### Non response to consent questions: The SAVE experience[[1]](#footnote-1)

Bettina Lamla, Munich Center for the Economics of Aging, email: lamla@mea.mpisoc.mpg.de

There is an increasing interest among social scientists in merging survey data with administrative records from social security institutions. Record linkage represents one way to combine different sources using a unique identifier such as the Social Security number. The informed consent of the respondents however is required, which in turn might induce bias to the consent question and even threaten stability in a panel study. Data from the longitudinal household survey “Saving and old-age Provision in Germany” (SAVE) are used for analysis of consent rates and patterns. Starting in 2001 the main focus of the study is to understand household saving behavior and asset choices. In the latest wave of the study participants have been asked to provide their written consent to link their answers to administrative data from the Federal Employment Agency. That is a tricky issue: willingness to give consent is not universal, thus reducing the number of linked records and potentially introducing bias (Jenkins et al., 2006; Sala et al., 2010). Furthermore, the request for consent might induce individuals to drop out of the study, thus compromising its longitudinal structure. Moreover, as SAVE is mainly a self-administered paper and pencil questionnaire while existing research is based on personal interviews. Only in those cases, in which the questionnaire was not send back after a second reminder, participants are contacted by an interviewer and the questionnaire is eventually completed in a personal interview. SAVE consists of two main subsamples which differ in their sampling scheme: the Random Sample – drawn using a random sampling technique and the Access Panel – drawn from a standing panel of households surveyed at regular intervals, operated by the company TNS Infratest TPI. Comparison between samples helps to quantify the trade-off between costs and representativeness.

A total of 1660 valid interviews could be obtained, representing 81% of the households giving an interview in the wave 2010. The response rate is clearly below the ones reached in the last waves, but it is still high in international comparison. We conclude that, neither the dramatic changes in the design (e.g. reduction of incentives by half) nor the request for consent to data linkage significantly damaged the panel. That can be partially confirmed when looking at the comments the respondents left at the end of the questionnaire. While about 47% of the respondents used the opportunity to leave a general comment about the questionnaire, an explicit complaint about the request of consent to data linkage is made in only 7 cases.

In the specific case of questions on consent to data linkage, it has been shown that asking for consent in the relevant context elicited higher consent rates than in the case the question was asked at the end of the interview (Killpack and Oskala, 2011). In SAVE 2011, the request for consent was placed right after a set of questions on old-age income and before a battery of questions on future entitlements to public pensions. That is because we motivated the request of informed consent with the need to collect data to better estimate the future pension needs. Out of the 1660 households, 956 gave an explicit consent to data linkage, representing 57.6% of the entire sample 2011. In almost 600 cases only a single consent was given, while in more than 400 cases (representing about 63% of the couples in the sample) two consents were accorded.

The SAVE data are linked with the data from the Federal Employment Agency using the SSN as a unique identifier. Thus, the consent form asked the respondents (and their partner) to report their SSN or alternatively a set of information (e.g. name, date of birth) which are necessary to reconstruct the individuals’ SSN. The quality of the information provided is good: in only 2% of the cases there are severe deficiencies in the reported data (wrong or missing SSN associated with incomplete criteria for reconstruction). In all the other cases, either the correct SSN (66% of the cases) or the complete set of criteria (32% of the cases) has been reported.

We look at possible selectivity in the pattern of consent, regressing the dichotomous variable “consent” – which equals 1 if at least on member of the household sent back a valid consent form, on a set of predictors. Besides controlling for a series of socio-demographic characteristics (age, gender of the respondent, household composition, location in East Germany, retired status, secondary and post-secondary education of the respondent, household net income quintile) another set of variables has been used to capture respondents’ differences in ability and willingness to provide consent. We expect respondents’ ability and motivation to play an important role in determining the decision to consent captured by a measure of cognitive ability (see Frederick, 2005 for more information on the ability test). In particular, good test scores are associated with less risk aversion and more patience (Oechssler et al., 2008). High cognitive abilities should help individuals in understanding the question and the scope of the consent, thus leading to higher consent rates. We expect that to be particularly true in a P&P context without an interviewer who can help the respondents in understanding the question and answer their possible concerns. We therefore interact the indicator for cognitive abilities with an indicator which equals 1 if the interview was conducted in the presence of an interviewer. Furthermore, as emphasized by Singer (2004), individuals’ attitudes toward risk and privacy are also relevant in the context of consent questions. We also use the information on health status to control for further factors which might affect the cognitive process of understanding and answering the consent question. In particular, we construct a binary variably which equals 1 if the reference person and/or his/her partner reported to suffer from long-lasting health problems in the survey 2010.We further control for the number of missing answers on sensitive items in the previous survey in 2010.[[2]](#footnote-2) High item-nonresponse may indicate both, lack of interest in the survey and concern with privacy issues. We expect therefore respondents with high item-nonresponse to those questions to be less likely to give consent. In addition, we introduce a dichotomous variable which takes value 1 if the household was receiving unemployment benefits provided by the Federal Employment Agency in the previous wave. The idea is that such individuals might be more concerned about the data linkage. Given that the IAB is the research institute of the employment agency, and the SAVE questionnaire collects (although not in the wave 2011) detailed quantitative information on households’ asset and income, those individuals might fear that their benefits (which are means tested) might be curtailed if their answers to the questionnaire are disclosed to the Federal Employment Agency. Finally, a dummy variable which equals 1 if the respondent belongs to the Access Panel captures possible differences in the consent patterns of the two subsamples.

Table 1 displays the results. There is evidence for mild consent bias. In particular, we find that older individuals, households living in East Germany and male respondents have a higher probability to give their consent. Secondary schooling is also related to the probability of agreeing to data linkage, with better educated individuals being more likely to consent. Household income is only moderately correlated with the probability to consent: only households in the middle of the income distribution appear to be significantly more likely to give their consent in comparison to low income households. Korbmacher and Schröder (2011) find a similar pattern when analysing the consent rate in SHARE. Individuals’ motivation and cognitive abilities are correlated with the probability to give consent. Interestingly, however, cognitive abilities become insignificant when the interview is conducted with an interviewer rather than as a self-administrated questionnaire. Furthermore, the interaction term with the personal interview is not statistically significant, indicating that in the presence of an interviewer other mechanisms rather than individuals’ cognitive ability determine the consent. Households suffering long-lasting health problems are also more likely to give their consent which is in line with previous research on consent patterns (e.g. Sala et al., 2010; Jenkins et al., 2006). Finally, individuals’ attitudes towards sharing sensitive information (here captured by the number of missing items to sensitive questions in the survey 2010) are significantly correlated with the probability of consenting: respondents with a high number of missing items to sensitive questions are less likely to consent to data linkage.

Altogether, the variables considered in the analysis explain only a small fraction of the total variance in the probability of giving consent (McFadden’s R2 = 0.0551)[[3]](#footnote-3).

Table 1: Propensity to consent to data linkage (probit regression)

|  |  |  |
| --- | --- | --- |
| Variable | Coefficient | Standard Error |
| Age <30 | Ref. |  |  |
| Age 30-39 | 0.278 |  | 0.212 |
| Age 40-49 | 0.108 |  | 0.201 |
| Age 50-59 | 0.463 | \*\* | 0.200 |
| Age 60+ | 0.638 | \*\*\* | 0.213 |
| RP is male | 0.190 | \*\*\* | 0.067 |
| RP has a partner | -0.050 |  | 0.086 |
| # Children at home | -0.035 |  | 0.042 |
| East Germany | 0.190 | \*\*\* | 0.078 |
| Household is retired | 0.015 |  | 0.106 |
| Unemployment benefits (t-1) | -0.083 |  | 0.138 |
| Hauptschule | Ref. |  |  |
| Realschule | 0.180 | \*\* | 0.081 |
| Abitur | 0.293 | \*\*\* | 0.110 |
| No post-secondary | Ref. |  |  |
| Vocational Training | -0.043 |  | 0.122 |
| University | -0.127 |  | 0.159 |
| Q1 | Ref. |  |  |
| Q2 | 0.098 |  | 0.121 |
| Q3 | 0.374 | \*\*\* | 0.135 |
| Q4 | 0.102 |  | 0.138 |
| Q5 | 0.189 |  | 0.147 |
| Access Panel | -0.027 |  | 0.067 |
| Personal Interview | 0.043 |  | 0.215 |
| Cognitive Ability | 0.055 | \* | 0.032 |
| Personal int. \* Cog. Ability | -0.214 |  | 0.168 |
| Long-lasting health problems | 0.120 | \* | 0.070 |
| # missings sensitive items | -0.009 | \*\*\* | 0.003 |
| Constant | -0.666 | \*\*\* | 0.217 |
|  |  |  |  |
| McFadden’s R2 | 0.0551 |  |  |
| Observations | 1,660 |  |  |

 Significance levels: \*\*\* p<0.01; \*\* p<0.05; \* p<0.10

 Source: SAVE 2010, 2011

**Points for discussion:**

* Only a small part of the variation can be explained by our analysis: Important variables missing?
* There are two stages in the decision making process: .1 Participation the survey and 2. Giving consent: Better to use Heckman selection model? What might be a good exclusion restriction?
* Couples can give 0, 1 or 2 consents. How to model this?
* SAVE is a mature panel. To what extent can results be conferred to other research?

**References**

Frederick, S. (2005): Cognitive reflection and decision making. Journal of Economic Perspectives 19, 25–42.

Jenkins, S., Cappellari, L., Lynn, P., Jäckle, A. and Sala, E. (2006): Patterns of Consent: Evidence from a General Household Survey. Journal of the Royal Statistical Society. Series A., 169(4), 701-722.

Killpack, C., Oskala, A. (2011): Understanding Society Innovation Panel Wave 4: Technical Report. National Centre for Social Research.

Korbmacher, J., Schröder, M. (2011): Consent when linking survey data with administrative data. Fourth Conference of the European Survey Research Association (ESRA), Lausanne.

Oechssler, J., Roider, A. and Schmitz, P.W. (2009): Cognitive Abilities and Behavioral Biases. Journal of Economic Behavior and Organization. 72, 147 – 152.

Sala, E., Burton, J. and Knies, G. (forthcoming): Correlates of Obtaining Informed Consent: Respondent, Interview and Interviewer Characteristics. Sociological Methods & Research.

Singer, E. (2003): Exploring the meaning of consent: participation in research and beliefs about risks and benefits. Journal of Official Statistics, 19, 273-285.

1. This paper is based on joint work with Michela Coppola. A longer version is published as a Discussion Paper and can be downloaded here: http://www.mea.mpisoc.mpg.de/uploads/user\_mea\_discussionpapers/1272\_258-12.pdf [↑](#footnote-ref-1)
2. Sensitive items are: household’s monthly income, value of real and financial assets, type and amount of outstanding debts. [↑](#footnote-ref-2)
3. Despite being slightly higher than the McFadden R2, the estimated R2 in a linear probability model is still low (R2=0.0733). [↑](#footnote-ref-3)