SHARE ERIC
SURVEY OF HEALTH, AGEING AND RETIREMENT IN EUROPE

SHARE WORKING PAPER SERIES

Can a responsive fieldwork design increase response rates and decrease response bias in the Survey of Health, Ageing and Retirement in Europe?

Michael Bergmann and Annette Scherpenzeel

Working Paper Series 27-2016

SHARE-ERIC | Amalienstr. 33 | 80799 Munich | Germany | share-eric.eu

mea | MAX-PLANCK-INSTITUT FÜR SOZIALRECHT UND SOZIALPOLITIK
Munich Center for the Economics of Aging
Horizon 2020
Federal Ministry of Education and Research
DFG
NIH National Institute on Aging

Can a responsive fieldwork design increase response rates and decrease response bias in the Survey of Health, Ageing and Retirement in Europe (SHARE)?

Michael Bergmann[*] *and Annette Scherpenzeel*

Munich Center for the Economics of Aging (MEA),
Technical University of Munich (Chair for the Economics of Aging)

Abstract: The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national face-to face panel study of the process of population ageing. For the sixth wave of data collection, a responsive fieldwork design was implemented in the German sub-study of SHARE. The aims of this design were, firstly, to improve the overall response rate in the German panel and, secondly, to decrease nonresponse bias. In this respect, responsive designs have been given a lot of attention in the recent survey methodology literature. These designs make use of background information to more efficiently allocate fieldwork resources to specific sample units. SHARE is especially suitable for such a strategy, because it already conducts a high level of fieldwork monitoring, has in place an advanced system to register fieldwork results, and possesses extensive information about the interviewer performance as well as the panel members and their response behavior in previous waves. Against this background, we implemented a responsive monitoring design for the German sub-study that served as a "dashboard" of response probabilities for relevant respondent characteristics during fieldwork. This allowed for immediate feedback to the survey agency and focused actions with regard to specific groups of respondents. Our analyses show that while the responsive measures seem to stimulate the overall response rate, the final wave 6 response probabilities were not more homogenous across respondent groups than in wave 5. The responsive fieldwork design hence did not lead to a reduction in nonresponse bias.

[*] Corresponding author: Michael Bergmann; E-mail: bergmann@mea.mpisoc.mpg.de.

Acknowledgment: This paper uses data from SHARE Waves 1, 2, 3 (SHARELIFE), 4 and 5 (DOIs: 10.6103/SHARE.w1.500, 10.6103/SHARE.w2.500, 10.6103/SHARE. w3.500, 10.6103/SHARE.w4.500, 10.6103/SHARE.w5.500), see Börsch-Supan et al. (2013) for methodological details.
The SHARE data collection has been primarily funded by the European Commission through the FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Additional funding from the German Ministry of Education and Research, 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) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

1. Introduction

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national panel study of the process of population ageing. It studies the different ways in which people over 50 live in 20 European countries from Sweden to Greece and Portugal to Estonia. Since 2004, data are collected face-to-face every two years, using a harmonized core questionnaire in all countries, as well as objective health measures such as hand grip strength and dried blot spots (Börsch-Supan and Jürges 2005, Börsch-Supan et al. 2013).

As for all panel surveys, retaining respondents over waves is crucial for SHARE. To study the process of ageing and changes in health and living conditions across Europe, respondents need to be observed across multiple points in time. Moreover, panel attrition not only decreases the number of longitudinal observations but can also lead to biased estimates since some groups of respondents are more likely to drop out than others. For a cross-national study like SHARE, the comparability across countries can be harmed if differential attrition and attrition bias occurs in countries. SHARE therefore devotes much attention to the motivation of longitudinal respondents, for example implementing incentive schemes for respondents and interviewers, extensive interviewer training and monitoring, and tracing and tracking of respondents between waves (Blom and Schröder 2010, Börsch-Supan, Krieger, and Schröder 2013, Kneip, Malter, and Sand 2015, Malter 2013). The present study is a continuation of these efforts, extending them to recent developments in survey methodology, which concentrate on the adaptation of fieldwork procedures to respondent characteristics. During the fieldwork of the sixth SHARE wave in Germany, we applied a so-called “responsive fieldwork design” focusing on specific groups with known low response probabilities.

Responsive as well as adaptive designs have been given a lot of attention in the recent survey methodology literature (e.g. Groves and Heeringa 2006, Schouten 2010, Schouten, Calinescu, and Luiten 2013, Couper and Wagner 2011, Lepkowski et al. 2013). The principle of such designs is that, instead of using the same design and same treatment for all sample units, the survey fieldwork strategy is tailored to different persons or households and adapted in response to different fieldwork results. The different, tailored strategies may be defined before the survey starts (adaptive designs, or static adaptive designs; Schouten, Calinescu, and Luiten 2013), but may also depend on data that is observed during data collection (responsive designs or dynamic adaptive designs; Schouten, Calinescu, and Luiten 2013). The choice of differential strategies is based on background variables that are known before data collection, for example sampling frame information or data from previous waves of data collection. In addition, paradata information which is collected *during* the fieldwork can be used. Since the concept of responsive designs evolved, many studies were published that applied such designs to existing surveys – with varying results on the response rates. Mohl and Laflamme (2007) describe how Statistics Canada has implemented some changes to their fieldwork design on the basis of response probabilities in previous surveys and what responsive measures could make this fieldwork still more efficient. Wagner et al. (2012) as well as Kirgis and Lepkowski (2013) defined a restricted period of a few weeks in the midst of their fieldwork, which was dedicated exclusively to a demographic group that was falling behind in response probabilities. This effort succeeded to increase the response of this group up to the level of the other groups. In the same study, Wagner et al. (2012) also implemented a one week period early in the fieldwork, in which screening contacts were prioritized to interviews, to counteract the interviewer preference for interviews over screenings. Luiten and Schouten (2013) adapted the choice of interviewing mode, call

schedule, and interviewer assignment on the basis of expected response probabilities for demographic groups, which resulted in slightly higher response rates and significantly better representativeness in comparison to a control sample. Peytchev et al. (2010) gave interviewers higher payments for completed interviews in the low propensity groups, but this manipulation did not affect the response rates and response rate variation in comparison to the control condition.

Numerous studies tried to optimize call schedules for interviewers on the basis of previously found response probabilities at different days and times. Lipps (2012) showed that in a panel study, the response probability can be increased by contacting panel members at the same day of the week and time of the day at which they were successfully interviewed in the prior wave. This approach was also followed by Kreuter and Müller (2015) in a German panel study using CATI interviews, but failed to show any significant effects on the contact and the cooperation probabilities. In this study, as well as in a similar study by Wagner (2013), it seemed that the main reason for this lack of results was non-compliance of the interviewers with the given instructions. Kreuter and Müller (2015) found that only 63% of their treatment group had indeed been contacted in the same window as they were scheduled on the basis of their prior participation. Wagner (2013) found a significant increase in CATI contact rates as a result of call schedule optimization during the fieldwork but no effect in a face-to-face survey. Moreover, he found that interviewers were not following the recommended contact windows in practice.

Since the main SHARE data collection is done by face-to-face interviews in all countries, we expected that the results of our efforts to implement a responsive fieldwork design would also depend strongly on interviewer compliance, in line with the results of Wagner (2013). We hence

decided to test a responsive fieldwork design in one country first before extending the complete operational management of SHARE with the extra efforts needed for such designs.

The main objective of the responsive fieldwork design hence was to improve the overall household response rate in wave 6 in Germany. In general, response fieldwork designs are meant to increase the homogeneity of response probability across respondent subgroups, thus improving the sample representativeness. This was also the secondary goal of the present study, since it had been shown that the response probabilities in SHARE were significantly different for subgroups in all countries, including Germany (Bristle et al. 2014). Finally, the objective of the test in Germany was to evaluate the feasibility and utility of implementing responsive fieldwork designs across all countries participating in SHARE.

2. Data and methods

2.1 Response rate definition

Since several types of response rates can be distinguished for longitudinal surveys, it is useful to define in detail which panel response rate we use in this study. Our definition is based on the "conditional cross-sectional" response rate for longitudinal studies, defined by Cheshire et al. (2011) and Lynn (2005) as: The proportion of sample members who respond in a given wave (including partial interviews) of those who responded in the immediately prior wave. We made an adaptation to this definition, since we include in the denominator only sample members who responded in the immediately prior wave and for whom at least one contact attempt has been made in the present wave. This adaptation made the response rate more useful for the monitoring during the early stages of the fieldwork, in which interviewers had not yet contacted all sample

members. Another distinction that is important to make is between the individual response rate and the household response rate. In the present study, we have chosen the household response rate, because most of the fieldwork adaptions which we implemented were aimed at households instead of individuals, for example the optimization of contact attempts. In short, in our analysis of the response rate in wave 6 conditional on wave 5, this rate[1] is defined as:

$$RR\ wave\ 6|\ wave\ 5 = \frac{number\ of\ households\ with \geq 1\ interview\ in\ wave\ 6}{\substack{number\ of\ households\ with \geq 1\ interview\ in\ wave\ 5 \\ and\ at\ least\ one\ contact\ attempt\ in\ wave\ 6}} \quad (1)$$

2.2 Sample

In the following, we use internal monitoring data (SHARE internal release 1.0.0 DE monitoring). The German subsample was originally generated with the help of the municipal registers to obtain a representative probability sample of the population aged 50 years and older. Data are collected face-to-face. The initial cross-sectional response rate for the baseline wave on the household level was 63.4% in the main sample (Börsch-Supan and Jürges 2005). The resulting baseline sample included 3008 individuals. Two years later, the first re-interview of panel members was done, followed by re-interviews every two years. Refreshment samples were recruited in wave 2 and wave 5. Because the first re-interview typically deviates in response patterns from subsequent panel waves we leave out the wave 5 refreshment sample from the analyses presented here and focus on the effect of the responsive fieldwork design on the response rates of the baseline sample of 2004 and the wave 2 refreshment sample of 2006. The conditional cross-sectional household response rates, as defined above, were significantly lower

[1] Following the standards set by the American Association of Public Opinion research (2016) this rate is most comparable with RR6.

in Germany than in other participating countries up to wave 4 (Kneip, Malter, and Sand 2015). Since other recent pan-European surveys (ESS5 2010, HFCS 2013, p. 43) show a similar pattern of comparatively low response rates in Germany, it is likely to be caused by the strict German data privacy law which prohibits to re-contact sample members who refused to participate for refusal conversion attempts or for subsequent panel waves. After wave 4, the decision was made to switch to another survey agency (TNS Infratest) for the fieldwork, which resulted in an improvement of the conditional cross-sectional household response rate in wave 5 as compared to the rate in wave 4 (Kneip, Malter, and Sand 2015)[2]. To further improve the German response rate in wave 6, the responsive fieldwork design described in this paper was developed.

2.3 Defining groups with differential response probabilities

SHARE is a study that is especially suitable for a response fieldwork design, because it already conducts a high level of fieldwork monitoring, has in place an advanced system across countries to register fieldwork results, and possesses extensive background information about the panel members and their response behavior in previous waves. A multilevel analysis of response rates in previous waves including both interviewer and respondent characteristics as predictor variables had shown already which groups of SHARE panel members are difficult to get or are more likely to drop out (Bristle et al. 2014). The participation probability of panel members in SHARE has been shown to depend on respondent characteristics such as age (nonlinear effect, 69 years peak), self-reported poor health (lower response probability), household composition

[2] Note that the response rates presented in Kneip et al. (2015) are conditional cross-sectional response rates with a different denominator then used in formula 1 and in our analyses: the authors use a definition in which all sample members that were interviewed in the prior wave or in any previous wave (and are still eligible) are included in the denominator.

(couple of which one partner not cooperated has lowest probability), working (lower response probability), type of house (single house = higher response probability), and self-reported activity level (no activities = lowest response probability). Furthermore, participation can also be predicted from paradata, such as the number of item missings in monetary questions (more= lower response probability) and the number of contact efforts needed in the last wave (higher = lower response probability).

We developed an analysis script on this prior knowledge of response probabilities, showing us on a two-weekly basis the response rates associated with the selected respondent characteristics during the fieldwork. Although we analyze the response probabilities for households in our analysis, some of the characteristics used as predictors are on the individual level. For those predictor variables we used the CAPI information from previous waves, in most cases from the immediately preceding wave. For cases with missing information in the preceding wave, due to item or case nonresponse, information from earlier waves was used to reduce the loss of observations. In a single household this procedure is straightforward. In households with more than one eligible person, we used the characteristics of the household member who answered the largest number of questions of the SHARE wave 5 questionnaire. When encountering ties, we selected the person who completed the household grid questionnaire about the household composition and changes in the household.

2.4 Analysis model

The analysis scripts developed for the fieldwork monitoring were structured in three steps: First, we looked at the development of the household response rate defined above in wave 6 as

compared to wave 5 at corresponding time points during the fieldwork period. Thus, the development in wave 5, in which no special fieldwork adaptions were done, served as a control condition to evaluate the performance of our responsive design as well as the extra fieldwork measures that are described in detail below. This control condition is not the same as a randomized experiment. However, since retaining the longitudinal sample size was of existential importance for the German SHARE sub-study in wave 6, we could not afford the risk to have a lower response rate in part of this sample. We do not compare wave 5 and wave 6 to prove that the overall response rate increased as a result of the responsive design. Differences in the overall response rate can be related to many other factors which we could not control for. However, the purpose of the comparison rather is to see whether changes in the response probabilities over the fieldwork period reflect the normal pattern for a specific subgroup or might be attributed to our wave 6 interventions. In wave 5, the numerator of formula 1 consisted of 760 households in which at least one individual interview was done at the end of the fieldwork period, the denominator included 890 attempted and eligible households in which at least one interview was realized in the previous wave (i.e. wave 4). In wave 6, the numerator consisted of 726 households, the denominator of 799 households.

Secondly, we calculated the absolute response probabilities related to the different respondent characteristics described above, as well as the deviation of these response probabilities from the overall mean household response rate. Furthermore, a logistic regression model was ran including the respondent characteristics as explanatory variables and the response rate defined in formula 1 as dependent variable. To account for interviewer effects, we used robust standard errors that are clustered over interviewers. As we are interested in the raw distribution of respondents participating in the survey or not, no specific weights have been applied. Finally, we

monitored whether response probabilities related to respondent characteristics became more homogeneous or more heterogeneous over time by continuously producing graphs of the development of the deviations in response probabilities over the entire fieldwork period.

2.5 Fieldwork adaptations

Three fieldwork adaptations were implemented for the respondent groups, which had low response probabilities in our analysis: 1) interviewer bonus incentives, 2) contact schedule optimization, and 3) individual interviewer feedback.

Interviewer bonus incentives: Since it was known from previous waves that people of 80 years or older had especially low conditional response probabilities, we put in place a specific strategy for this group in advance. Before the fieldwork started, interviewers were promised a bonus incentive of five euros extra on top of their normal per interview payment for each interview with a person of 80 years or older. This can also be regarded as a (static) adaptive design measure (Schouten, Calinescu, and Luiten 2013).

Contact schedule optimization: This measure was developed during the fieldwork monitoring, when the results showed that the youngest age group between 50 and 65, had lower than average contact rates as well as respondents who were still working. We proved on the basis of the incoming contact data that the early evening was the best time to successfully contact these groups, showed this finding to the interviewers, and urged them to do so.

Individual interviewer feedback: The SHARE monitoring system allowed the monitoring of contact and response rates obtained by individual interviewers in each country. Using this

system, we provided intensive feedback about the contact and response rates of individual interviewers to the agency, who in turn would contact the underperforming interviewers. This measure was in line with two other measures in our program, since we especially looked at the interviewer rates for specific respondent groups as well, such as the youngest working group and the oldest group over 80 years.

In addition to the responsive and adaptive measures that were implemented for specific groups, some general motivational measures for all respondents were implemented. Besides the usual high respondent incentives which SHARE Germany pays conditionally on participation, we used a small unconditional gift in wave 6: A postage stamp booklet including three postage stamps was included in the advance letter for all panel members. Furthermore, the interviewers were paid an extra five euros for each completed interview if they reached their individual target of interviewing 80% of the sample members assigned to them.

3. Results

In table 1, we present the results of the analysis script at the first and the last monitoring time point, hence at the earliest stage of the fieldwork and after the fieldwork for the longitudinal sample in wave 6 was finished. In the first line of the table, we present the overall household response rate as defined in formula 1. At the first monitoring moment, two weeks after the start of the wave 6 fieldwork, contact attempts had been made in 266 households and in 37.6% of these households at least one interview had already been done. Similarly, at the end of the fieldwork period, contact attempts had been made in 799 households and in 90.9% of these households an interview was done. The next lines of the table show what these response rates

were for specific groups of panel members, as compared to the average rate in the first line. For each group we present the observed (univariate) deviation from the overall mean response probability. In addition, the B coefficients resulting from a multivariate logistic regression model including all characteristics are given.

Table 1: Monitoring of respondent characteristics in wave 6

	Start of fieldwork			End of fieldwork		
	N	Deviation from obs. response probability	b (se)[1]	N	Deviation from obs. response probability	b (se)[1]
Household response rate (based on attempted HH)	266	37.6		799	90.9	
Gender						
female (ref.)	53	-0.5		370	-1.5	
male	47	0.6	-0.04 (0.32)	356	1.6	0.31 (0.29)
Age						
50-64	13	-11.6	-0.19 (0.48)	180	2.4	0.17 (0.40)
65-69 (ref.)	25	1.5		156	0.4	
70-79	38	-0.7	-0.10 (0.29)	269	1.9	0.16 (0.42)
80+	24	11.4	0.27 (0.40)	121	-7.4	-0.57 (0.40)
Education						
low	52	4.1	0.17 (0.32)	335	-1.3	0.08 (0.29)
medium (ref.)	26	-2.9		217	-0.4	
high	20	-4.8	0.07 (0.35)	161	3.3	0.45 (0.40)
Composition of household						
single hh (ref.)	34	4.9		201	-2.3	
hh with 2 persons	57	0.2	0.04 (0.31)	443	1.2	0.21 (0.30)
hh with more than 2 persons	9	-11.9	-0.20 (0.51)	82	-0.8	-0.26 (0.55)
Working status						
retired (ref.)	80	3.3		490	-0.8	
paid work	12	-11.0	-0.39 (0.56)	152	1.8	0.22 (0.44)
other	8	-5.7	-0.41 (0.61)	83	1.4	0.33 (0.48)
Housing						
other (ref.)	44	7.0		249	-1.6	
detached house	55	-4.2	-0.44 (0.24)	475	0.9	-0.14 (0.33)
Urbanization						
rural (ref.)	61	-1.4		471	1.0	
urban	38	2.5	-0.05	253	-1.8	-0.41

			(0.40)			(0.31)
Household income						
1st quartile	29	13.3	0.67 (0.45)	161	1.1	0.06 (0.38)
2nd quartile (ref.)	20	-2.5		154	0.3	
3rd quartile	20	-0.6	0.14 (0.39)	162	1.7	0.13 (0.48)
4th quartile	22	1.0	0.36 (0.33)	153	2.4	0.08 (0.44)
no answer	9	-15.6	-0.64 (0.53)	96	-8.1	-0.81* (0.37)
Health						
good or better (ref.)	51	-4.7		416	0.4	
fair	37	4.9	0.25 (0.28)	235	-0.1	0.08 (0.26)
bad	12	12.4	0.12 (0.52)	75	-1.6	0.22 (0.40)
# contact attempts for a CI						
1 (ref.)	50	3.0		311	2.3	
2-4	42	-0.8		323	-1.8	
5-8	6	-10.1		73	-0.7	
more than 8x	1	-12.4		15	-2.6	
Activities						
no activities	9	9.8	0.07 (0.49)	46	-10.2	-0.66 (0.43)
light activities (ref.)	41	4.7		288	-2.0	
time consuming activities	50	-4.3	-0.29 (0.33)	392	2.9	0.52* (0.22)
Constant			-0.30 (0.59)			2.05** (0.65)
McKelvey and Zavoina's R^2			0.11			0.12
N (multivariate model)			258			782

[1] Cluster-robust standard errors in parentheses (accounting for interviewer level).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

As an illustration, we look at the following example: For the age group between 50 and 64 years, the response rate at the beginning of the fieldwork was 11.6%-points lower than average, hence 26% (37.6% - 11.6%-points). The B coefficient of this group was -0.19 at that time point, indicating that the probability of an interview in this age group was slightly lower than the probability of an interview in the reference age group of people between 65 and 69 years, when controlling for all other respondent characteristics in the model. This difference in response probability was not significant in the model at the first time point, which included only small numbers of households in the groups. At the end of the fieldwork period, the response probability of the youngest age group in comparison to the other groups had significantly

changed: The deviation in probability (2.4%-points) as well as the B coefficient (0.17) were now positive and small, indicating that the probability of an interview in this group, after final contact attempts, was equal or slightly (but not significantly) higher than in the reference age group. We can see more such changes in effect signs, for example for the age group over 80 years old, the lowest and the highest educated groups and the single person households. The largest and significant effects were found, at the end of the fieldwork, for respondents who had not answered the income question in the previous wave questionnaire (lower interview probability) and respondents who had reported time consuming activities (higher interview probability).

However, the monitoring script was not only run at the beginning and the end of the fieldwork, but bi-weekly during the whole fieldwork period. It is therefore more interesting to look at the complete pattern of the changes in response probabilities over the fieldwork period. Figure 1 gives an overview of the development of the household response rate over the total fieldwork period in wave 5 and wave 6. The fieldwork period in wave 5 lasting from February to September 2013 was somewhat longer (31 weeks) than in wave 6 lasting from February to July 2015 (23 weeks), but in the final eight weeks of wave 5 the response of the longitudinal sample barely increased anymore[3]. The response in wave 6 starts at about the same rate as in wave 5 but accelerates after week 8, and ends at a considerably higher rate than in wave 5. The first objective of the special fieldwork design in wave 6 was hence realized: The final wave 6 response rate is 4.6 percentage points higher (90.9%) than the final wave 5 response rate (85.4%). A logistic regression to test the overall difference in response over time, using a continuous date variable, a wave dummy, and an interaction of both, showed a significant difference between the two waves from week 15 onwards.

[3] In wave 5, a large refreshment sample was recruited in addition to the fieldwork for the longitudinal panel sample, see section Sample. The extra weeks were merely needed for the fieldwork in this new sample.

Figure 1: Development of the overall household response rate over the wave 5 and wave 6 fieldwork period

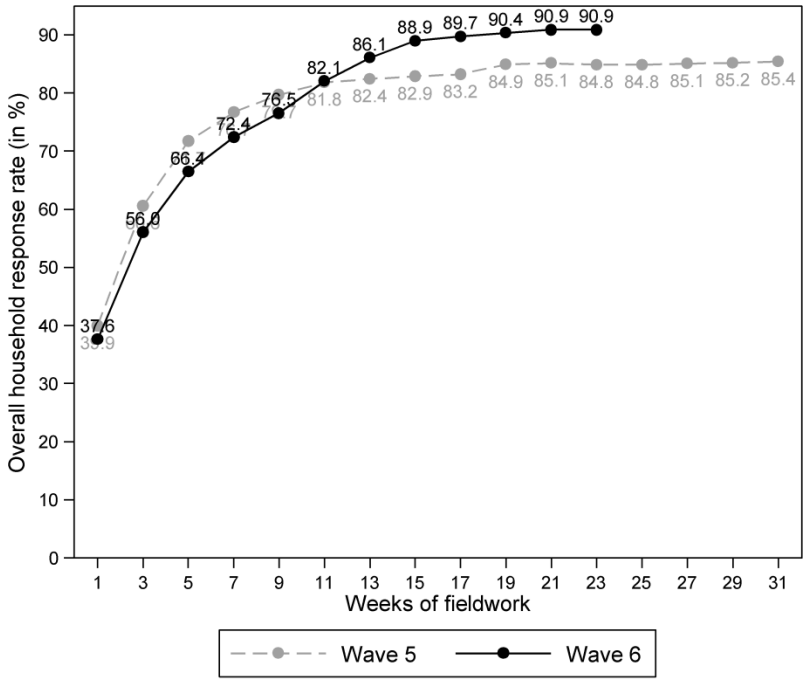


Figure 2 shows the response probabilities over time for the two age groups for which we observed a large shift in the response deviation over time in table 1: the youngest age group (50-64) and the oldest age group (80+). The two patterns of response seem to be consistent with the overall graph in figure 1. However, we get a very different result from depicting the *deviations* in response probabilities for these groups, as we used in table 1. Figure 3 shows, at the left, a large negative deviation from the average response rate for the youngest age group at the start of the fieldwork in wave 6. A negative but much smaller deviation was also found for this group at the start of the wave 5 fieldwork. Since we observed a similar low response probability for the respondent group that was working, we hypothesized that the lower response rate was related to a lower contact probability of this group and that we could increase this probability by optimizing contact timing. Tests on the first fieldwork data indeed showed that contact attempts

during the early evening (17.00 - 19.00 h.) had a significantly higher interview probability in this specific age group. In co-operation with the staff of the fieldwork agency, we informed the interviewers about our finding in fieldwork week 4 as well as in week 7, and recommended them to contact the youngest age group more often in the evenings. In fieldwork week 9, we wrote a midterm motivational letter for the interviewers, which also stressed this recommendation and reminded the interviewers that this could increase the likelihood of reaching their individual target of 80% response rate and thus getting the bonus payment. Finally, between week 7 and 12, we monitored the individual performance of interviewers in contacting the younger households, as well as their number of contact attempts in the recommended time window. Based on the combined information on respondent age, interviewer success rate and timing of contact attempts, we provided the survey agency with detailed lists of which interviewers[4] were to given feedback or re-instructions.

Although figure 3 seems to confirm that a change took place after week 3 of fieldwork, we cannot test whether it was really our recommendation that caused the increase in response rate for the youngest age group. The wave 5 response rate for the same group seemed to increase somewhat around that time of the fieldwork as well, even though these effects are much smaller. An indication of a possible causal effect can be found in the interviewer compliance data (not shown): Before we gave our recommendations, 26% of all contact attempts with younger households were done in the time window between 17.00 and 19.00 h. This number increased to 30% after our first and to 42% immediately after our second recommendation. The interviewers hence followed our recommendations. However, we cannot directly relate this finding to the increased response rate. The final deviation in response probability of the youngest age group

[4] The interviewers were anonymized for us: For our analysis and lists we used encrypted interviewer ID's for which the key was only known to the survey agency.

was very similar to the final deviation in wave 5 (figure 3), where no fieldwork adaptions have been proposed. However, figure 2 showed that the absolute response probability of this group was considerably higher in wave 6 (93.3%) than in wave 5 (87.3%), consistent with the overall response rates in figure 1. This difference is statistical significant at the 5%-level.

Figure 2: Development of the household response rate for the youngest age group (50-64) and the oldest age group (80+)

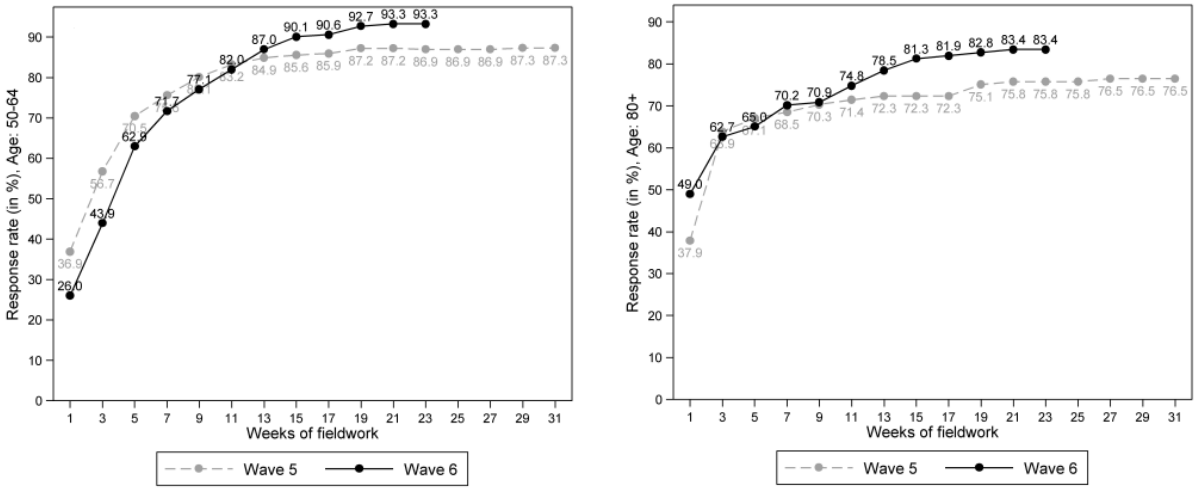
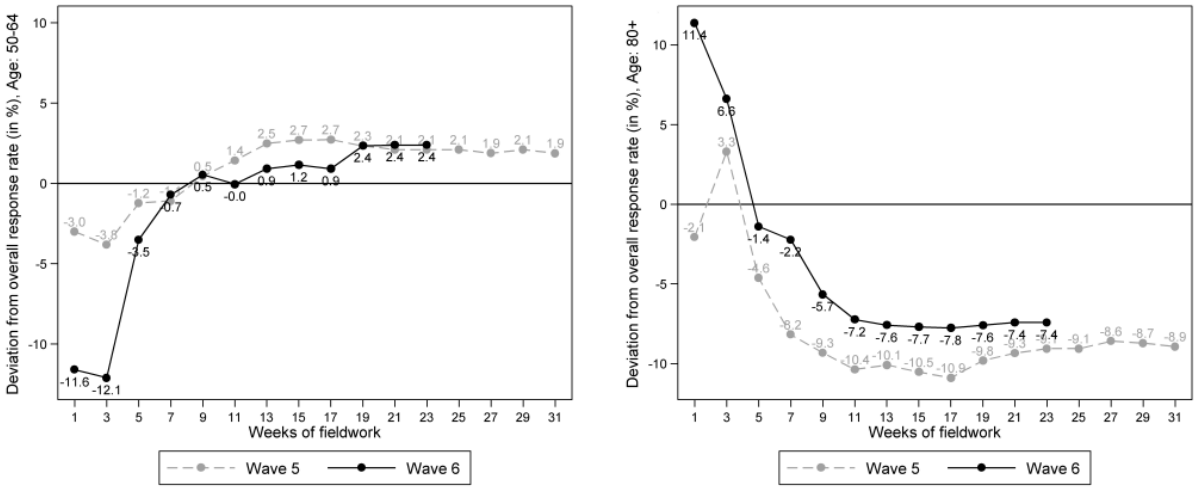


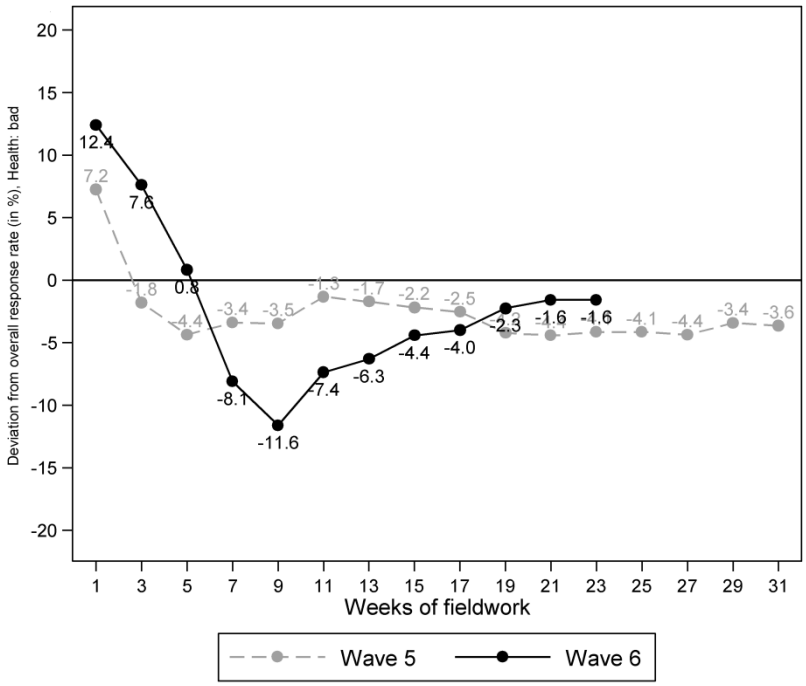
Figure 3: Deviations from the overall household response rate for the youngest age group (50-64) and the oldest age group (80+)



The right part of figure 3 shows that for the oldest age group of 80 years or older the pattern of deviations in response probability started with a large positive deviation, as was also shown in table 1. This was remarkable since we knew from previous waves that this age group normally had relatively low response probabilities. However, after a few weeks the pattern changed and the deviation indeed became negative. Our hypothesis was that, due to the bonus incentive for interviewing people over 80 years, interviewers were giving priority to contacting the oldest sample members first – perhaps also at the expense of the youngest age group. After a few weeks, they most probably had interviewed the easiest to reach 80+ group and were left with the oldest, difficult remaining cases. We still tried to stimulate the response probability of this group again after week 9 by reminding interviewers of the bonus incentive and by offering them help from the fieldwork staff to contact people in nursing homes. However, the response deviation remained negative, as in wave 5. The final deviation was, however slightly smaller than in wave 5 (-7.4%-points versus -8.9%-points) and the final absolute response rate was significantly higher (83.4% in wave 6, 76.5% in wave 5) at the 10%-level.

The pattern of deviations observed for the oldest age group, in figure 3, is to a certain extent reflected in the patters we observed for the group of sample members who self-reported to be in bad health in the previous wave questionnaire. Figure 4 shows that this group also starts with a relatively high response probability that drops around week 6 or 7 and stays low after that. Both groups indeed show some overlapping: From those persons reporting bad health over one third is over 80 years old, while only one fifth is under 65.

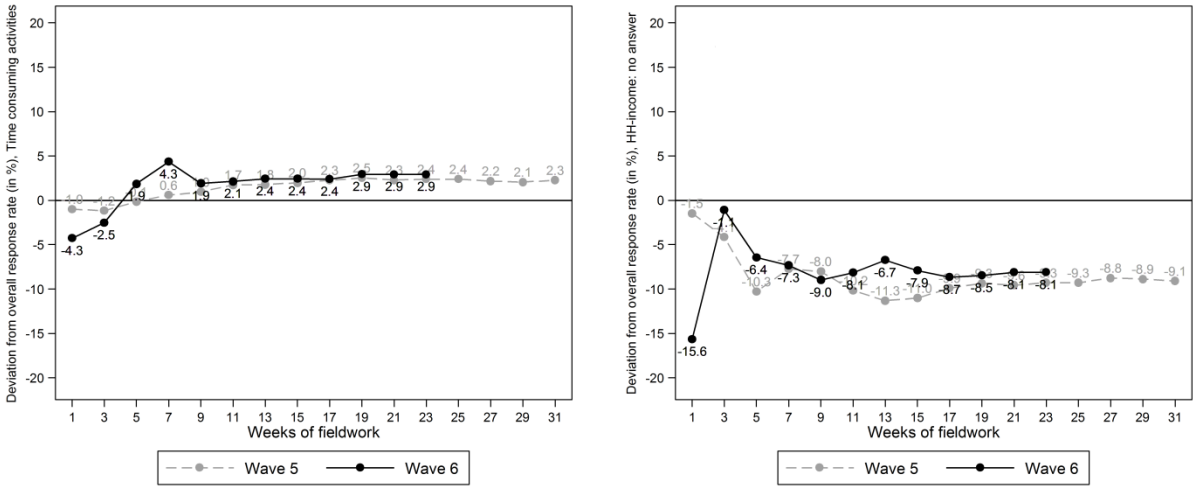
Figure 4: Deviations from the overall household response rate for respondents who reported to be in bad health



Two additional remarkable groups are depicted in figure 5, both having significant deviations in response probabilities at the end of the fieldwork period (see table 1). At the left of figure 5 is depicted the group of sample members who reported to have time consuming activities in the previous wave. This group has a slightly lower response probability at the start of the fieldwork in wave 6, increase to the mean level in week 4-5 and end somewhat higher than the average, similar to the final rate in wave 5. Since the wave 5 pattern was, however, more consistently positive, this group might also have suffered slightly from the extra efforts interviewers were devoting to the oldest age group at the start of the wave 6 fieldwork. We did not direct any special fieldwork efforts to this group, for which the response probabilities already recovered early in the fieldwork period.

Finally, the right part of figure 5 shows the deviations of the group who gave no answer to the income question in the previous wave. This group started with a much lower response probability than any other group and mostly remained low despite some fluctuations due the small number of observations in the first weeks. This pattern could be seen in wave 5 as well. Although we considered this to be a group for which responsive measures are especially worthwhile, it was impossible to design effective extra measures during the course of the wave 6 fieldwork without knowing the common cause of the income nonresponse in one wave and unit nonresponse in the next wave.

Figure 5: Deviations from the overall household response rate for respondents who reported time consuming activities and respondents who did not answer the income question



4. Conclusion

The first objective of implementing a responsive fieldwork design in the German SHARE sub-study was to improve the overall response rate. Since the final wave 6 response rate (90.9%) was significantly higher than the final wave 5 response rate (85.4%), we can conclude that this

first aim was attained even though we cannot unquestionably attribute this total increase to the responsive design. As we described in the method section, we not only used responsive measures for specific subgroups, but also implemented several response stimulation measures for the total sample, such a small unconditional gifts and interviewer incentives for reaching an individual response target. Each of the measures might have contributed to the overall result in wave 6, although we would expect the general measures to stimulate the response evenly over the fieldwork period. The same expectation applies to possible other differences between wave 5 and wave 6, such as more experienced interviewers and fieldwork staff or slightly increased attrition bias. However, the response in wave 6 started at about the same rate as in wave 5, accelerated after week 8 and was significantly higher than in wave 5 from week 15 onwards. The most intensive efforts to redirect the contact attempts of the interviewers for specific subgroups of households and feedback to interviewers were done between week 8 and week 12, which might be an indication that the responsive measures might have stimulated the overall response rate. Furthermore, we observed a substantial increase in the contact attempts with younger households in the recommended time window after our request to interviewers. Hence, in contrast with the findings of Wagner (2013) and Kreuter and Müller (2015), our interviewers seemed to comply with the recommendations given to them during the fieldwork, thus making it plausible to find a positive effect on response.

The secondary goal of the responsive fieldwork design in this study was to increase the homogeneity of response probability across respondent subgroups. This second aim has not been attained very well with our design. Despite the effect that the responsive measures seemed to have on the absolute response rate in the subgroups, the final wave 6 response probabilities in these groups were not closer to the average than in wave 5. The final deviation in response

probability of the youngest as well as the oldest age group, for example, was very similar to the final deviation in wave 5. Furthermore, some large response deviations were observed for respondent characteristics that cannot easily be counteracted by common fieldwork reactions such as incentives, intensified efforts or re-scheduled contact attempts. Specifically related to the target population of the SHARE survey is the (cognitive or physical) inability to participate caused by old age and poor health. In addition, the relationship we found between income item nonresponse in the previous panel wave and response probability in wave 6 is difficult to translate into effective measures during the course of the fieldwork without knowing more about this causal relation. This is line with Tourangeau (2015), who formulated the paradox that respondent characteristics that are suitable for responsive fieldwork measures might in fact be of limited use for true response bias reduction.

The third objective of the responsive fieldwork design in wave 6 in Germany was to evaluate the feasibility and utility of implementing such designs across all countries participating in SHARE. Our conclusion with respect to this third aim is that the responsive design at this moment is not yet efficient enough to outweigh the large investment of time and effort needed for the continuous monitoring, analyzing, and implementing of extra measures in all countries at the same time. We consider doing some in-depth studies of specific groups with low response probabilities before the design adaptations will be extended to all other countries. For example, it is worthwhile to find out more about the possible common cause of the income nonresponse in one wave and unit nonresponse in the next wave by specifically interviewing panel members about their reasons for not answering the income questions. This could result in better adapted strategies for this subgroup, such as different advance letters, specific incentives, or specially trained interviewers. Another study would be directed to the group of 80+ years-olds and the

people in bad health. When contacting these older or ill panel members, SHARE interviewers often encounter so-called “gate-keepers”, i.e. family members, adult children, or care givers who are hesitant to let the interviewer in or allow the interview. Such contact attempts could possibly be aided by studies finding out what information would convince the gate-keepers. In conclusion, we will continue to further differentiate the fieldwork strategy on the basis of respondent characteristics, but will focus more on the advance development of adaptive strategies for specific groups than on responsive measures during the fieldwork period. In this way, we can make better use of the large amount of respondent information that is available in a longitudinal study like SHARE, and reduce the costs of the monitoring effort.

Acknowledgements

We would like to thank Frederic Malter and Gregor Sand of the SHARE operations department as well as Johanna Bristle for their efforts in developing the monitoring system and helping us to use it as the basis for the present study.

References

Blom, A. G., and Schröder, M. 2010. "Sample Composition 4 Years On: Retention in SHARE Wave 3," by M. Schröder, *Retrospective Data Collection in the Survey of Health, Ageing and Retirement in Europe. Sharelife Methodology*, 55-61. Mannheim: MEA.

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, 992-1001. doi: 10.1093/ije/dyt088.

Börsch-Supan, A., and Jürges, H. 2005. *The Survey of Health, Ageing and Retirement in Europe – Methodology*. Mannheim: MEA.

Börsch-Supan, A., Krieger, U., and Schröder, M. 2013. "Respondent Incentives, Interviewer Training and Survey Participation," SHARE Working Paper (12-2013), Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Bristle, J., Celidoni, M., Dal Bianco, C., and Weber, G. 2014. "The Contribution of Paradata to Panel Cooperation in SHARE," SHARE Working Paper (19-2014), Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Cheshire, H., Ofstedal, M. B., Scholes, S., and Schröder, M. 2011. "A Comparison of Response Rates in the English Longitudinal Study of Ageing and the Health and Retirement Study," *Longitudinal and Life Course Studies*, 2, 127-144. doi: 10.14301/llcs.v2i2.118.

Couper, M. P., and Wagner, J. 2011. "Using Paradata and Responsive Design to Manage Survey Nonresponse," *ISI World Statistics Congress*, Dublin.

ESS5. 2010. ESS5 - 2010 Fieldwork Summary and Deviations.

Groves, R. M., and Heeringa, S. G. 2006. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs," *Journal of the Royal Statistical Society*, 169, 439-457. doi: 10.1111/j.1467-985X.2006.00423.x.

HFCS. 2013. "The Eurosystem Household Finance and Consumption Survey. Methodological Report for the First Wave," Statistics Paper Series 1, Frankfurt a. M.: European Central Bank.

Kirgis, N. G., and Lepkowski, J. M. 2013. "Design and Management Strategies for Paradata-Driven Responsive Design: Illustrations from the 2006-2010 National Survey of Family Growth," by F. Kreuter, *Improving Surveys with Paradata. Analytic Uses of Process Information*, 121-144. Hoboken, New Jersey: John Wiley & Sons.

Kneip, T., Malter, F., and Sand, G. 2015. "Fieldwork Monitoring and Survey Participation in Fifth Wave of SHARE," by F. Malter and A. Börsch-Supan, *SHARE Wave 5: Innovations & Methodology*, 102-159. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Kreuter, F., and Müller, G. 2015. "A Note on Improving Process Efficiency in Panel Surveys with Paradata," *Field Methods*, 27, 55-65. doi: 10.1177/1525822X14538205

Lepkowski, J. M., Mosher, W. D., Groves, R. M., West, B. T., Wagner, J., and Gu, H. 2013. "Responsive Design, Weighting, and Variance Estimation in the 2006-2010 National Survey of Family Growth," *Vital Health Statistics*, 2, 1-52.

Lipps, O. 2012. "A Note on Improving Contact Times in Panel Surveys," *Field Methods*, 24, 95-111. doi: 10.1177/1525822x11417966.

Luiten, A., and Schouten, B. 2013. "Tailored Fieldwork Design to Increase Representative Household Survey Response: An Experiment in the Survey of Consumer Satisfaction,"

Journal of the Royal Statistical Society, 176, 169-189. doi: 10.1111/j.1467-985X.2012.01080.x.

Lynn, P. 2005. "Outcome Categories and Definitions of Response Rates for Panel Surveys and Other Surveys Involving Multiple Data Collection Events from the Same Units," *16th International Workshop on Household Survey Nonresponse*.

Malter, F. 2013. "Fieldwork Management and Monitoring in SHARE Wave Four," by F. Malter and A. Börsch-Supan, *SHARE Wave 4: Innovations & Methodology*, 124-139. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Mohl, C., and Laflamme, F. 2007. "Research and Responsive Design Options for Survey Data Collection at Statistics Canada," in *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 2962-2968.

Peytchev, A., Riley, S., Rosen, J., Murphy, J., and Lindblad, M. 2010. "Reduction of Nonresponse Bias through Case Prioritization," *Survey Research Methods*, 4, 21-29. doi: 10.18148/srm/2010.v4i1.3037.

American Association of Public Opinion Research. 2016. *Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys*. Available from http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf.

Schouten, B. 2010. "Dynamic Adaptive Survey Designs: From Paradata to Data," *21st International Workshop on Household Survey Nonresponse*, Nürnberg.

Schouten, B., Calinescu, M., and Luiten, A. 2013. "Optimizing Quality of Response through Adaptive Survey Designs," *Survey Methodology*, 39, 29-58.

Tourangeau, R. 2015. "Nonresponse Bias: Three Paradoxes," *Keynote presentation at the 26th International Workshop on Household Survey Nonresponse,* Leuven.

Wagner, J. 2013. "Adaptive Contact Strategies in Telephone and Face-to-Face Surveys," *Survey Research Methods,* 7, 45-55. doi: 10.18148/srm/2013.v7i1.5037.

Wagner, J., West, B. T., Kirgis, N., Lepkowski, J. M., Axinn, W. G., and Ndiaye, S. K. 2012. "Use of Paradata in a Responsive Design Framework to Manage a Field Data Collection," *Journal of Official Statistics,* 28, 477-499.