Respondent incentives, interviewer training and survey participation

Axel Börsch-Supan, Ulrich Krieger and Mathis Schröder

Respondent incentives, interviewer training and survey participation

Axel Börsch-Supan[a], Ulrich Krieger[b]
and Mathis Schröder[c]

January 28th, 2013

Abstract

This paper provides an overview of an experimental study implemented in the German sub-sample of the Survey of Health, Ageing, and Retirement in Europe (SHARE). We conclude that in our setting pre-paid incentives are effective in increasing response rates, where higher incentives lead to higher response rates in an almost linear fashion. However, pre-paid incentives (regardless of the amount) lead to higher initial refusals even before the interviewer tried to contact the household. In terms of cost-effectiveness, no final conclusion can be drawn – paying the incentive is of course more expensive, but the increases in the response rate and the implied reduction in the initially necessary gross sample may outweigh these costs. Beyond mere monetary conclusions, the incentive groups did not prove to be biased compared to the control group, whereas they are significantly closer to the sampling frame in terms of their age distribution. Furthermore, using a doorstep questionnaire we show that – as far as measurable – there is no non-response bias in key variables of interest. Due to complications during the fieldwork, one initially planned experiment on Refusal Avoidance Training could not be conducted. Still we learned that interviewers do not target those respondents with highest incentive first, leading to the conclusion that their behavior is not much influenced by the incentive the respondents receive.

[a] Munich Center for the Economics of Aging, Max - Planck Institute for Social Law and Social Policy
[b] University of Mannheim
[c] German Institute for Economic Research, Berlin
Address correspondence to Ulrich Krieger, e-mail: Ulrich.Krieger@uni-mannheim.de

1. Introduction

Surveys are dependent on the individuals willing to participate in them. While there is no official account of the actual survey burden per individual, the general notion is that individuals increasingly face requests from multiple sources to participate in studies (e.g. Groves and Couper 1998, Table 1.1). As the capability to process any kind of individual data increases, the demand for such data also has grown, both for commercial and scientific purposes. The increased burden to potential respondents may affect their willingness to participate. One central measure of willingness to participate is the response rate – roughly the fraction of people participating in a study compared to those who were supposed to participate (The American Association for Public Opinion Research 2008, for details). During the last decades, response rates have been found to decline in many surveys all over the world (e.g. de Leuw and de Heer 2002, Singer 2006, Brick and Williams 2013). Since high rates of non-response may have serious consequences with regard to the representativeness of the actual population and thus on the ability to draw conclusions from the data, ways to improve participation rates in surveys are of interest as well as specifying the amount of bias introduced due to non-response.

Countermeasures against these potentially harmful developments have been taken on the dimensions of an interview as developed by Groves and Couper (1998): the survey design, respondent and household settings, the interviewers, and the interviewer-respondent interaction. While many of these measures have been tested experimentally in the United States, evidence on most of them was not very common in Germany. This discussion paper reports the efforts of a project implemented in the Survey of Health, Ageing, and Retirement in Europe (SHARE), where three controlled experiments were combined in the 4th wave of the German sub-sample. Specifically, the aim of the project was threefold (details on implementation follow in section 3):

- examine the effects of respondent incentives on survey participation,
- measure the nonresponse bias by conducting an nonresponse follow up study, and
- investigate how additional training during the fieldwork affects interviewer performance.

The SHARE survey is a multidisciplinary and cross-national European research infrastructure collecting longitudinal micro data in 20 countries. The survey targets the population of participating countries which are 50 years old or older. The main topics are health, socioeconomic status and social networks (see Börsch-Supan and Jürges 2005, for SHARE's methodological details). Four panel waves with more than 80,000 participating individuals have been completed so far, with a fifth wave starting in 2013.

In the fourth wave, a refreshment sample was added to the German SHARE study. This refreshent sample had a target size of 4,000 individuals based on a register sample. The experiments from this project were conducted using sample members of this refreshment sample. With the SHARE survey already funded, implemented and electronically monitored,

no extra fieldwork cost had to be covered by the German Science Foundation, the funders of this project. In addition, this setting had the advantage of testing the projects' hypotheses in a realistic survey environment.

This paper proceeds with a brief literature overview, followed by chapters reporting the implementation of the project, challenges met during the fieldwork and the main results from the experiments. The last section then concludes with a brief summary.

2.   Literature review

2.1.   Incentives in Surveys

Incentives (of various kinds) are one of the possibilities to target the respondent directly when trying to influence the respondent's decision to participate (see Groves and Cooper 1998). The use and the effect of respondent incentives vary on three dimensions:

- the type of survey (interviewer-facilitated or not),
- the type of incentive (monetary or not),
- the mode of incentive (conditional or not)

The use of incentives has been shown to be effective in the United States, both for mail surveys (Church 1993) as well as in interviewer-mediated surveys (Singer et al. 1999). Both studies are meta-analyses comparing 37 (Church) and 39 (Singer et al.) studies, respectively. The main findings strongly support the use of incentives (see Singer 2002):

In mail surveys,

- Unconditional incentives increase response rates (compared to conditional ones),
- Unconditional monetary incentives are more effective than unconditional gifts,
- Response rates increase (linearly) with the amount of money.

In interviewer-mediated surveys,

- Incentives increase response rates in telephone and face-to-face surveys in a linear fashion, although the effects are smaller than in mail surveys,
- Unconditional incentives do not differ significantly from conditional incentives (although the evidence is weak due to the small number of studies with unconditional incentives),
- Money is more effective than gifts,
- Although the effect of the incentive increases with the burden, even in low-burden studies the incentive is effective,
- Incentives are more effective in studies with otherwise low response rates, which could be evidence that incentives can provide a motive for participation.

While these two meta-analyses focus on the US literature, there have been several studies over the last years in German speaking regions of Europe which are concerned with the effects of incentives. The US findings for incentives in mail surveys could be replicated in

several studies. Harkness et al. (1998) report on the experience in two ISSP studies using different "near-cash" incentives – one with stamps, the other with lottery tickets. They find increases in response rates of five and three percentage points, respectively. Arzheimer and Klein (1998) find that the in-kind incentive (a telephone card) not only increases the response rate by about eleven percentage points, but also increases the rate at which the survey is sent back. For Switzerland, Diekmann and Jann (2001) report a similar result when using a phone card as incentive in their mail survey. They find that the unconditional incentive yields significantly higher response rates than both the control group and the group with conditional incentives.

More recently, Scherpenzeel and Toepoel (2012) report results from a study in the Netherlands aiming at the response rate when setting up an online panel. In their experiment, not only the interview mode varied between a telephone and a face-to-face approach, but also the sum of the (purely monetary) incentive was varied between 10, 20 and 50 Euros. The authors find that unconditional incentives are more effective compared to conditional incentives, but they could not replicate the linear relationship between incentive size and response rate found in the US. Their conclusion is that even a low prepaid incentive amount (10 Euros) suffices in increasing the response rates, while being more cost effective than all other variations (including the one without incentives). In another application of a face-to-face study in Germany, Castiglioni et al. (2008) also test the use of incentives in an experiment. Interestingly, and in contrast to other results, they find that the conditional incentive actually yields higher response rates than both the unconditional and the no-incentive treatments. They try to explain this rather counterintuitive finding with the role of the interviewer, and point at the importance of training: "[…] poor quality in the interviewer pool cannot be compensated by incentives".

All in all, while many studies report positive effects on incentives, the applicability of these results may still depend on the survey environment (content, sample, country) and hence Singer and Kulka's (2000) recommendation to test whether incentives work in a specific setting was followed in this project. The original distribution included 4,000 cases in three different groups with pre-paid incentives of 10, 20 and 40 Euros, as well as a control group without any incentives.

2.2. The role of the interviewer

Interviewers in interviewer-mediated studies receive basic training in how to approach a household, how to read different cues during the initial contact and how to convince potential respondents to participate. But they also receive specific training on the study they are conducting, so that they can inform the respondent about the goals and purpose of the study. However, the remaining interviewer idiosyncrasies are likely to influence survey participation. Since there is no personal interaction (either by phone or face-to-face) with the respondent, the issue of interviewer effects does not arise for mail or internet surveys.

Interviewer training is an essential part for interviewer-mediated surveys. As Groves and Couper (1998, Chapter 7) point out, the interviewer has a potentially large influence on the respondent's decision to participate in a survey. Groves and McGonagle (2001) have tested the influence of training and especially the "Refusal Avoidance Training" (RAT). They found not only that the response rates increase after such a training, but also that the variation between interviewers is reduced dramatically. Mayer and O'Brien (2001) find that interviewers who went through an extra training concerning refusals also have a significantly higher response rate on first contact. (Their experiment does not allow an evaluation of the full response rate.) For Germany, Schnell and Trappmann (2006) find that a two-hour RAT for the European Social Survey yields a response rate difference of 59.3 percent vs. 56.8 percent (significant on a 10% level) in favour of the extra trained. However, while the RAT-sample reaches a 7.2% lower refusal rate for those individuals contacted, these interviewers have a 5.7% significantly higher non-contact rate. They explain this effect by the non-randomness of the allocation of interviewers across sample points and especially the higher workload of RAT interviewers. They conclude that "[…] to clarify the effects of a Refusal Avoidance Training, a further experiment is needed. A higher number of interviewers and a random allocation would be desirable" (Schnell and Trappmann 2006, 9). Even though SHARE conducts extensive interviewer trainings in every wave (see for example, Alcser and Benson 2005), RAT procedures are only a small subsection and their effectiveness has not been tested. In this project, we originally planned to randomly choose interviewers to participate in an RAT after the initial four weeks of fieldwork, to afterwards elicit the RAT effect on response rate and interviewer variation.

2.3. Non-response studies

The main concern of low response rates is the potential for non-response bias, i.e. biases brought about for key variables of the survey. However, as Keeter et al. (2000) showed in their experiment, even differences in response rate as large as 25 percentage points may not lead to significant differences in key variables. Similarly, Groves and Peytcheva (2008) show in their meta-analysis that the size of the bias does not necessarily depend on the amount of non-response to a survey. Groves (2006) mentions non-response follow-up studies as one of several possibilities to assess and quantify the non-response bias. This involves a detailed analysis of the contact patterns and the early result codes for all respondents, comparing those who initially were reluctant with those who immediately agreed to be interviewed. In addition, follow-up studies can entail contacting (a random subset of) non-respondents to assess reasons for non-response and some key variables of the study (e.g. Beaulieu 2006, Willimack and Dalzell 2006). Rässler and Schnell (2003) suggest that instead of a non-response follow-up study that tries to re-interview non-respondents, multiple imputations may be more cost effective. However, the authors also acknowledge that their recommendation needs to be tested in other datasets and for different variables. In this project, the goal of the non-response study was to elicit potential non-response biases on a few important variables.

3. Experimental design

3.1. Incentive Experiment

When the project was started in September 2010, the threefold design of the experiments was finalized in cooperation with the team running the SHARE survey at the MEA and the agency contracted for fielding the survey, the Institut für angewandte Sozialforschung, Bonn (infas).

The first part of the study evaluates if prepaid cash incentives do indeed increase cooperation rates and how these rates are influenced by different amounts of cash incentives. The implementation of this experiment is no trivial matter as unconditional prepaid cash incentives are uncommon as a tool to increase response rates in Germany. In contrast to the original plan of running the incentive experiment double blind (i.e. neither interviewer nor respondents would know about the experiment), we decided to run the experiment half blind, informing interviewers of the treatment condition of respondents, but leave the respondents uninformed. The reasons for this decision were twofold. First, the agency had doubts about the practical implications of not informing the interviewers about such a considerable design feature. They deemed it important for the interviewers to be fully informed about all aspects of the survey to present the study correct and positive at the doorstep interaction when establishing contact. If respondents with questions about the incentives they received meet an interviewer who is completely unaware of this particular study feature, this may reflect badly on the interviewer, the agency, the survey sponsor and the study as a whole.

Second, informing interviewers about the incentives creates a more realistic scenario. When using incentives in subsequent surveys (without experimental control), interviewers are well aware of the presence of incentives. They can refer to the incentives and use those as a reference when making contact. So in a real setting, incentives will always jointly affect interviewers and respondents. Running the experiment blind on the respondents' side only increases the transferability of our findings to future survey research projects.

The entire refreshment sample was divided in four different batches to be fielded sequentially. The batches were sized so that the target size of 4,000 cases would be realized if a response rate of 55 percent (5,247 addresses), 45 percent (1,050 additional addresses) or 40 percent (840 additional addresses) could be realized during fieldwork. 2,098 addresses were kept as a reserve. To allow sample members in the treatment group as much time as possible in the field, the experiment was being run in the first batch of addresses that were send out right at the beginning of field time.

Sample members from sample points in communities with less than 9,000 inhabitants were excluded from the experiment because there was a considerable risk of respondents finding out about other respondents' incentives, and thus confounding the treatment conditions in these smaller towns and villages. Out of a total of 210 sample points the refreshment sample was drawn from, 54 sample points were excluded for this reason.

After the addresses had been drawn from the register, the survey agency delivered the sample to SHARE and the project team. Experimental conditions and control group were randomly allocated in all of the 156 larger sample points in the survey. Therefore, all sample points in

the experiment contained sample members of the three experimental conditions and the control group. As a consequence, the experiment was run in all German regions and all interviewers had sample cases in all experimental conditions, as interviewers usually work in a certain sample point. Table one gives a summary of the experimental setup, where the size of the treatment groups was chosen according to considerations of statistical power for the significance tests.

Table 1: Summary of sample sizes

	Sample points	**Addresses**
Full sample	210	9,235
First batch	210	5,247
Excluded (small communities)	54	1,347
Part of the experiment	156	3,900
Experimental conditions		
40 Euros	156	750
20 Euros	156	750
10 Euros	156	1,375
No prepaid Incentive	156	1,025

3.2. Nonresponse follow-up

As second part of the project, the interviewers administered a short 1-page questionnaire to those respondents who refused to participate in the main SHARE survey. Drawing on work by Lynn (2003) on the Pre-Emptive Doorstep Administration of Key Survey Items (PEDAKSI) method and following the example set by a study on a doorstep questionnaire survey (DQS) in the European Social Survey (Matsuo et al. 2010), this questionnaire included basic demographics and five items from the main survey. The goal of this short questionnaire was to compare responses of respondents in the SHARE survey to respondents that only answered the questions of the short questionnaire. Without the doorstep questionnaire these respondents would have fallen in category of non-respondents, thus the approach allows to draw limited conclusions on non-respondents and nonresponse bias. In the planning stage, four questions (one with a follow up) were selected that closely resemble key survey questions from the main survey:

- How many persons live in this household?
    - If more than 1 person: Are children or a partner among them?
- How many children do you have?
- How would you rate your health status?
- What would best describe your current occupational status?

The questionnaire was designed to fit on less than one page to signal respondents the ease and little effort to complete. Interviewers were advised to approach all sample members who finally refused participation in the main survey with a request to participate in the

nonresponse study. Completed questionnaires were then sent back to the survey agency, coded electronically and made available to the project team.

3.3. Interviewer training

An experiment on additional in-field interviewer training was originally planned as the third part of the project. We successfully trained — in cooperation with infas — a total of 183 interviewers for the SHARE study. The contract for the study demanded those interviewers to be experienced in face-to-face interviews and trained in general interviewing techniques. During the three-day SHARE-training those interviewers were reminded of those general interviewing techniques, measures to build rapport with respondents and present the study accurately and positively. They were advised on good practice tips to avoid initial refusal and increase respondents' cooperation while respecting the voluntary decision of each respondent to participate or not.

As a third experimental treatment in this project, a refreshment training was scheduled four weeks into the fieldwork. By then, all interviewers should have been actively involved in collecting data and thus have first-hand experience with the 4th wave of the SHARE survey. About half of the interviewers were to be randomly selected and supposed to be invited by infas for a full day of interviewer training. In this training, the interviewers should have been reminded of interviewer techniques, review their approach on initial contact with survey respondents, and discuss best practice procedures in gaining cooperation with the respondents and conducting the interview. The outcome measure of the experiment was planned to be the difference in cooperation rates among the interviewers prior to the treatment and those with an additional day of training. A control group of interviewers without additional training was planned to serve as a reference group to control for effects of the duration of the fieldwork as later contacts may not be an exact replica of addresses issued earlier in fieldwork.

3.4. Ethics review

It is important to point out that all experiments described so far (as all other SHARE fieldwork procedures in general) have been submitted to, and approved by, the ethics committee at the University of Mannheim, which was the legally responsible entity for SHARE at that point.

4. Realization of the project

Unfortunately, the survey operations conducted by infas did not proceed as planned. We encountered severe capacity limitations, huge delays, and a massively reduced number of interviewers supplied by infas to SHARE. Similar and sometimes even more severe restrictions were imposed on parallel surveys -- such as the Panel on Household Finances commissioned by the German Bundesbank and the Panel Study "Labour Market and Social Security" (PASS) commissioned by the Institute for Employment Research.

From the very beginning, the fieldwork progress was slow. Of the 183 trained interviewers only 49 showed activity as of May 2011, meaning that they reported contact attempts or other events to the agency. All others were "silent". SHARE and the project team raised concerns early, warning about the slow field progress and the possible consequences for the refreshment sample. In late May, senior members at infas described the panel fieldwork as 'behind plan' but stated that the necessary steps would be taken to ensure that all addresses would be contacted by July 31st. To accommodate further difficulties, SHARE then rescheduled the end of main fieldwork for the refreshment sample to August 30th. In subsequent conversations with infas and observing the progress rate of fieldwork it became clear that even those targets could not be reached. The main reason for this was, again, the low levels of activity from interviewers in the field. For example, between June 10th and June 24th, only 30 interviewers (out of 183 trained) were actively working for the SHARE survey.

Both contact and cooperation rates were subsequently suffering from this fact. In a telephone conference in mid-July, infas admitted having problems with the quality of their interviewers. Only about 60 interviewers were actively working and producing the desired results. The agency had little hope of reactivating interviewers that were not working. Keeping in mind that interviewers at infas and other survey agencies are not employed by the agency but work as freelancers ("Werkvertragnehmer"), there is no direct way for the agencies to tell interviewers when to conduct their interviews. The decision was reached to concentrate all efforts on the active interviewers and push for higher activity in the month of August. As a result of efforts by infas, a surge in interviewer activity and completed interviews could be observed in late July. For a short time, 88 (out of 183) interviewers were actively working for SHARE. Considering the total number of trained interviewers this number was still insufficient to meet fieldwork deadlines.

In August 2011 the German SHARE team and the SHARE central coordination team came to the conclusion that under the conditions provided by the fieldwork agency, the refreshment sample could not be completed in the foreseeable future. Because SHARE is an international survey enterprise with closely matched project schedules across 20 countries, there was no room to allow for an indefinite procrastination. Data production deadlines had to be met and the start of the next panel wave in January 2013 meant a limited time for the completion of the fieldwork. Of major concern was the threat to the completion of the longitudinal sample (which was in the field parallel to the refreshment sample). Because SHARE is a longitudinal infrastructure project, the longitudinal information collected from sample members who are with the survey for up to eight years has to be prioritized. Given the sub-optimal performance by the survey agency, the lack of improvement of the situation and the fear for the continuity of the longitudinal sample, the decision was reached to interrupt the fieldwork in the refreshment sample as of August 30th and advise infas to instead focus entirely on the longitudinal sample. Progress in the panel sample remained slow and was completed in June 2012, 15 months after fieldwork commenced. In March 2012, the SHARE coordination thus came to the conclusion to not resume fieldwork in the refreshment sample as the German sample was out of the common SHARE schedule.

In addition to these interviewer related issues, there was another unforeseen development in the field: From the start of the fieldwork in early June, when the first advance letters reached the sampled households, the survey agency registered an unusually high volume of calls to the hotline mentioned in the letter. During the first week of fieldwork, a total of 85 calls were answered. In a memo sent to the SHARE team, infas reported about calls of anxious respondents who were unhappy with the inclusion of cash in their advance letter. Most of those people calling the hotline questioned the legitimacy of the survey, some even felt pressured into the survey – quite opposite to the effect intended. Note, however, that the overwhelming majority of the respondents (almost 94 percent, see below) did not voice concerns. In response to the complaints, the hotline staff spoke to respondents and also to police, lawyers, the Arbeitskreis deutscher Markt- und Meinungsforschungsinstitute (the business association representing the private-sector market and social research agencies) and representatives of municipalities that received complaints by sample members. One complaint to a member of the Bundestag reached high levels at the Federal Ministry of Education and Research, the funders of SHARE in Germany.

Although the infas staff could reassure most callers of the sincere motives of the project, at the time infas informed the project team at SHARE, already 3,065 advance letters of the 5,421 respondents from the first batch had been mailed to the households, of which 1,723 contained a cash incentive. As a result of the infas memo, the advance letter was changed, putting even more emphasis on the voluntary nature of the survey and the unconditional gift character of the incentive. In addition, a second advance letter was mailed to the 3,065 sample members who had already received the original version of the advance letter. There was little indication that this new letter helped to reduce skepticism among potential respondents: Hotline calls remained frequent during the mailing of the remainder of the advance letters and only dimmed off 5 weeks into the field period when all letters had reached sample members.

Although the main goals of the project could be achieved, this necessary decision had severe implications for the research project. First and foremost, the number of observations is lower than planned, as only 1,900 rather than 3,900 sample cases have been contacted. Moreover, not all SHARE quality standards have been adhered to by infas (e.g. a minimum of 8 personal visits to sample members). This hurts the incentive experiment as well as the nonresponse follow-up study. A second major implication of the fact that interviewers did not work as planned was the cancellation of the interviewer training experiment. There were three reasons for this decision:

- The number of active interviewers working on the SHARE survey was consistently very low, so the cell sizes in the experiment would have been too low.
- During the time window for the experiment in mid-July the agency was preoccupied with trying to save the study as a whole.
- Once the decision was taken to halt the fieldwork in August there would have been no time to measure the effectiveness of the training on response rates of the interviewers.

In summary: While parts one (incentive experiment) and two (nonresponse follow-up) could be realized despite the reduction in sample size, the capacity limitations of infas made the realization of the third part (interviewer training) impossible.

5. Results

5.1. Incentive Experiment

The main effect of unconditional prepaid cash incentives on response rates was overall positive, and this much more than expected. Controlling for age, gender and population size of the community of the sampling point, sample members receiving an incentive agreed more to the survey request than those respondents from the control group. This analysis is restricted to addresses that have a final disposition code (out of sample, refusal or interview) or could not be contacted during the field period while exhausting the minimum number of 8 personal visits to the housing address. Out of the 3,900 addresses originally in the experiment this condition is met for 2,241 households. As variable of interest we choose the cooperation of households. While SHARE is a survey of individuals and incentives have been targeted to the respondent drawn from the register, the survey also targets the cohabiting household members. As the incentive may also have an influence on partners we look at cooperation of at least one household member. Table 2 shows the results from the three treatment conditions and the control group.

Table 2: Effect of incentives on cooperation

Incentives		Not Cooperating	Cooperating	Total
Control	**(N)**	398	150	548
	(%)	72.6	27.4	100
€10	**(N)**	494	307	801
	(%)	61.7	38.3	100
€20	**(N)**	258	178	436
	(%)	59.2	40.8	100
€40	**(N)**	209	247	456
	(%)	45.8	54.2	100

The proportion of sample members cooperating with the survey request increases monotonously with the amount of incentives paid. The control group has the lowest cooperation propensity of the four groups. The difference of the 40€ group to the control group is about 25 percentage points with regard to the proportion of households that cooperate. Paying 10€ cash unconditionally to sample members still increase the likeliness of taking part in the survey by over 10 percentage points compared to the control group that received no incentive upfront. While the difference between 10€ and 20€ unconditional cash incentive is not statistically significant (the difference being about 2 percentage points in likelihood to participate) the results show clearly that paying larger incentive amounts increases the participation to a larger extent than the smaller cash amounts. Moreover, and

11

most importantly, all differences to the control group are statistically significant. We observe a percentage point difference of over 13 percent between the 20€ and the 40€ incentive treatments. This contradicts the work by Scherpenzeel and Toepoel (2012) who found a diminishing return of large incentive offerings in the LISS panel. Our findings are in line with a meta study conducted by Singer et al. (1999) that found a significant linear effect for incentive sizes.

To check if these findings hold in multivariate analysis a fixed-effects logistic regression on household cooperation with the interview request was conducted. As households are nested in interviewers, we use interviewer fixed-effects. Explanatory variables in the regression are gender, age group, having received a second version of the advance letter (see section 4), and the incentive treatment. Population size of the municipalities of the sampling points is controlled by adding dummy variables for the lower third (9,000-35,0000 inhabitants) and the upper third of the population distribution (more than 200,000 inhabitants) respectively. Table 3 shows the results.

Table 3: fixed-effects logistic regression on household cooperation

Odds ratio coefficients (Standard Errors in brackets)

10 €incentive	1.91***
	(0.27)
20 €incentive	2.18***
	(0.38)
40 €incentive	4.07***
	(0.72)
Male target person	1.25*
	(0.14)
50-54 years old	1.10
	(0.14)
2nd Version of advance letter	0.89
	(0.20)
Smaller sample point	0.49*
	(0.15)
Large sample point	0.93
	(0.25)

N: 2,751, McFadden's R²: 0.05, 107 interviewers

Besides the effects of the incentive treatment, two control variables yield significant results: gender and living in smaller municipalities. Households of male target persons have 25 percent higher odds to cooperate than if the target person is female. The odds for cooperation of households in communities with less than 35,000 inhabitants are reduced by about 50 percent compared to households from towns that have between 35,000 and 200,000 inhabitants. Regarding the main effect, we show the contrasts of the control group to the treatment groups. All those effects are positive and significant, meaning that the odds of responding are larger for all incentive groups than the control group. For the 10€and 20 €, the

12

odds are about twice as large as in the control group, for the 40 € group the odds of responding are four times as large as in the control group.

The positive effect of incentives on response propensity can thus also be shown in multivariate analysis. But given that fieldwork had to be terminated early, one possible explanation for these positive effects could be a selective interviewer effort, if interviewers try to primarily contact the sample members from the incentive conditions first. Given that interviewers are aware of the incentive treatments and they are primarily paid by completed cases, interviewers have strong incentives to focus their efforts on those cases that have been offered an incentive due to the anticipated higher chances of a successful interview.

We therefore analyze the effect of the incentive treatment on the contact propensity of households. If interviewers have prioritized contacting incentivized households we expect to find effects of incentives in this analysis. As above, we use gender, age, the advance letter version and sampling point population as control variables. Individual differences in interviewers' contact success are again controlled for by using interviewer fixed effects. Table 4 below shows the results of the fixed-effects logit regression on contact. For this analysis all 3,900 cases from the experiment are included in the analysis, however, 778 addresses from 29 interviewers have to be dropped from the analysis due to lack of variation on the dependent variable: these interviewers either contacted all addresses or no addresses at all.

Table 4: Fixed-effects logistic regression on household contact

Odds ratio coefficients (Standard Errors in brackets)

10 €incentive	1.05
	(0.18)
20 €incentive	0.96
	(0.18)
40 €incentive	1.01
	(0.20)
Male target person	0.87
	(0.10)
50-54 years old	0.58***
	(0.07)
2nd Version of advance letter	1.16
	(0.27)
Smaller sample point	0.62
	(0.25)
Large sample point	1.70
	(0.72)

N: 2,824, McFadden's R²: 0.02, 88 interviewers

The only significant effect in this analysis is the negative effect of being in the age cohort of 50 to 54 years. The odds of those respondents being contacted are 38 percent smaller than those of the older respondents. There is no significant effect of the incentive treatment groups,

indicating that the contact propensity does not depend on whether a prepaid incentive has been sent or not. Hence it seems unlikely that the interviewers used the information on the incentive treatment for their contact decision. This is plausible since we know from talking to the agency and the interviewers that driving distances between households is their foremost consideration when planning their work. While interviewers get some compensation for travel cost, their main income comes from finalized interviews. Thus, they try to optimize driving distances between addresses. Given that incentivized addresses are distributed randomly, incentives did not affect contact strategies by interviewers.

As mentioned in section 4, the unconditional incentive treatment also had negative consequences: an unusual number of respondents were worried about the sincerity of the project and called – among other places – the hotline mentioned in the advance letters. A major concern about these hotline calls were sample members who refused their participation right away. In those cases, no interviewer can be sent out to try to gain participation. 197 respondents from the 3,900 addresses in the experiment refused via the hotline, accounting for about 50 percent of total call volume to the hotline. The distribution in regard to treatment groups is shown in Table 5.

Table 5: Hotline refusal by treatment groups

	Control	10€	20€	40€	Total
No. of refusals	16	86	45	50	197
Number of addresses	1,050	1,350	750	750	3,900
Percentage of hotline refusals	1.5	6.4	6.0	6.7	5.1

Table 5 clearly shows that refusals via hotline occurred more often among the incentive groups. The proportion of sample members refusing is about equal across treatment groups at about 6.5 percent, the differences being not significant. This result may be interpreted as an indication that the increased resentment towards the survey in this group is due to the sending of cash in principle, not the amount of cash. Refusal in the control group is considerably lower at 1.5 percent of sample members. The loss of theses sample members has to be put into perspective to the large gains in cooperation in treatment conditions. Moreover, while one has to take the hotline refusals seriously, the great majority (almost 95 percent of sample members) did not voice concerns about the incentive treatment and cooperated more than the members of the control group with the survey request.

To examine the cost effectiveness of each additional Euro spend on response rates, we show the marginal cost of additional interviews realized in the incentive groups and the average interviewer cost dependent on incentive group. Here we standardized the size of the experimental groups to 1,000 contacts each.

Marginal cost of each additional interview realized are the amount we spend more per additional interview realized in the incentive groups in comparison to the control group.

Figure 1 shows the results. For the 10 € group each additional interview compared to the control group cooperation costs on average 92 € This value is obtained as follows. Suppose we start with a gross sample of 1,000 individuals. The costs for the cash incentives are then 10,000 € The net sample without incentives would be 274 individuals according to the response rates in Table 2, while with a 10€ incentive it would be 383. Thus, 109 additional interviews would be generated by the incentives. Dividing the costs (10,000€) by the difference of the number of realized interviews (383 – 274 =109) yields 92€

The procedure is the same for the other incentive amounts. By coincidence the marginal cost of additional interviews in the 20€ and 40€ are identical at 149€ Figure 1 shows that the 20€ treatment is not cost efficient as it yields a lower cooperation rate than the 40€ group given identical marginal costs.

Figure 1: Marginal cost for add. interview over incentive groups
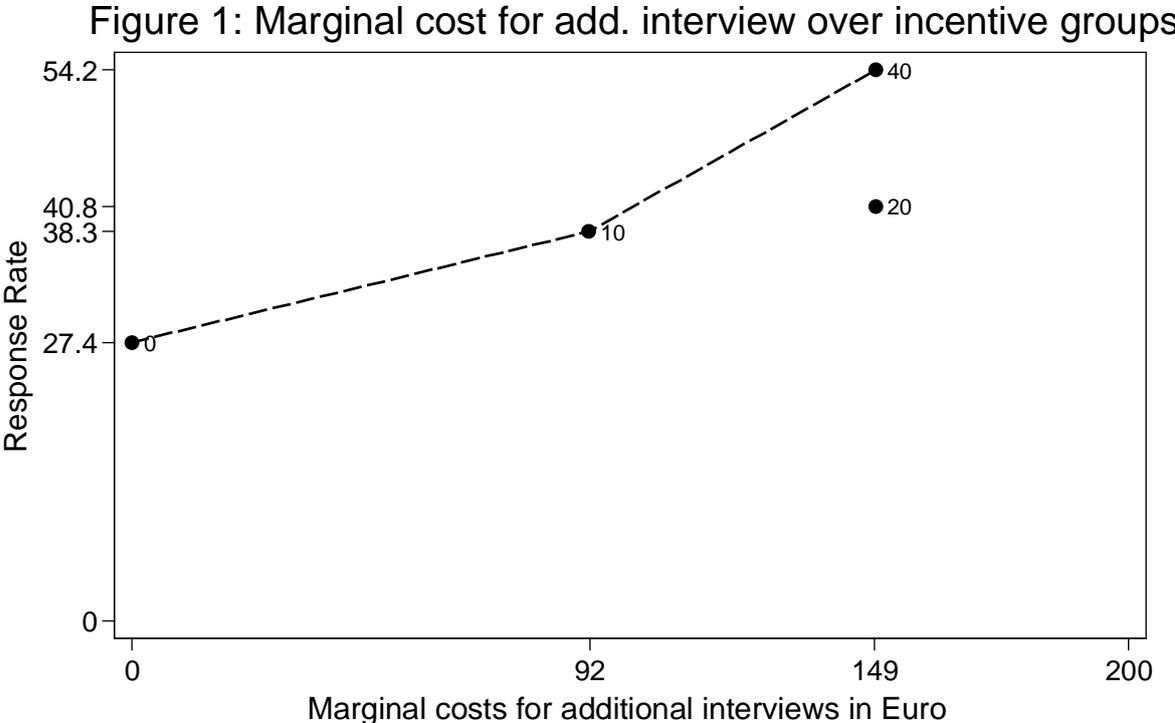
The average cost of a SHARE interview in Germany — without the incentive experiment — is about 230€ This number includes all fix and variable cost at the survey agency as well as the SHARE team in Munich. We add to this number the increased average cost of realized interviews in the incentive groups. Results are shown in Figure 2.

Spending 10 € on unconditional incentives increases the average cost of realized interviews by 26€ (1,000 contacts, 10,000€ spent on incentives, 383 realized interviews). Spending 40€ on incentives adds 74€ to the average interview costs, increasing the cooperation rate to 54.2 percent.

With higher response rates the survey cost will be lower as sample sizes and interviewer staff can be smaller. However, these potential cost savings are not forwarded towards the contractors as of now. The fixed and variable cost of interviews will typically be the same irrespective of the incentive offered.

In summary, the experiment shows an impressive increase in the cooperation rate, doubling it from about 27 to about 54 percent. At the same time, incentives carry substantial costs in terms of money and complaints. The tradeoff between response rates and costs is therefore a very serious one.


Figure 2: Average interview cost over incentive groups
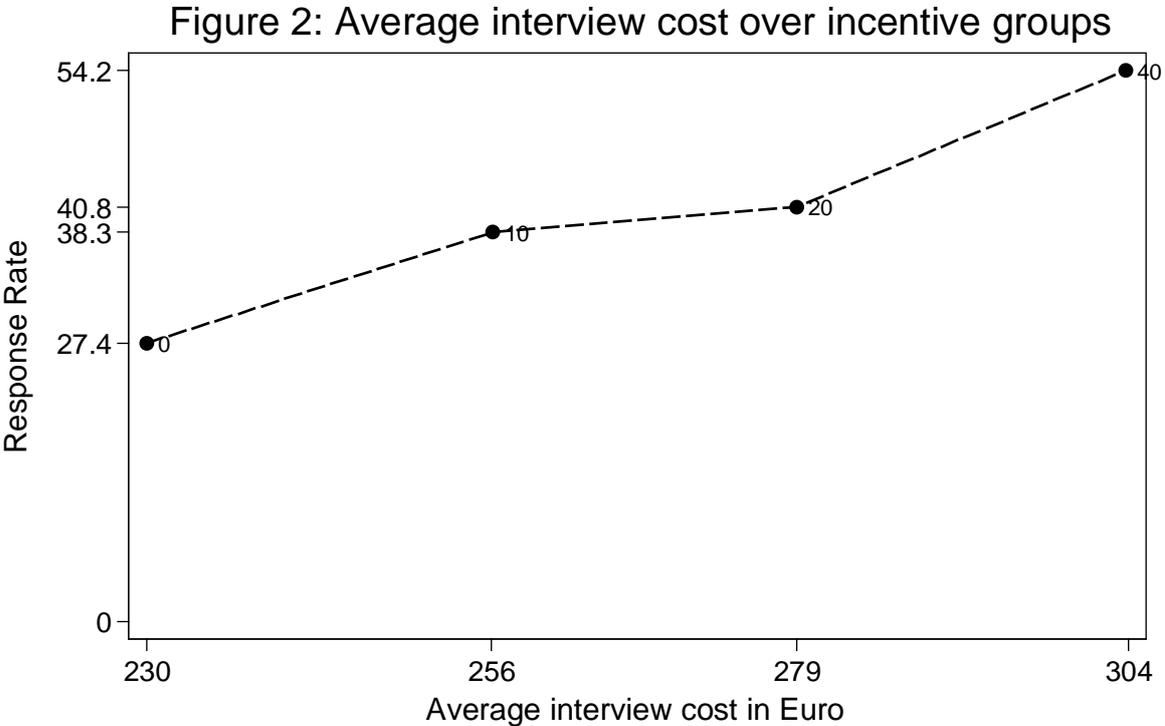
5.2. Nonresponse follow-up

Data from the nonresponse follow-up were delivered to the project team in late June 2012. As in the incentive experiment, we only analyze finalized cases. Out of the 2,990 finalized addresses, 889 addresses were eligible for the "Doorstep Questionnaire Survey" (DSQ) as a consequence of a hard and final refusal to the SHARE survey. Of all cases eligible, 197 respondents completed the DQS. This yields to a response rate of 22.2 percent in the DQS survey. This response rate is similar to the findings of Lynn (2003), who reported a cooperation rate of 25 percent. It is lower than the response rate in the DQS survey run in the European Social Survey of 44.7 percent. We conjecture that the early end of fieldwork has played an important role here as well, as hard refusals and requests for a doorstep questionnaire survey participation do typically occur in later stages of the fieldwork when all other attempts to gain respondent cooperation to the main survey have failed.

For a comparison of SHARE and DQS respondents we focus on the sample of those addresses that have been part of the experiment. Figure 3 gives an overview of the results split by the experimental conditions: Incentivized cases are compared to the control group.

Figure 3: Result of SHARE and Doorstep Questionnaire surveys.
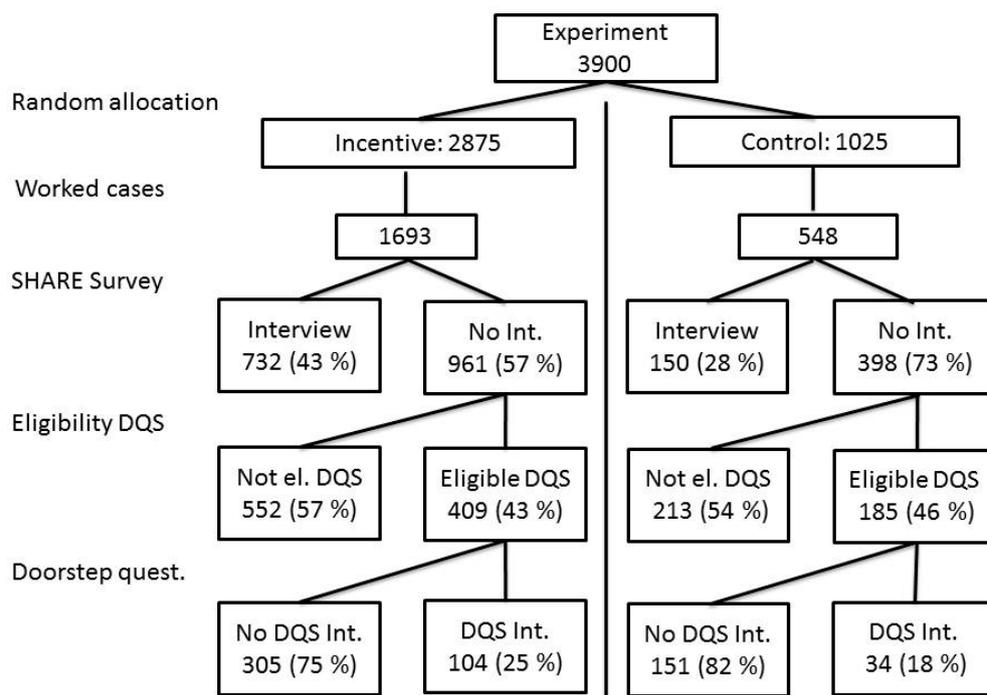
59 percent (1,693) of incentive cases have been worked in the survey, compared to 53 percent (548) of control group cases. The slight difference in the proportion of addresses that have been worked by the interviewers can be explained by the higher cooperation rate among the incentivized cases as those cases are not at risk for still being in the contact process.

When comparing those sample members who took part in the DQS to those that refused, we found no significant effects of age and gender (analysis not shown). The cooperation rate is larger among those respondents that received an incentive.

Having information from the respondents of the DQS allows for a comparison of their answers to SHARE respondents. As DQS respondents are SHARE non-respondents this allows to examine the size of the nonresponse bias in the survey. We can show the difference of answers to key survey questions between respondents and non-respondents. In combination with the incentive experiment this brings the added bonus to compare non-respondents to SHARE respondents with and without a prepaid incentive treatment. We can thus examine, if the difference between respondents and non-respondents is greater when incentives are paid.

Table 6 gives the result of a comparison of means between SHARE respondents with and without incentive treatment, the DQS respondents and those that refused both requests. These three groups are split into treatment and control group. The information about gender and age are available for all targeted households as this information was provided with the register sample. Health, occupational status and number of children are information taken from the SHARE interview or the Doorstep Questionnaire. Household income is available for those respondents that completed the SHARE interview only.

Table 6: Comparison of means between groups (standard deviations in parentheses)

	Incentive			No Incentive		
	SHARE	DQS	Frame	SHARE	DQS	Frame
Male	.49 (.50)	.49 (.50)	.44 (.50)	.49 (.50)	.45 (.51)	.43 (.50)
Age 50-55	.17 (.38)	.20 (.40)	.18 (.38)	.24 (.43)	.21 (.42)	.18 (.38)
Health: fair	.30 (.46)	.21 (.41)	-	.24 (.43)	.15 (.36)	-
Health: bad	.13 (.33)	.17 (.38)	-	.12 (.33)	.15 (.36)	-
Working	.31 (.46)	.22 (.42)	-	.36 (.48)	.29 (.46)	-
Number of children	1.86 (1.35)	1.65 (1.28)	-	1.89 (1.39)	1.4 (.88)	-
Low school education (Hauptschule)	.49 (.50)			.45 (.50)		
High school (Abitur)	.17 (.37)			.18 (.39)		
Vocational training	.54 (.50)			.58 (.49)		
University education	.21 (.41)			.28 (.45)		
Voluntary work	.20 (.40)			.19 (.39		
Social activities	.40 (.49)			.41 (.49)		
Left right Scale	4.96 (1.90)			4.88 (2.05)		
Praying often	.25 (.43)			.22 (.41)		
Cognitive ability	.73 (.44)			.73 (.44)		
retired	.51 (.50)			.43 (.50)		
HH income	4737 (7635)	-	-	5028 (7088)	-	-
Observations	732	104	754	150	34	364

When comparing SHARE respondents with and without incentives, a significant difference in age composition can be observed. The proportion of younger sample members (50 to 55 years) is larger and thus further from the sampling frame proportion among the control group (t=1.9). There are no significant mean differences between the two DQS groups. When comparing the incentive group with the non-incentive group, two larger differences are noteworthy: The proportion of retired respondents and those without university degree are smaller among the incentive group, the difference being 8 and 7 percentage points, respectively. These differences are significant on a ten percent level (t=-1,79, t=1,95, two-sided tests). Incentivized respondents report a mean household income that is almost 300€ lower than the mean household income reported by control group households, although this difference is not significant. Further research with the data will be necessary do determine if this difference is an indication of low income populations being brought into the survey as mentioned in the literature (e.g. Singer and Kulka 2000).

Are respondents different from non-respondents? When comparing gender and age composition of SHARE and DQS respondents to those that did not complete an interview, the only significant mean difference is the larger portion of male target persons among SHARE respondents without incentives compared to the non-respondents (t=1.9).

The mean differences between SHARE respondents and DQS respondents are not significant when controlling for experimental condition. Once incentive and control group are combined, the DQS group shows a significantly lower percentage (by 9 percentage points) of respondents who rate their health condition as fair (t=-2.3).

In summarizing the results we conclude that there is little evidence for nonresponse bias amongst the variables we looked here. However the small number of observed cases, especially in the control group does hurt this analysis. Most importantly, the evidence presented here does not support the notion that the incentive treatment introduced a bias into the study.

6. Summary

This paper provides an overview of an experimental study implemented in the German sub-sample of the Survey of Health, Ageing, and Retirement in Europe (SHARE). We conclude that in our setting pre-paid incentives are effective in increasing response rates, where higher incentives lead to higher response rates in an almost linear fashion. However, pre-paid incentives (regardless of the amount) lead to higher initial refusals even before the interviewer tried to contact the household. In terms of cost-effectiveness, no final conclusion can be drawn – paying the incentive is of course more expensive, but the increases in the response rate and the implied reduction in the initially necessary gross sample may outweigh these costs. Beyond mere monetary conclusions, the incentive groups did not prove to be biased compared to the control group, whereas they are significantly closer to the sampling frame in terms of their age distribution. Furthermore, using a doorstep questionnaire we show that – as far as measurable – there is no non-response bias in key variables of interest. Due to complications during the fieldwork, one initially planned experiment on Refusal Avoidance Training could not be conducted. Still we learned that interviewers do not target those respondents with highest incentive first, leading to the conclusion that their behavior is not much influenced by the incentive the respondents receive.

7. Acknowledgements

8. Literature

Alcser, Kirsten H., and Grant D. Benson, 2005. "The SHARE Train-the-Trainer Program." In: Axel Börsch-Supan and Hendrik Jürges (eds) *The Survey of Health, Ageing and*

Retirement in Europe. Methodology, Ch. 6. Mannheim, Mannheim Research Institute for the Economics of Aging (MEA).

Arzheimer, Karl, and Markus Klein. 1998. "Die Wirkung materieller Incentives auf den Rücklauf einer schriftlichen Panelbefragung." *ZA-Informationen* 43: 6-31.

Beaulieu, Martin. 2006. "A Study of Nonrespondents in the Canadian Vehicle Survey." 2006 Proceedings of the Survey Research Methods Section, American Statistical Association.

Börsch-Supan, Axel and Hendrik Jürges (eds), 2005. "The Survey of Health, Ageing and Retirement in Europe. Methodology." Mannheim, Mannheim Research Institute for the Economics of Aging (MEA).

Brick, J. Michael, and Douglas Williams. 2013. "Explaining Rising Nonresponse Rates in Cross-Sectional Surveys." *ANNALS of the AAPSS* 645: 36–59.

Castiglioni, Laura, Klaus Pforr, and Ulrich Krieger. 2008. "The Effect of Incentives on Response Rates and Panel Attrition: Results of a Controlled Experiment." *Survey Research Methods* 2: 151-158.

Church, Allan H.. 1993. "Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis." *Public Opinion Quarterly* 52: 62-79.

De Leuw, Edith and Wim de Heer, 2002. "Trends in Household Survey Nonresponse: A Longitudinal and International Comparison." In: Robert M. Groves, Don A. Sillman, John L. Eltinge and Roderick J.A. Little (eds), *Survey Nonresponse*. New York, Wiley, pp. 41-54.

De Leuw, Edith, and Wim de Heer. 2002. "Trends in Household Survey Nonresponse: A Longitudinal and International Comparison." In *Survey Nonresponse*, edited by Robert M. Groves, Don A. Sillman, John L. Eltinge, and Roderick J.A. Little, 41-54. New York: Wiley.

Diekmann, Andreas, and Ben Jann. 2001. "Anreizformen und Ausschöpfungsquoten bei Postalischen Befragungen. Eine Überprüfung der Reziprozitätshypothese." *ZUMA-Nachrichten* 48: 18-27.

Groves, Robert M., and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York, Wiley.

Groves, Robert M., and Katherine A. McGonagle. 2001. "A Theory-Guided Interviewer Training Protocol Regarding Survey Participation." *Journal of Official Statistics* 17: 249-265.

Groves, Robert M.. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 72, Special Issue: 646-675.

Groves, Robert M., and Emilia Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias. A Meta-Analysis." *Public Opinion Quarterly* 72: 167-189.

Harkness, Janet, Peter Mohler, Michael Schneid, and Bernhard Christoph. 1998. "Incentives in Two German Mail Surveys 1996/97 & 1997." In: Koch, Achim and Richard Porst (eds), *Nonresponse in Survey Research*. ZUMA, Mannheim, ZUMA-Nachrichten Spezial 4.

Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M. Groves, and Stanley Presser. 2000. "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* 64: 125-148.

Lynn, Peter. 2003. "PEDAKSI; Methodology for Collecting Data about Survey Non-Respondents." *Quantity and Quality* 37: 239-261.

Matsuo, Huideko, Jaak Billiet, Geert Loosveldt, Frode Berglund, and Oyven Kleven. 2010. "Measurement and Adjustment of Non-Response Bias Based on Non-Response Surveys: The Case of Belgium and Norway in the European Social Survey Round 3." *Survey Research Methods* 4: 166-178.

Mayer, Thomas S., and Eileen O'Brien. 2001. "Interviewer Refusal Aversion Training to Increase Survey Participation." *Proceedings of the Annual Meeting of the American Statistical Association*, August 5–9, 2001.

Rässler, Susanne, und Rainer Schnell. 2003. "Multiple Imputation for Unit-Nonresponse versus Weighting Including a Comparison with a Nonresponse Follow-Up Study." Working Paper, Center for Quantitative Methods and Survey Research University of Konstanz, Germany.

Scherpenzeel, Annette, and Vera Toepoel. 2012. „Recruiting a Probability Sample for an Online Panel. Effects of Contact Mode, Incentives, and Information." *Public Opinion Quarterly* 76: 470-490.

Schnell, Rainer, and Mark Trappmann. 2006. "The Effect of the Refusal Avoidance Training Experiment on Final Disposition Codes in the German ESS-2." Working Paper, 3/2006, Center for Quantitative Methods and Survey Research University of Konstanz, Germany.

Singer, Eleanor, John Van Hoewyk, Nancy Gebler, Trivellore Raghunathan, and Katherine McGonagle. 1999. "The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys." *Journal of Official Statistics* 15: 217-230.

Singer, Eleanor, and Richard A. Kulka. 2000. "Paying Respondents for Survey Participation." Survey Methodology Program Working Paper No.092, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbour.

Singer, Eleanor. 2002. "The Use of Incentives to Reduce Nonresponse in Household Surveys." In *Survey Nonresponse*, edited by Robert M. Groves, Don A. Sillman, John L. Eltinge, and Roderick J.A. Little, 41-54. New York: Wiley.Singer, Eleanor. 2006. "Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70: 637-645

The American Association for Public Opinion Research. 2008. "Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*." 5th edition*. Lenexa, Kansas: AAPOR.

Willimack, Diane K., and Jeffrey L. Dalzell. 2006. "An Examination of Non-Contact as a Source of Nonresponse in a Business Survey." 2006 Proceedings of the Survey Research Methods Section, American Statistical Association.