooperation is not clear *a priori*, therefore we decided to use two variables: the average number of contact by interviewer and the number of contact per household to capture two possible competing effects.

**Figure 3. Number of contacts by cooperation**



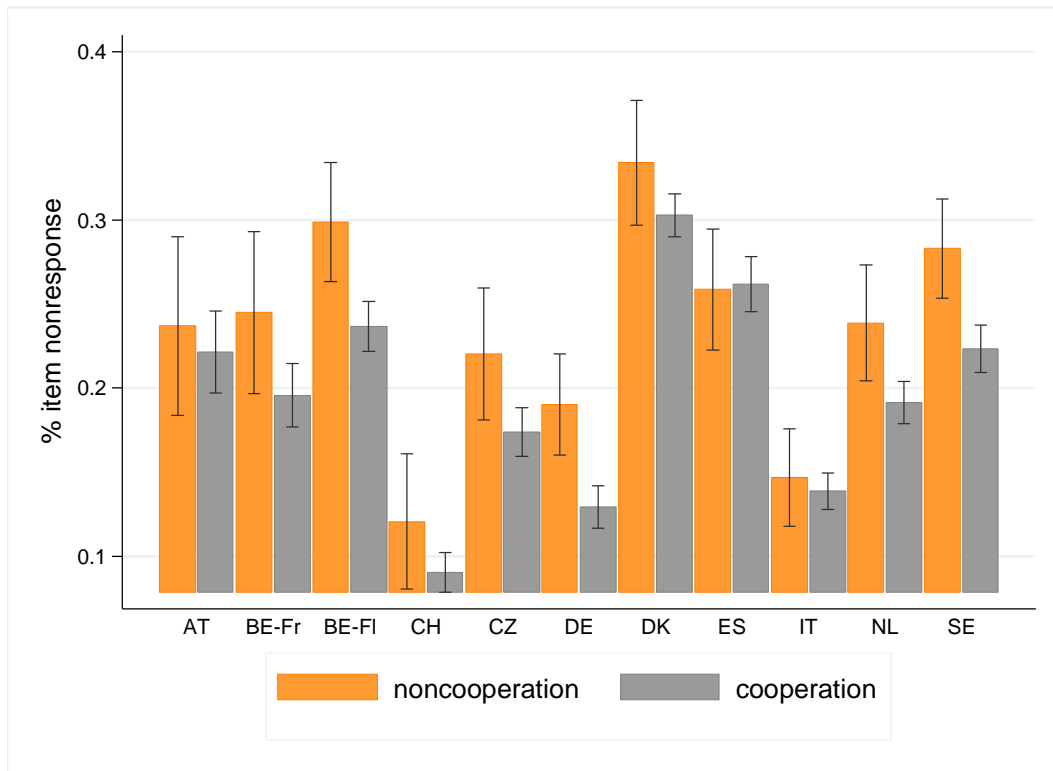
In Figure 4, we show the average interview length in wave 3 by country and cooperation outcome. We can see that only in some cases there is a statistically significant difference: In BE-FI (Flemish part), Denmark and the Netherlands interviews were longer for those cooperating in wave 4.

**Figure 4. Interview length by cooperation**



In Figure 5, it can be noticed that in four survey agencies, i.e. Belgium - BE-FI -, Germany, Netherland and Sweden, on average those who cooperate in wave 4 have a lower percentage of item nonresponse in wave 3 compared to non-cooperating individuals.

**Figure 5. Percentage of item nonresponse to monetary items, by cooperation**



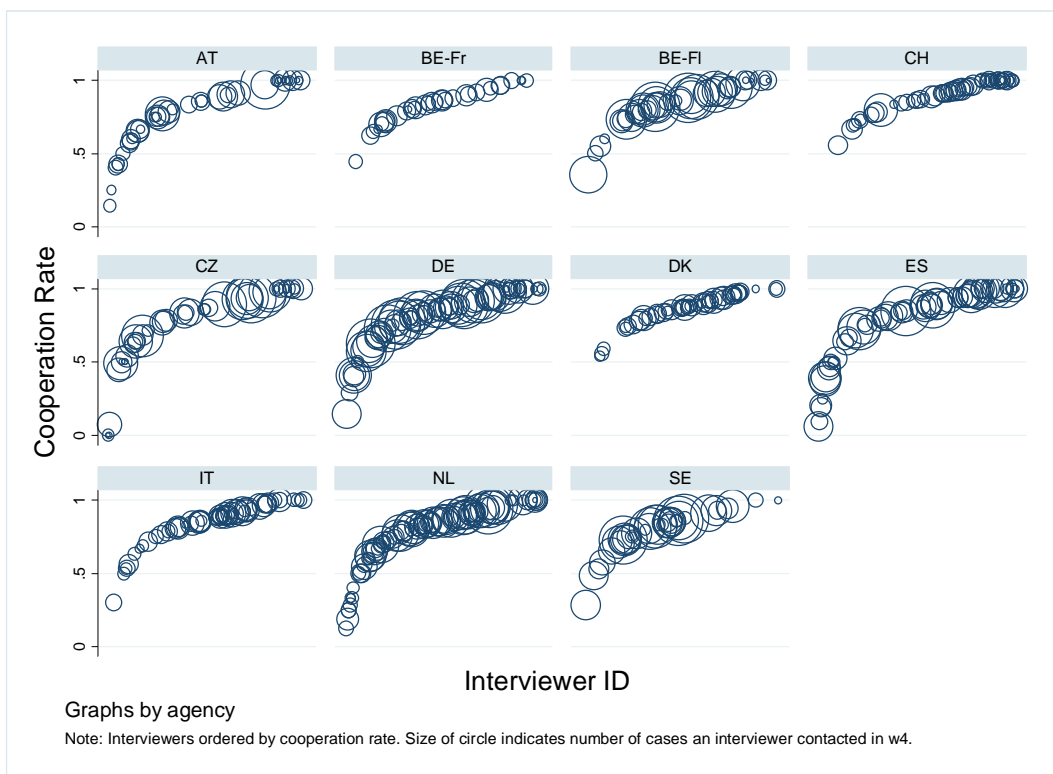
## 4.2 Interviewer effects in cooperation

We now look at cooperation rate heterogeneity among survey agencies, interviewers and their workload (Figure 6). We ordered interviewers, within each survey agency, according to their respective cooperation rate, reported on the vertical axis. One circle represents one interviewer. The size of the circles in the graph represents interviewers' workload for cooperation attempts in terms of number of cases the interviewer contacted in wave 4. From the figure, we can highlight the following aspects:

- There is variation in interviewer performance within countries: Interviewers differ substantially in their cooperation rates.
- The extent of variation between interviewers varies across countries: There are cases in which cooperation rates range from 0.5 to 1 and only very few interviewers show poor performance. For other survey agencies there is more variation between the interviewers according to their performances in terms of cooperation rates;

- The fieldwork strategy in terms of workload assigned varies across countries: Looking at the size of the circles, we can see that in some countries the workload is equally distributed among interviewers (for instance cases BE-Fr, CH, DK or IT), and other survey agencies where the workload varies (e.g. AT, DE, ES or SE); for those cases in which the number of cases is not distributed equally among interviewers, it is not clear whether survey agencies allocate the workload depending on the characteristics/ability of interviewers in gaining cooperation: if that was the case, we would expect larger circles on the right-hand side of the graph, but it could be argued that survey agencies allocate hard-to-convince cases to capable or adequate interviewers, so that there is no clear positive association between workload and interviewers' cooperation rates.

**Figure 6. Interviewer-specific cooperation rate by survey agency**



After having provided some evidence about raw associations of our variables of interest, we now turn to the description of the empirical strategy.

## 5. Empirical Strategy

We estimate a multilevel logit model<sup>9</sup> to account for correlation in cooperation probabilities for respondents interviewed by the same interviewer, and interviewers working for the same survey agency. This estimation strategy specifies the hierarchical structure of the data and allows us to avoid underestimation of standard errors and therefore incorrect inference (Couper and Kreuter, 2013). Given that we are interested in understanding how different levels contribute to explain cooperation, we start by estimating a random-intercept model (null model). We then enrich this baseline specification by stepwise inclusion of covariates at the individual, interviewer and survey agency level respectively. This bottom-up procedure has the advantage of keeping the model simple (Hox, 2010). Our outcome of interest is cooperation, denoted with  $y_{ijk}$ , which takes value 1 if the respondent  $i$ , interviewed by interviewer  $j$  of survey agency  $k$ , participates in wave 4 conditional on having participated in wave 3 and at least one earlier wave (either wave 1 or wave 2 or both).

The null model can be specified as follows

$$y_{ijk}^* = \beta_{jk} + \beta_k + e_{ijk} \quad \text{Level-1} \quad (1.1)$$

$$\beta_{jk} = \gamma_0 + u_{jk} \quad \text{Level-2} \quad (1.2)$$

$$\beta_k = \lambda_0 + v_k \quad \text{Level-3} \quad (1.3)$$

where the continuous latent response variable  $y_{ijk}^*$  is such that

$$y_{ijk} = \begin{cases} 1 & \text{if } y_{ijk}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

Equation 1.1 specifies the first level where  $e_{ijk}$  are level-1 residual terms, distributed as standard logistic,  $e_{ijk} \sim L(0; \sigma_e^2)$ ,  $\beta_{jk}$  are interviewer-specific intercepts and  $\beta_k$  are survey agency-specific intercepts. In a logit model the error variance at the first level is fixed,  $\sigma_e^2 = \pi^2/3$ , in order to fix the scale of  $y_{ijk}^*$  (Rabe-Hesketh and Skrondal, 2005). Thus in the multilevel extension no level-1 variance will be estimated.

Equation 1.2 and 1.3 specify the model for level-2 and level-3 specific intercepts. In equation 1.2 and 1.3,  $\gamma_0$  and  $\lambda_0$  are the mean intercepts whereas  $u_{jk}$  and  $v_k$  are deviations from the interviewer-specific and survey agency-specific intercepts, with  $u_{jk} \sim N(0; \sigma_u^2)$  and  $v_k \sim N(0; \sigma_v^2)$  respectively (Skrondal and Rabe-Hesketh, 2004). The third level is necessary, based on a standard likelihood ratio test criterion<sup>10</sup>.

---

<sup>9</sup> We estimate the multilevel logit model with the command of STATA *xtmelogit* and rescale parameters by means of the command *meresc* developed by Enzmann and Kohler (2012). Results are confirmed using *melogit* instead of *xtmelogit* with mean-variance adaptive Gauss–Hermite quadrature and mode-curvature adaptive Gauss–Hermite quadrature integration methods.

<sup>10</sup> The test statistic takes the value of 29.31 – under the null that level 3 is not required, i.e. the third level variance component is zero, it is asymptotically distributed as a mixture of a Chi-Squared with zero and a Chi-Squared with one degree of freedom (Self and Liang, 1987).

We then compare the models 1-4, where covariates on the three different levels are introduced in a stepwise procedure, to this null model to understand the role of each group of variables in reducing heterogeneity at different levels.

The first specification (Model 1) includes a set of controls for individual level socio-demographic characteristics,  $\mathbf{x}_{ijk}$ . Among these variables we include SHARELIFE information on demographics such as age, gender, years of education, household composition dummies - the reference category is the case in which there is a partner who has not been interviewed -, employment status, health status and controls for current household income, with dummy variables for the top three equivalent household income quartiles. We also control for residential information. A first dummy variable takes value one if the building is located in a urban area and a second dummy takes value one if the respondent lives in a detached or semi-detached house to control for the type of residential building.

In the second specification (Model 2) we add a set of paradata indicators ( $\mathbf{z}_{ijk}$ ) at the individual level: two dummy variables control for participation in the previous waves, in particular whether the individual was interviewed only in wave 1 (and 3), or in all the three previous waves, the reference category is therefore being interviewed only in wave 2. To account for the influence of previous interview duration, wave 3 interview length in hours and its square are added. At this stage, we include also the percentage of item nonresponse to monetary items, willingness to answer and whether the respondent did not ask for clarification. Finally, we also control for the number of contacts with the household (in wave 4) until an interview was conducted or a final refusal was received.

In the third specification (Model 3) we include controls at the interviewer level ( $\mathbf{s}_{jk}$ ). Especially interviewer's age and gender, interviewer's experience (a dummy which takes value one if the interviewer has already participated in at least one previous wave of SHARE) and the average number of contacts per household registered by the interviewer (before the interview or a final refusal). We include also interviewer quality indicators: a dummy which identifies interviewers who round least and another dummy for interviewers who round most. Finally, Model 4 controls for the survey agency level covariates ( $\mathbf{t}_k$ ) project prioritization and communication between interviewers and the survey agency.

The complete model, Model 4, can be specified as

$$y_{ijk}^* = \beta_{jk} + \beta_k + \boldsymbol{\beta}'_{1jk}\mathbf{x}_{ijk} + \boldsymbol{\beta}'_{2jk}\mathbf{z}_{ijk} + e_{ijk} \quad \text{Level-1} \quad (2.1)$$

$$\beta_{jk} = \gamma_{0k} + \boldsymbol{\gamma}'_{1k}\mathbf{s}_{jk} + u_{jk} \quad \text{Level-2} \quad (2.2)$$

$$\beta_k = \lambda_0 + \boldsymbol{\lambda}'_1\mathbf{t}_k + v_k \quad \text{Level-3} \quad (2.3)$$

where  $\mathbf{x}_{ijk}$  and  $\mathbf{z}_{ijk}$  vectors are individual level socio-demographic and paradata controls,  $\mathbf{s}_{jk}$  is a vector of interviewers' covariates and  $\mathbf{t}_k$  a vector of survey agency controls.

As already pointed out, in the logistic model the variance of the lowest-level residuals is fixed at a constant ( $\pi^2/3$ ). The main consequence is that in each of the models, the underlying scale is standardized to the same standard distribution, i.e. the residual variance cannot decrease when adding controls to the model. Moreover the value of the regression coefficients associated with the



included controls and the value of the higher-level variances are also rescaled. As a consequence it is not possible to compare the null model parameters with the following enriched specifications or to investigate how variance components change. Hox (2010) extends the rescaling procedure of Fielding (2004) to the multilevel setting and suggests the construction of scaling factors to be applied to parameters of the fixed part and random effects in order to make changes in these variables directly interpretable. In the case of a multilevel logistic regression model the scale correction factor is given by  $\sqrt{\sigma_0^2/\sigma_m^2}$  where the numerator is the total variance of the null model ( $\sigma_0^2 = \sigma_e^2 + \sigma_u^2 + \sigma_v^2$ ) and the denominator is the total variance of the model  $m$  ( $m=1,\dots,4$ ) including the first level predictor variables,  $\sigma_m^2 = \sigma_F^2 + \sigma_e^2 + \sigma_u^2 + \sigma_v^2$  with  $\sigma_F^2$  variance of the linear predictor of model  $m$  obtained using the coefficients of the predictors of the fixed part of the equation.

## 6. Results

### 6.1 Variance component analysis

Table 2 reports the results of various specifications of random-intercept models, without and with covariates, in terms of estimated variance components, intraclass correlations and model fit statistics<sup>11</sup>.

According to Davis & Scott (1995), the intraclass correlations at the interviewer and survey agency level are defined as:

$$\rho_{iwer} = \frac{\sigma_u^2}{\sigma_e^2 + \sigma_u^2 + \sigma_v^2} \quad \rho_{agency} = \frac{\sigma_v^2}{\sigma_e^2 + \sigma_u^2 + \sigma_v^2}$$

These definitions are in fact the most appropriate given that we are interested in how much variance is explained at each level.

Looking at the intraclass correlations we notice that survey agencies contribute about 1.8 % of the variation whereas interviewers account for about 24.7 %. The intraclass correlations in Table 2 suggest that most of the variation in cooperation, 73.5 %, is at the individual level. Table 2 reports also AIC values as a measure of goodness of fit for each successive specification. Reductions in the AIC show improvements in the model fit. Examining log-likelihoods yields similar conclusions, whereby the full model is to be preferred. The likelihood ratio test shows that adding respondent-level paradata improves the model strongly and reduces the scaled variance at the respondent level by 5 %.

---

<sup>11</sup> A similar approach can be found in a paper about interviewer effects on nonresponse in the European Social Survey (Blom et al. 2011). Although the approach is similar, we refrain from comparing the findings across SHARE and ESS here. Nonresponse processes can differ substantially between cross-sectional cooperation and cooperation in a later wave of a panel.

**Table 2: Estimated variance components, intraclass correlations and model fit statistics for different specifications of the multilevel models for cooperation.**

Model	0	1	2	3	4
	Intercept-Only	Respondent	R-Paradata	Interviewer	Agency
Variance components					
<i>Not scaled:</i>					
$\sigma_e^2$ (individual level)	3.29				
$\sigma_u^2$ (interviewer level)	1.104	1.106	1.122	0.939	0.944
$\sigma_v^2$ (agency level)	0.080	0.092	0.095	0.100	0.020
<i>Scaled:</i>					
$\sigma_e^2$ (individual level)		3.228	3.070	3.069	3.069
$\sigma_u^2$ (interviewer level)		1.085	1.047	0.876	0.881
$\sigma_v^2$ (agency level)		0.090	0.089	0.094	0.018
ICC_Iwer	0.247	0.246	0.249	0.217	0.222
ICC_Agency	0.018	0.020	0.021	0.023	0.005
Log likelihood	-7054.5606	-6992.9191	-6837.9853	-6811.1777	-6805.2108
LR-Test against previous model (degrees of freedom, p-value of LR-test)		123.28 (13; 0.000)	309.87 (8; 0.000)	53.62 (6; 0.000)	11.93 (2; 0.003)
Model fit statistic AIC	14115.12	14017.84	13723.97	13682.36	13674.42

Observations: 17112 Respondents, 643 interviewers, 11 agencies. ICC=Intraclass Correlation, AIC=Aikaike Information Criteria.

If we compare Model 2 and 3, where we introduce interviewers' characteristics, it can be seen that this set of interviewer level fixed effects account for a modest proportion of the variation at that level. Comparing the variance,  $\sigma_u^2$ , between models, it can be seen that about 16 % of the variation is captured by age, gender, experience, interviewers' average number of contacts and rounding behaviour. The likelihood ratio test reveals that adding interviewers' characteristics as predictors for cooperation results in a statistically significant improvement in model fit ( $p < 0.0001$ ).

Finally in Model 4, we add survey agency fieldwork strategies. The inclusion of survey agency related variables captures a large part of the variation at the third level, comparing  $\sigma_v^2$  between Model 3 and 4, we notice that we are able to explain about the 81 % of the variation. But we need to take into account that the variation at the survey agency level in total is rather small in comparison to the variance at the interviewer level. In fact, if we look at the intraclass correlation, the survey agencies contribute about 1.8 % of the variation whereas interviewers account for about 26.5 %. According to the likelihood ratio test, adding survey agencies' characteristics as predictors for cooperation improves the model fit ( $p = 0.0026$ ). We describe the effects of each set of covariates in the next subsection.

## 6.2 Predictors from multilevel analysis

We report in Table 3 the rescaled estimated coefficients for the stepwise specifications, where we add respondents', interviewers' and survey agencies' controls. The effects for each set of variables are described in the following subsections. We comment mainly on our preferred specification that is the complete specification reported in column 4.

**Table 3: Estimated multilevel models including respondents', interviewers' and agency's characteristics (dependent variable: cooperation).**

Model	0	1	2	3	4
	Intercept-Only	Respondent	R-Paradata	Interviewer	Agency
<b><i>Respondent characteristics</i></b>					
Female		0.068 (0.047)	0.076* (0.046)	0.075 (0.046)	0.077* (0.046)
Age		0.144*** (0.028)	0.100*** (0.028)	0.100*** (0.028)	0.101*** (0.028)
Age2		-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Being in poor health		-0.186*** (0.049)	-0.153*** (0.049)	-0.152*** (0.048)	-0.151*** (0.048)
HH composition – partner interviewed <i>(reference: partner without interview)</i>		0.374*** (0.075)	0.387*** (0.075)	0.383*** (0.075)	0.383*** (0.075)
HH composition – single <i>(reference: partner without interview)</i>		0.515*** (0.087)	0.498*** (0.087)	0.498*** (0.087)	0.497*** (0.087)
Years of Education		0.009 (0.006)	0.004 (0.006)	0.004 (0.006)	0.005 (0.006)
HH Income – 1st quartile		0.030 (0.066)	-0.012 (0.066)	-0.012 (0.066)	-0.007 (0.066)
HH Income – 2nd quartile		0.138** (0.067)	0.081 (0.066)	0.076 (0.066)	0.080 (0.066)
HH Income – 3rd quartile		0.156** (0.065)	0.110* (0.064)	0.108* (0.064)	0.110* (0.064)
Working		-0.171*** (0.065)	-0.147** (0.065)	-0.147** (0.064)	-0.146* (0.064)
Living in a urban area		0.008 (0.071)	-0.004 (0.070)	0.018 (0.070)	0.015 (0.070)
Living in a single house		0.264*** (0.056)	0.267*** (0.055)	0.264*** (0.055)	0.267*** (0.055)
<b><i>Paradata at the respondent level</i></b>					
Interviewed in w1 and w3 <i>(reference: interviewed in w2 and w3)</i>			-0.575*** (0.088)	-0.560*** (0.088)	-0.567*** (0.088)
Interviewed in w1, w2, w3 <i>(reference: interviewed in w2 and w3)</i>			0.937*** (0.080)	0.919*** (0.079)	0.920*** (0.079)
Item nonresponse in monetary questions			-0.496*** (0.085)	-0.503*** (0.084)	-0.505*** (0.084)
Length of interview (hours)			1.229*** (0.257)	1.203*** (0.256)	1.211*** (0.256)
Length of interview^2 (hours)			-0.390*** (0.107)	-0.382*** (0.106)	-0.386*** (0.106)
Willingness to answer			0.401*** (0.085)	0.403*** (0.085)	0.408*** (0.085)
Did not ask for clarification			0.252*** (0.064)	0.253*** (0.064)	0.256*** (0.064)
# of contacts with HH in wave 4			0.018 (0.018)	0.033* (0.018)	0.033* (0.018)

- Table 3 continued -

<b>Interviewers' characteristics (w4)</b>					
Age				-0.006 (0.005)	-0.005 (0.004)
Female				0.103 (0.099)	0.095 (0.098)
Experience with working on previous SHARE waves				0.591*** (0.106)	0.576*** (0.107)
Interviewer specific mean of contacts with HH until cooperation/refusal				-0.185*** (0.068)	-0.167** (0.069)
Rounding to a multiple of 5 at grip strength measure (too many)				-0.170* (0.098)	-0.191* (0.099)
Rounding to a multiple of 5 at grip strength measure (too few)				-0.699*** (0.229)	-0.689*** (0.228)
<b>Agency control variables</b>					
Priority decided by Survey Agency					0.152 (0.138)
Everyday contacts					0.611*** (0.168)
Constant	1.833*** (0.101)	-3.631*** (1.012)	-3.565*** (1.012)	-3.136*** (1.048)	-3.452*** (1.046)
$\sigma_e^2$ (individual level - scaled)	3.29	3.228	3.070	3.069	3.069
$\sigma_u^2$ (interviewer level - scaled)	1.104	1.085	1.047	0.876	0.881
$\sigma_v^2$ (agency level - scaled)	0.080	0.090	0.089	0.094	0.018
Deviance	14109.121	13985.838	13675.971	13622.355	13610.422
N	17112	17112	17112	17112	17112

Standard errors in parentheses (\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ )

Looking at Table 3, we can see that respondents' characteristics are highly predictive of cooperation in wave 4. While gender does not influence cooperation in wave 4, age instead is an important demographic control to consider particularly in a study about the older population as SHARE (Blom and Schröder 2010). According to our estimates, age has a non-linear effect on the probability of cooperation. Both regressors, *Age* and *Age2*, are statistically significant: up to about 69 years of age there is a positive association between participation and age, after that it becomes negative and the older the respondent the more likely is non-cooperation in the following wave – controlling for health conditions. The increase of participation with age for younger ages is probably explained by the increase in leisure associated to retirement (that typically occurs in the late fifties and early sixties). The decrease past age 69 may be related to the onset of cognitive decline: there is evidence that the magnitude of age-related cognitive decline accelerates at older ages (Salhouse, 2009), especially, according to a longitudinal analysis in Aartsen et al. (2002), cognitive decline might start after midlife, but most often occurs at higher ages (70 or higher). The same result was previously found in Schaie (1989): *most abilities tend to peak in early midlife, plateau until the late fifties or sixties, and then show decline, initially at a slow pace, but accelerating as the late seventies are reached*. Also SHARE data show a similar pattern for cognitive abilities (Celidoni et al, 2013). Therefore, even if there is still no consensus about when the rate of decline in abilities begins, there is evidence that after 60, more precisely around 70, the magnitude of age-related decline accelerates and this can affect cooperation in surveys too.

If the respondent reported *Being in poor health* in wave 3, this has a negative and statistically significant effect on the probability to cooperate. This is not surprising but at the same time inconvenient for a survey on health and ageing. To diminish the negative effects on attrition due to health problems, SHARE allows proxy interviews.

Compared to couples with non-responding partners, singles are more likely to participate in the next wave as well as individuals whose partner took part in the survey. Education, expressed here in terms of years of education, is not statistically significant; income quartiles instead, based on wave 3 information, are marginally significant. Social disparities do not seem to be particularly relevant for cooperation in an advanced panel wave, even if others studies have shown that higher educated people are more likely to stay in the panel (Lutgig, 2014, Watson and Wooden 2009). Compared to individuals in a non-working condition (retired, unemployed, sick or disables and housewives), workers have a lower probability to cooperate in the next wave, maybe due to time constraints. SHARE in fact has a rather long and detailed questionnaire that has the advantage of collecting a rich set of information, but as such it requires quite some time to answer all questions. This might create problems when respondents are still in employment and it has to be kept in mind when looking at statistics such as employment rates later in life for which attrition (or survey participation) could be an issue. Other factors, such as the characteristics of the area where the respondent lives and the building type might play a role in predicting cooperation. Living in an urban area in our case is not significant whereas living in a detached or semi-detached house increases the chances of cooperation in wave 4.

In addition to this standard set of respondents' characteristics, we use respondent-level paradata. Compared to participation in wave 2 (given that all respondents by design participated in wave 3), those who participated in both waves 1 and 2 (*Interviewed in w1, w2, w3*) are, as expected, more likely to participate also in wave 4, whereas individuals who participated only in wave 1 (*Interviewed in w1 and w3*) are less likely to cooperate in wave 4. In other words, respondents with an interrupted response pattern are less likely to participate again than respondents with continuous participation. We can observe also that a very good or good willingness to answer, and not having asked for clarifications during the interview in wave 3, are highly significant and associated with a higher probability of cooperation. As already explained earlier, we use the percentage of missing information in monetary amount questions as a predictor of cooperation in wave 4. This variable significantly affects the probability of cooperation. More precisely, the higher the percentage of missing information in SHARELIFE, the less likely is cooperation in wave 4. As paradata at the respondent level, we use also the length of the whole interview in wave 3. Both the length of the interview in hours and its square are highly statistically significant, showing a u-shaped effect, therefore the interview length has a positive association with cooperation up to a certain point, roughly 1.6 hours, when the probability of cooperating starts decreasing. We will show in the robustness section that using interview pace rather than length does not change our results. Among respondents' paradata we included also the number of contacts at the household level, before the first interview: the higher the number of contacts made by the interviewer, the more likely is cooperation in wave 4.

To explain the variation at the interviewer level, we add some socio-demographic controls, age and gender, a variable indicating interviewing experience in previous SHARE waves, the average number of contacts per interviewer and two dummies capturing interviewers' quality based on grip strength rounding behaviour. Age and gender do not significantly affect cooperation in wave 4, whereas experience does play a role, more precisely having experience with previous SHARE waves increases the likelihood of gaining respondents' cooperation. This is partly in line with previous findings by Jäckle et al (2013) on the effect of experience, measured in years working for the survey agency. Regarding the average number of contacts, it can be seen that an interviewer that on average registers many contacts is less likely to gain cooperation. This can be seen as a contextual effect as introduced in Skrondal and Rabe-Hesketh (2009), see also Paccagnella (2006) for a discussion about centering in multilevel models. It highlights the presence of two effects: a pure number of contacts effect which is positively related to cooperation and an interviewer-contact effect suggesting that an interviewer who tends to register several contacts might reduce or even cancel the positive effects of re-approaching the household. It can be also noticed that the two variables measuring interviewers' quality in terms of diligent interviewing behaviour are significant (the 'too often rounding' indicator is significant at 10%), with expected signs. If interviewers round grip strength scores to 5 or 10 too often or too few times, then gaining cooperation is less likely compared to cases in which the rounding percentage is as expected. This finding is in accordance to Korbmacher and Schroeder (2013) on consent to record linkage.

The final set of covariates in Table 3 is related to harmonised information collected at the survey agency level. We consider whether the priority of the projects is decided by the survey agency compared to situations in which interviewers can totally or partly choose how to organize their work. This can be seen as a variable capturing to what extent interviewers are autonomous and free to choose among several projects they are currently working on (e.g. working on SHARE or working on another survey on a specific day). The second variable captures the frequency of communication between survey agency and their interviewers. When including the priority variable, as the only third level covariate, we find that it influences the propensity to cooperate (results not shown here), but looking at Table 3, we can see that the decision about priority does not significantly affect cooperation when controlling for daily contact. Having daily contact with interviewers increases the chances to gain cooperation of respondents. This result hints at the importance of communication between survey agency coordinators and interviewers to successfully conduct surveys.

### 6.3 Robustness analysis

In this subsection we show that our results are robust to various changes in specification. The estimates in Tables 4 and 5 are not rescaled given that they are often obtained on a different subsample and thus not directly comparable with the original null model.<sup>12</sup>

In particular we test whether:

1. results change by redefining the *sample of analysis*. In column 1 of Table 4 the analysis is carried out at the household (rather than respondent) level and in column 2 we drop interviewers with less than ten interviews. With the first specification we consider the nested structure of SHARE interviews, where several respondents per household can be interviewed. With the second specification we want to address the potential inaccuracy of estimates when group sizes are small (Hox 2010).
2. results change when modifying the *set of controls*. In column 3 of Table 4 we use interview pace rather than interview length, in column 4 and 5 we include additional interviewer controls (summary statistics for these additional covariates are reported in Appendix B);
3. results regarding the effects of *survey agency practices* are robust. In column 1 of Table 5 we use as third level the Eurostat Nomenclature of Territorial Units for Statistics (NUTS1) instead of survey agency; in column 2 we estimate a two level model and control for group of countries.

**Table 4: Robustness analysis - Multilevel model estimates (dependent variable: cooperation).**

	1	2	3	4	5
	Household level	# Interviews >10	Pace	Short Introduction	Interviewer education
<i>Respondent characteristics</i>					
Female	0.131** (0.058)	0.073 (0.049)	0.088* (0.048)	0.079* (0.048)	0.043 (0.053)
Age	0.148*** (0.035)	0.108*** (0.029)	0.106*** (0.029)	0.105*** (0.029)	0.093*** (0.033)
Age2	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Being in poor health	-0.165*** (0.060)	-0.182*** (0.051)	-0.148*** (0.050)	-0.157*** (0.050)	-0.158*** (0.055)
HH composition – partner interviewed (reference: partner without interview)	0.252** (0.098)	0.391*** (0.080)	0.384*** (0.078)	0.396*** (0.078)	0.374*** (0.088)
HH composition – single (reference: partner without interview)	0.541*** (0.093)	0.508*** (0.093)	0.509*** (0.090)	0.514*** (0.090)	0.482*** (0.101)
Years of Education	-0.001 (0.007)	0.003 (0.006)	0.008 (0.006)	0.006 (0.006)	0.002 (0.007)
HH Income – 1st quartile	-0.016 (0.081)	-0.045 (0.070)	-0.012 (0.068)	-0.007 (0.068)	-0.029 (0.076)
HH Income – 2nd quartile	0.022 (0.081)	0.091 (0.070)	0.082 (0.068)	0.083 (0.068)	0.080 (0.076)
HH Income – 3rd quartile	0.060 (0.079)	0.103 (0.068)	0.112* (0.067)	0.113* (0.067)	0.111 (0.073)
Working	-0.068 (0.081)	-0.141** (0.068)	-0.166** (0.067)	-0.151** (0.067)	-0.216*** (0.074)
Living in a urban area	0.001 (0.082)	-0.014 (0.074)	0.019 (0.072)	0.015 (0.072)	0.048 (0.081)
Living in a single house	0.260*** (0.066)	0.290*** (0.058)	0.262*** (0.057)	0.277*** (0.057)	0.293*** (0.063)

<sup>12</sup> Rescaling do not substantially change the estimated parameters, scaled results are available upon request.

- Table 4 continued -

<b>Paradata at the (first) respondent level</b>					
Interviewed in w1 and w3 (reference: interviewed in w2 and w3)	-0.912*** (0.112)	-0.603*** (0.093)	-0.589*** (0.091)	-0.588*** (0.091)	1.335*** (0.289)
Interviewed in w1, w2, w3 (reference: interviewed in w2 and w3)	1.007*** (0.098)	0.939*** (0.085)	0.960*** (0.082)	0.953*** (0.082)	-0.463*** (0.121)
Item nonresponse in monetary questions	-0.597*** (0.106)	-0.528*** (0.089)	-0.519*** (0.087)	-0.522*** (0.087)	-0.534*** (0.096)
Length of interview (hours)	0.533*** (0.161)	1.374*** (0.270)		1.257*** (0.265)	1.335*** (0.289)
Length of interview <sup>2</sup> (hours)	-0.099*** (0.038)	-0.442*** (0.112)		-0.400*** (0.110)	-0.463*** (0.121)
Pace of interview			5.017*** (1.122)		
Pace of interview <sup>2</sup>			-7.414*** (2.003)		
Willingness to answer	0.375*** (0.108)	0.407*** (0.090)	0.449*** (0.088)	0.423*** (0.088)	0.338*** (0.094)
Did not ask for clarification	0.301*** (0.078)	0.246*** (0.068)	0.264*** (0.066)	0.264*** (0.066)	0.288*** (0.072)
# of contacts with HH in wave 4	0.109*** (0.024)	0.021 (0.019)	0.035* (0.019)	0.034* (0.0190)	0.033 (0.021)
<b>Interviewers' characteristics (w4)</b>					
Age	-0.001 (0.004)	-0.008 (0.005)	-0.004 (0.005)	-0.005 (0.005)	-0.006 (0.005)
Female	0.104 (0.100)	0.066 (0.105)	0.104 (0.102)	0.100 (0.102)	0.093 (0.118)
Interviewer education (iscsd 5-6)					-0.005 (0.131)
Short introductions				0.054 (0.122)	
Experience with working on previous SHARE waves	0.611*** (0.108)	0.747*** (0.115)	0.598*** (0.111)	0.591*** (0.111)	0.655*** (0.128)
Interviewer specific mean of contacts with HH until cooperation/refusal	-0.192*** (0.069)	-0.192** (0.075)	-0.171** (0.071)	-0.172** (0.071)	-0.179** (0.078)
Rounding to a multiple of 5 at grip strength measure (too many)	-0.241** (0.100)	-0.233** (0.105)	-0.204** (0.102)	-0.200* (0.102)	-0.152 (0.117)
Rounding to a multiple of 5 at grip strength measure (too few)	-0.728*** (0.237)	-1.100*** (0.304)	-0.713*** (0.236)	-0.714*** (0.236)	-0.747*** (0.259)
<b>Agency control variables</b>					
Priority decided by Survey Agency	0.150 (0.108)	0.189* (0.112)	0.157 (0.145)	0.156 (0.141)	-0.102 (0.209)
Everyday contacts	0.554*** (0.131)	0.564*** (0.136)	0.633*** (0.177)	0.628*** (0.173)	0.397 (0.252)
Constant	-4.854*** (1.311)	-3.479*** (1.094)	-3.604*** (1.085)	-3.618*** (1.087)	-3.007** (1.228)
$\sigma_u^2$ (interviewer level)	0.767	0.890	0.944	0.942	1.025
$\sigma_v^2$ (agency level)	<0.001	<0.001	0.021	0.019	0.027
Deviance	9598.726	12932.675	13621.792	13610.222	11140.992
N	12147	16422	17112	17112	13704

Standard errors in parentheses \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$



### *Sample of analysis*

The first column of Table 4 reports multilevel estimates at the household level (the sample size is now 12147 households) in which demographics refer to the first respondent. Interview length is defined as sum of the single interview lengths and this explains why it peaks at 2 hours and 40 minutes (expressing it in terms of single interview length for a first respondent that usually has a longer interview - 1.5 times the second respondent interview length - we obtain results similar to Table 3, cooperation probability increases up to 1 hour and 40 minutes and then decreases). All the other results do not change if the analysis is carried out at the household level. Comparing the second column of Table 4 with the fourth column of Table 3, we can see that results are robust to the exclusion of interviewers with a reduced number of interviews (lower than 10).<sup>13</sup>

### *Set of controls*

The third column of Table 4 presents the estimates when we include interview pace instead of interview length (similar to Korbmacher and Schroeder 2013, also see Loosveldt and Beullens 2013<sup>14</sup>). Interview pace, measured as length of interview divided by the number of items asked, is an alternative way of capturing the potential burden experienced by respondents in the previous wave or the motivation to respond by allowing for the time needed per question. In other words, here length is adjusted for differences in instrument length due to different types of respondents in SHARE. The results are consistent with those obtained using length.

Since part of the variability at the interviewer level remains unexplained, in the last two columns of Table 4 we include two additional controls, *short intro* and interviewer's education. The *short\_intro* variable should capture interviewers who are likely to skip section introductions that should be carefully read during the interview to ensure comparability across countries. Harmonization in SHARE is obtained also via standardised procedures in terms of questionnaire design and training. All interviewers are instructed, among several aspects, also to read carefully the whole CAPI question as it appears in the computer screen. However, beyond language differences, within countries there are interviewers that do not follow instructions; especially, when we compare keystroke data about section introductions, it is possible to notice that there are interviewers who skip them quickly. We use this information to construct an additional variable for the interviewer's quality. More precisely, we first compute the median reading time, by interviewer, for selected sections' introductions, that are relatively long, such as social network, activities, financial transfers and income from work and pension. If this value is lower than the country median in at least one case, then *short\_intro* takes value 1 and zero otherwise. The second additional variable is *Iwer education (isced 5-6)* a dummy that takes value 1 if the interviewer has tertiary education. This information is not available for the entire estimation sample, for 3408 observations we have missing data about interviewers' education, either because they did not provide such information to the survey agency or because the survey agency itself did not collect or provide it for all interviewers. Both variables are not significant and do not improve results. It should be noticed however that in

---

<sup>13</sup> Interviewers with less than 10 interviews are 126.

<sup>14</sup> Loosveldt and Beullens (2013) investigate interview speed measured as number of items per minute. They differentiate this concept from "speed of interviewing", which refers to the time needed per question. The latter concept is what we term interview pace in this paper.

the last column, when controlling for education, results at the survey agency level are no longer significant. This result is likely to be driven by the reduction in sample size. The variance that remains unexplained at the interviewer level could also capture regional differences that we are not able to disentangle from interviewer effects, an issue that is encountered also in other surveys, see for instance Couper and Kreuter (2013) who analyse the National Survey of Family Growth.

**Table 5: Robustness of survey agency practices effects - Multilevel model estimates (dependent variable: cooperation).**

	1	2
	NUTS 1 as third level	2-level model grouped countries
<b><i>Respondent characteristics</i></b>		
Female	0.136* (0.057)	0.081 (0.048)
Age	0.131*** (0.035)	0.107*** (0.029)
Age2	-0.001*** (0.000)	-0.001*** (0.000)
Being in poor health	-0.176** (0.060)	-0.160** (0.050)
HH composition – partner interviewed (reference: partner without interview)	0.362*** (0.094)	0.391*** (0.078)
HH composition – single (reference: partner without interview)	0.546*** (0.109)	0.515*** (0.090)
Years of Education	0.015* (0.007)	0.007 (0.006)
HH Income – 1st quartile	0.088 (0.083)	0.004 (0.068)
HH Income – 2nd quartile	0.009 (0.082)	0.087 (0.068)
HH Income – 3rd quartile	0.046 (0.079)	0.118 (0.066)
Working	-0.127 (0.080)	-0.149* (0.067)
Living in a urban area	0.063 (0.086)	-0.003 (0.071)
Living in a single house	0.239*** (0.067)	0.273*** (0.057)
<b><i>Paradata at the respondent level</i></b>		
Interviewed in w1 and w3 (reference: interviewed in w2 and w3)	-0.670*** (0.109)	-0.596*** (0.090)
Interviewed in w1, w2, w3 (reference: interviewed in w2 and w3)	1.069*** (0.100)	0.948*** (0.082)
Item nonresponse in monetary questions	-0.585*** (0.106)	-0.514*** (0.087)
Length of interview (hours)	1.247*** (0.318)	1.217*** (0.265)
Length of interview^2 (hours)	-0.416** (0.133)	-0.385*** (0.110)
Willingness to answer	0.464*** (0.104)	0.419*** (0.088)
Did not ask for clarification	0.221** (0.079)	0.277*** (0.066)
# of contacts with HH in wave 4	0.057* (0.023)	0.034 (0.019)

- Table 5 continued -

<b>Interviewers' characteristics (w4)</b>		
Age	-0.001 (0.005)	-0.007 (0.005)
Female	0.212 (0.113)	-0.010 (0.096)
Experience with working on previous SHARE waves	0.707*** (0.122)	0.619*** (0.111)
Interviewer specific mean of contacts with HH until cooperation/refusal	-0.189* (0.074)	-0.136* (0.066)
Rounding to a multiple of 5 at grip strength measure (too many)	-0.071 (0.114)	-0.207* (0.102)
Rounding to a multiple of 5 at grip strength measure (too few)	-0.895*** (0.249)	-0.713** (0.236)
<b>Agency control variables</b>		
Priority decided by Survey Agency	0.129 (0.143)	0.167 (0.111)
Everyday contacts	0.603*** (0.176)	0.592*** (0.145)
Southern countries		0.035 (0.169)
Central countries		0.015 (0.120)
Survey Agency dummies		
_cons	-4.928*** (1.285)	-3.580*** (1.083)
$\sigma_u^2$ (interviewer level)	0.840	0.948
$\sigma_v^2$ (NUTS 1 level)	0.031	
Deviance	9640.173	13639.634
N	12688	17112

Standard errors in parentheses \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.001$

### Survey agency practices

In order to understand whether a small number of groups at the third level is problematic, we run the analysis based on NUTS 1 regions<sup>15</sup>, increasing the number of group at the highest level from 11 to 44. Looking at column 2 of Table 5, we can see that estimates are very similar to column 4 of Table 3. In column 2 of Table 5, we run a two-level model and include also controls for groups of countries to capture cultural effects and to understand whether *agency control* variables are still significant after controlling for other possible country determinants. We grouped countries as follows: the dummy *south* takes value one for Italy and Spain and zero otherwise, the dummy *centre* takes value one for Belgium, Switzerland, Germany, Czech Republic and Austria whereas *north*, the reference group, equals one for Denmark, Sweden and the Netherlands. We finally run the complete model by dropping countries one-by-one: results, available upon request, do not change.

<sup>15</sup> The information about NUTS1 regions has been derived from the accommodation history section in SHARELIFE where individuals are asked to report the region for each house where they have lived. Typically interviewers move within NUTS2 regions, but some of them cover larger areas, to preserve the nested structure of the model, we dropped interviewers, 136, who operated in more than one NUTS 1 region.

## 7. Conclusions

Nonresponse has been a longstanding issue in survey research with several studies seeking to identify factors that affect attrition in panel surveys. Our analysis focuses especially on the role of paradata in providing additional information to predict cooperation. We are especially interested in factors affecting cooperation propensity that are “under the researcher control”: survey agencies fieldwork strategies, features of interviewers and the respondent-interviewer interaction. We investigate which paradata from SHARE wave 3 and 4 help to predict cooperation in wave 4, regarding (1) previous experience with interviews, (2) current interviewer characteristics, and (3) agency-level fieldwork indicators. Using multilevel models, we found that factors at all three levels (respondent, interviewer and survey agency) influence cooperation.

Panel respondents are allowed to base their cooperation decision on their previous interview experience. The respondent-interviewer interaction from wave 3 is predictive for future cooperation. Our findings highlight for instance that item nonresponse predicts cooperation in the next wave, i.e. respondents who answered most of the monetary items are more likely to participate in wave 4 than those who refused to provide answers to a considerable number of questions. Item nonresponse in monetary questions can be used for operational purposes to anticipate cases that are likely to be problematic. Interview length is another factor that is associated with cooperation in wave 4. Our results suggest that very long interviews discourage individuals in participating in later waves, and this has to be kept in mind when designing or modifying the questionnaire. But up to a certain point, longer interviews are associated positively with further cooperation and this might reflect engagement of the respondent in the study. The way interview length affects future cooperation is a gap in the current literature and further research is needed.

Previous experience with working as a SHARE interviewer seems to matter for convincing respondents more than interviewer’s socio-demographic characteristics such as age, gender or education. Interviewers who perform well on survey tasks which require diligence are also more successful in gaining cooperation. This again reflects the importance of high-quality training and selecting diligent individuals as interviewers. Moreover, the number of successful contacts can be used as an indicator to monitor interviewers, since our results highlight two competing effects. Given a specific interviewer, the higher the number of contacts the more likely cooperation is in wave 4, but interviewers who, on average, tend to register several contacts reduce or even cancel the positive effects of re-approaching the household. These indicators can be used to identify outliers that need further quality evaluations to assure that the prescribed protocol was followed. In general, more detailed measures of diligence or a combination of measures might help to investigate the relationship between work quality and survey participation outcomes. Further research should focus on the interaction between the different levels of respondents’ prior interview experience, current interviewers’ abilities and the overarching fieldwork strategies applied by survey agencies. Especially at the intersection of respondents and interviewers there is room for improvement in assignment strategies; strategies that do not only rest on regional proximity but also on demographics such as age and gender or interviewers’ abilities in approaching reluctant sample units.

Finally, when looking at survey agency related controls, we found that having contact everyday with interviewers increases the chances to gain cooperation of respondents. This result highlights the importance of communication between survey agency coordinators and interviewers to successfully conduct surveys. Given that this analysis represents a wave-specific description of response behaviour in SHARE we have to be cautious when extending our results to the whole gross sample. The response process in later waves might differ in important ways from participation patterns in earlier waves of a panel. To better understand the phenomenon and suggest more robust operational practices it will be useful to consider the whole longitudinal gross sample, i.e. all individuals that have been interviewed at least once. This will also allow for separating between retention and recovery. The underlying mechanisms for subsequent participation on the one hand (retention) and interrupted participation on the other hand (recovery) might differ. In our analysis we are able to explain only part of the variability at the interviewer level: additional information is needed to understand what determines cooperation propensity at the interviewer-respondent level. This highlights the importance of collecting data about personality traits (known also as ‘big 5’) or attitude questions on persuasion strategies as done in the European Social Survey (see Blom et al. 2011) or interviewers working for the UK National Centre for Social Research (Jäckle et al. 2013). This was partly already done within the SHARE interviewer survey for Germany (Korbmacher and Blom, 2011) and will be extended to the whole SHARE sample asking about interviewers’ general attitudes, own behaviour, experience with measurements and expectations in the future.

## References

- Aartsen, M. J., Smiths, C. H. M., Van Tilburg, T., Knopscheer, K. C. P.M. and D.J.H Deeg (2002). Activity in older adults: Cause or consequence of cognitive functioning? A longitudinal study on everyday activities and cognitive performance in older adults, *Journal of Gerontology: Psychological Science*, 57,153–162.
- Blom, A.G. (2012). Explaining cross-country differences in survey contact rates: application of decomposition methods. *Journal of the Royal Statistical Society Series A*, 175, Part1, 217–242.
- Blom, A.G. and M. Schröder (2010). Sample Composition 4 Years on: Retention in SHARE Wave 3. In Schröder (Ed.) *Retrospective Data Collection in the Survey of Health, Ageing and Retirement in Europe: SHARELIFE Methodology*. Mannheim: MEA, Universität Mannheim.
- Blom, A.G., Leeuw, E.D.D., and J.J. Hox (2011). Interviewer Effects on Nonresponse in the European Social Survey. *Journal of Official Statistics*, 27, 359-377.
- Blom, A.G., P. Lynn and A. Jäckle (2008). Understanding Cross-National Differences in Unit Non-Response: The Role of Contact Data. ISER Working Paper 2008-01.
- Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S. and S. Zuber (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, 42, 992-1001.

- Branden, L., Gritz, R.M., and M.R. Pergamit (1995). The Effect of Interview Length on Attrition in the National Longitudinal Survey of Youth. NLS Discussion Paper. Report: NLS 95-28.
- Burton, J., Laurie H. and Moon N. (1999). Don't ask me nothin' about nothin', I just might tell you the truth - The Interaction between unit nonresponse and item nonresponse. Paper presented at the International Conference on Survey Nonresponse, Portland, Oregon.
- Celidoni, M., Dal Bianco, C., and G. Weber (2013). Early retirement and cognitive decline. A longitudinal analysis on SHARE data. Marco Fanno working paper n. 174- 2013.
- Couper, M.P., and F. Kreuter (2013). Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society Series A*, 176, 271-286.
- De Leeuw, E.D., Hox, J.J., and M. Huisman (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics*, 19, 153–176.
- Durrant, G. B., R. M. Groves, Staetsky, L. and F. Steele (2010). Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys. *Public Opinion Quarterly*, 74, 1–36.
- Durrant, G., and F. Kreuter (2013). The Use of Paradata in Social Survey Research. *Journal of the Royal Statistical Society Series A*, 176, 1-3.
- Enzmann, D. and U. Kohler (2012). MERESC: Stata module to rescale the results of mixed nonlinear probability models. Statistical Software Components S457400, Boston College Department of Economics.
- Fielding, A. (2004). Scaling for Residual Variance Components of Ordered Category Responses in Generalised Linear Mixed Multilevel Models. Quality and Quantity, *The European Journal of Methodology*, 38, 425-433.
- Fricke, S., Creech, B., Davis, J., Gonzalez, J., Tan, L., and N. To (2012). Does Length Really Matter? Exploring the Effects of a Shorter Interview on Data Quality, Nonresponse, and Respondent Burden. Paper presented at the Federal Committee on Statistical Methodology 2012 Research Conference, Washington DC, USA.
- Groves, R.M, and M.P. Couper (1998) Nonresponse in Household Interview Surveys. New York: John Wiley & Sons, Inc.
- Hox, J., and E. de Leeuw (2002). The Influence of Interviewers' Attitude and Behavior on Household Survey Nonresponse: An International Comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little (Eds.), Survey nonresponse. New York: Wiley
- Hox, J.J. (2010). Multilevel Analysis. Techniques and Application, (Second Edition). New York: Routledge.
- Jäckle, A., Lynn, P., Sinibaldi, J., and S. Tipping (2013). The Effect of Interviewer Experience, Attitudes, Personality and Skills on Respondent Co-Operation with Face-to-Face Surveys. *Survey Research Methods*, 7, 1-15.
- Kneip, T. (2013). Survey Participation in the Fourth Wave of SHARE, Malter, Frederic und Börsch-Supan, Axel, SHARE Wave 4: Innovations & Methodology, MEA, München. 140-155.
- Korbmacher, J. M., and M. Schroeder (2013). Consent When Linking Survey Data with Administrative Records: The Role of the Interviewer. *Survey Research Methods*, 7, 115-131.

- Kreuter, F. (2013). *Improving Surveys with Paradata: Analytic Uses of Process Information*. Hoboken, New Jersey: John Wiley & Sons.
- Krosnick, J.A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5, 213-36.
- Lepkowski, J. M. and M.P. Couper (2002), Nonresponse in the second wave of longitudinal household surveys. In *Survey Nonresponse* (eds R. M. Groves, D. A. Dillman, J. L. Eltinge and R. J. A. Little). New York: Wiley.
- Lipps, O., and G. Benson (2005). Cross National Contact Strategies. Paper presented at the AAPOR - ASA Section on Survey Research Methods.
- Lipps, O., and A. Pollien (2011). Effects of Interviewer Experience on Components of Nonresponse in the European Social Survey. *Field Methods*, 23, 156-172.
- Loosveldt, G., and K. Beullens (2013). The impact of respondents and interviewers on interview speed in face-to-face interviews. *Social Science Research*, 42, 1422-1430. doi: <http://dx.doi.org/10.1016/j.ssresearch.2013.06.005>.
- Loosveldt, G., Pickery, J. and J. Billiet (2002). Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey. *Journal of Official Statistics*, 18, 545-57.
- Lutig, P. (2014). Panel Attrition: Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers. *Sociological Methods & Research*, 14, 699-723.
- Lynn P., Kaminska, O. and H. Goldstein (2011). Panel attrition: How important is it to keep the same interviewer? ISER Working Paper Series, 2011-02, Institute for Social & Economic Research, ESRC.
- Lynn, P. (2008). The problem of nonresponse. In: De Leeuw et al. (Eds.). *International Handbook of Survey Methodology*. London & New York: Lawrence Erlbaum Associates.
- Lynn, P. (2013). Longer interviews may not affect subsequent survey participation propensity. *Understanding Society Working Paper Series: 2013-07*.
- Nicoletti, C., and F. Peracchi (2005). Survey response and survey characteristics: Microlevel evidence from the European Community Household Panel. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168, 763-781.
- Paccagnella O. (2006). Centering or Not Centering in Multilevel Models? The Role of the Group Mean and the Assessment of Group Effects. *Evaluation Review*, 30, 66-85.
- Pickery J., Loosveldt G., and A. Carton (2001). The Effects of Interviewer and Respondent Characteristics on Response Behavior in Panel Surveys: A Multilevel Approach. *Sociological Methods & Research*, 29, 509-23.
- Rabe-Hesketh, S. and A. Skrondal (2005). *Multilevel and Longitudinal Modeling Using Stata*. Second Edition. College Station, Texas: Stata Press.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* 63, 581-592.
- Salthouse, T. A. (2009). When does age-related cognitive decline begin?, *Neurobiology of Aging*, 30,507-514
- Schaie, K. W. (1989). Individual differences in rate of cognitive change in adulthood, in *The Course of Later Life: Research and Reflections* (Eds.) V. L. Bengtson and K. W. Schaie, Springer, New York.

- Self, S. G., and K.Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82, 605-610.
- Sharp, L.M. and J. Frankel (1983). Respondent Burden: A Test of Some Common Assumptions, *Public Opinion Quarterly*, Vol. 47, 36-53.
- Skrondal, A. and S. Rabe-Hesketh (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society: Series A*, 172, 659-687.
- Skrondal, A. and Rabe-Hesketh, S. (2004). Generalized latent variable modeling: Multilevel, longitudinal and structural equation models. Boca Raton, FL: Chapman & Hall/CRC.
- Watson, N. and M. Wooden (2009). Identifying Factors Affecting Longitudinal Survey Response. In *Methodology of Longitudinal Surveys*, edited by P. Lynn. Chichester, England: John Wiley.

## APPENDIX A

Table A1: Descriptive statistics of Variables at Respondent Level (N= 17112)

Variable	Mean	SD	Min	Max	Description
Cooperation	0.83	0.37	0	1	Cooperation in w4 (Outcome)
Female	0.56	0.50	0	1	Gender (reference: male)
Age	68.06	9.79	35	102	Age of Respondent in years
Being in poor health	0.39	0.49	0	1	Self-reported poor health
HH composition – partner interviewed	0.67	0.47	0	1	Partner in HH, and interviewed in w3 (ref: partner without interview)
HH composition – single	0.24	0.43	0	1	Single HH (ref: partner without interview)
Years of Education	10.73	4.75	0	25	
HH Income – 1 <sup>st</sup> quartile	0.25	0.43	0	1	HH Income, 1.Quartile by country
HH Income – 2nd quartile	0.25	0.43	0	1	HH Income, 2.Quartile by country
HH Income – 3rd quartile	0.25	0.43	0	1	HH Income, 3.Quartile by country
Working	0.27	0.45	0	1	If R declares to be employed or self employed
Living in an urban area	0.23	0.42	0	1	Small town or rural area (ref: urban)
Living in a single house	0.70	0.46	0	1	Living in a single house (ref: flat)
Interviewed in w1 and w3	0.64	0.48	0	1	Interviewed in w1 and w3
Interviewed in w1, w2, w3	0.57	0.50	0	1	Interviewed in w1,w2,w3
Item nonresponse in financial questions	0.21	0.28	0	1	Percentage of item nonresponse to monetary items in w3
Length of interview	0.90	0.38	0.24	2.55	Length of interview in w3 (in hours)
Willingness to answer	0.92	0.26	0	1	Willingness to answer in w3
Did not ask for clarification	0.83	0.37	0	1	Did not ask for clarifications in w3
# of contacts with HH in wave 4	2.45	1.40	1	27	Times hh was contacted in w4

Data: SHARELIFE release 1, SHARE wave 4 release 1.1.1 and internally available paradata.



Table A2: Descriptive statistics of Variables at Interviewer Level (N= 643)

Variable	Mean	SD	Min	Max	Description
Interviewer's Age	54.10	12.21	19	79	
Interviewer's Female	0.62	0.49	0	1	
Interviewer's Experience with working on previous SHARE waves	0.54	0.50	0	1	
Interviewer's Average # of contacts in wave 4	2.52	0.80	1	7	Average # of contacts in wave 4 by interviewer
Rounding to a multiple of 5 at grip strength measure (too many)	0.39	0.49	0	1	If the interviewer's percentage of rounding is below (above) the lower (upper) cut-off of the 90% confidence interval centred around the statistically expected value of 20.8%
Rounding to a multiple of 5 at grip strength measure (too few)	0.06	0.23	0	1	

Data: SHARE and internally available interviewer information, wave 4

Table A3: Descriptive statistics of Variables at Agency Level (N= 11)

Variable	Mean	SD	Min	Max	Description
Priority_byagency	0.4	0.51	0	1	agency decides the priority of projects agency monitors and has contact with the
Everyday_contact	0.2	0.41	0	1	interviewers on a daily basis

Data: SHARE internally available agency information, wave 4.

## APPENDIX B

Table B1: Descriptive statistics of Variables at Respondent Level (N= 17112)

Variable	Mean	SD	Min	Max	Description
Pace of interview	0.20	0.30	0.04	0.84	Pace = Length of interview in w3 (in hours)/# of item

Data: SHARELIFE release 1, SHARE wave 4 release 1.1.1 and internally available paradata.

Table B2: Descriptive statistics of Variables at Interviewer Level

Variable	Mean	SD	Min	Max	N	Description
Interviewer education (iscsed 5-6)	0.33	0.47	0	1	517	Interviewer has tertiary education
Short introductions	0.78	0.42	0	1	643	Interviewer has at least one short introduction (i.e. time recoded lower than a country-specific median)

Data: SHARE and internally available interviewer information, wave 4