Interviewer Effects on Respondents' Willingness to Provide Blood Samples in SHARE

Julie M. Korbmacher

Working Paper Series 20-2014

Interviewer Effects on Respondents' Willingness to Provide Blood Samples in SHARE

Julie M. Korbmacher

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

Abstract

This paper deals with the influence of the interviewer on respondents willingness to provide blood samples in the German part of the Survey of Health, Aging and Retirement in Europe (SHARE). A multilevel approach highlights the importance of the interviewer: the empty model shows an intraclass correlation (ICC) of 36% meaning that 36% of the variance are at the interviewer level. Information coming from an additional interviewer survey is used to identify determinants which can explain these effects, focussing on interviewers' experience in the job and with the measurement as well as the expectations regarding their own success at getting respondents' consent. The results show that interviewers' experiences as well as their expectations are important determinants of consent. The ICC can be reduced substantially to 9%.

Keywords: Biomeasures, interviewer effects, interviewer survey, multilevel model

***Acknowledgement**: I would like to thank Johanna Bristle, Stephanie Eckman, Barbara Felderer, Antje Kirchner, Thorsten Kneip, Frauke Kreuter, Ulrich Krieger, Joe Sakshaug, and Luzia Weiss for valuable comments and suggestions. I received helpful comments at the 5th Conference of the European Survey Research Association, Ljubljana - (July 2013), the 68th Annual Confernece of the American Association for Public Opinion Research, Boston, USA - (May 2013), the Workshop: Explaining Interviewer Effects in Interviewer Mediated Surveys, Mannheim, Germany - (April 2013), and at the International Panel Survey Methods Workshop, Melbourne Australia - (July 2012). All remaining errors are my own. This paper uses data from SHARE wave 4 release 1.1.1, as of March 28th 2013 (DOI: 10.6103/SHARE.w4.111) or SHARE wave 1 and 2 release 2.6.0, as of November 29 2013 (DOI: 10.6103/SHARE.w1.260 and 10.6103/SHARE.w2.260) or SHARELIFE release 1, as of November 24th 2010 (DOI: 10.6103/SHARE.w3.100). The SHARE data collection has been primarily funded by the European Commission through the 5th Framework Programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life),through the 6th Framework Programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5- CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Programme (SHARE-PREP, N° 211909, SHARE-LEAP, N° 227822 and SHARE M4, N° 261982). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 and OGHA 04-064) and the German Ministry of Education and Research as well as from various national sources is gratefully acknowledged (see www.share-project.org for a full list of funding institutions). In addition, I gratefully acknowledge financial support from the Munich Center for the Economics of Aging (MEA),and the VolkswagenStiftung for funding the pilot studies of the collection of biomeasures.*

1 Introduction

Interviewers play a key role in the process of conducting a survey. Their diverse tasks begin in some studies with building the sampling frame, followed by making the first contact with the sampled unit, gaining cooperation from the sampled person(s), asking all survey questions, answering the respondent's questions, recording the answers and results, conducting potential measurements, meanwhile maintaining the respondent's motivation throughout the whole process (Blom and Korbmacher, 2013; Schaeffer et al., 2010; Groves et al., 2009). For all of these tasks, different skills are relevant and there is great potential for interviewers to (consciously and subconsciously) systematically affect the survey outcomes (Blom and Korbmacher, 2013; Durrant et al., 2010; Groves et al., 2009; Fowler, 2009), resulting in potentially biasing interviewer effects.

Due to new technologies, surveys are getting more complex, for example, by implementing new measurements and techniques. One example of such an innovation is a relatively new strand of research: the combination of medical studies and social surveys. Over the past few years, more and more studies have started the collection of biomeasures in social surveys as objective measurements of the respondents' health. A very promising new biomeasure is the collection of dried blood spots, as this new technology allows analyzing meaningful and objective blood parameters from a few drops of blood. As respondents' consent is usually necessary for the collection of dried blood spots, the list of tasks interviewers have to preform has been extended to obtaining this consent and conducting the measurements.

As for all survey requests which require the respondent's consent, the consent process itself is a very important element. First, if not all respondents consent, this decreases the number of cases and therewith the statistical power. Second, systematic differences between respondents who consent and those who do not can lead to bias (e.g. Korbmacher and Schröder, 2013; Sakshaug, 2013; Sakshaug et al., 2010). With regard to the request for consent, the interviewers are of special interest, assuming that they influence the respondents' decision whether to consent or not. In addition, interviewers are under the researchers' control (see Groves and Couper, 1998), meaning that the characteristics of the interviewer can be influenced either by the selection process or by training. Therefore, understanding the mechanisms behind interviewer effects is essential for increasing the usefulness of the data and decreasing any potential consent bias.

The goal of this paper is first to quantify interviewer effects on the respondent's decision to consent to the collection of dried blood spots in the *Survey of Health, Ageing and Retirement in Europe* (SHARE). In a second step, an additional survey, collecting detailed information on the SHARE interviewers, will be used to explain the interviewer effects by the characteristics of the interviewers. Two of the characteristics are of special interest: experience and the interviewers' expectations in the consent rates they will reach.

The following Section summarizes the idea of biomeasures in social surveys and the role of the interviewer (Section 2). The research question and a description of the data are provided in Section 4 and 5, respectively. Sections 5.1 and 6 discuss the method and the results, and the paper closes with a brief discussion of the results in Section 7.

2 The Collection of Biomeasures in Social Surveys

The integration of biology and the social sciences has become increasingly important in recent years, so that social surveys have started to collect biological characteristics and measurements (Sakshaug et al., 2014). The advantage of this development compared to classical medical surveys is that social surveys are typically based on probability samples with a large number of observations. Medical surveys are mostly based on small and non-random samples and miss collecting contextual information which is important for social scientists (Schnell, 2009). Collecting biological information in a social survey is a fruitful way to combine these two important disciplines.

The specific measures implemented in different surveys vary, as do the terms used for that class of measurements. Schnell (2009) differentiates between biometric attributes, biological attributes, and biomarkers, whereas Jaszczak et al. (2009) use the term *biomeasure* to summarize the "biological, anthropometric, functional, and sensory measurement" (Jaszczak et al., 2009, p.5) which can be collected in a survey. The following study adapts Jaszczak's terminology of *biomeasures* to summarize all physical measurements (such as grip strength and lung power test), measurements of the body (such as height, weight, or waist circumference), as well as the collection of bodily fluids. The advantage of these biomeasures compared to respondents' self-reports is first that they allow also for detecting undiagnosed diseases, for example diabetes, and second, that they can provide objective information about the health status without the measurement error due to respondents' misreporting (Jaszczak et al., 2009). Which biomeasures can be collected in a survey is not only a question for the usefulness for the research questions, but also of feasibility. One has to bear in mind that these measurements are typically collected by interviewers without medical training.

Two of the pioneer studies in implementing biomeasures within social surveys are the *Health and Retirement Study* (HRS) and the *National Social Life, Health, and Aging Project* (NSHAP).[1] HRS is a face-to-face panel survey in the U.S. conducted since 1992. It collects data from people aged 50 and older every two years (Sakshaug et al., 2010). It first piloted biomeasure collection in 2001, and has fully implemented these measurements since 2006 (e.g., physical measures, blood pressure, blood spots, and the collection of saliva) (Weir, 2008). The NSHAP started in 2005 and collects data from U.S. adults

[1] For a detailed overview of surveys collecting biomeasures, see Sakshaug (2013).

aged 57 to 85 face-to-face with the goal of studying "the links between health and sexuality in the lives of older American" (O'Muircheartaigh et al., 2009, p.12). The NSHAP collects a battery of twelve measures (including blood spots, saliva, vaginal swabs, and measures of weight, waist, and blood pressure) and, as in the HRS, all are conducted by trained interviewers without medical degrees (Jaszczak et al., 2009; Weir, 2008). In recent years, European social surveys have also implemented the collection of biomeasures in interviewer-mediated surveys. The British survey *Understanding Society* started collecting biomeasures such as height, weight, waist circumference, blood pressure, and blood in its second wave (2010), employing nurses (McFall et al., 2012). It switched to trained interviewers within a sub-sample in 2011 (McFall et al., 2014), demonstrating the pros and cons of using nurses vs. interviewers. In 2011, the *Survey of Health, Ageing and Retirement in Europe (SHARE)* tested new biomeasures in the German sub-sample (see Schaan (2013)). The following study will be based on the results of that study.

Biomeasures in the Survey of Health, Ageing and Retirement in Europe

The *Survey of Health, Ageing and Retirement in Europe* (SHARE) is a multidisciplinary and longitudinal survey collecting micro-data on the health, socio-economic status, and social and familial networks to learn more about the process of population aging (see Börsch-Supan et al., 2013). As health is a key aspect of aging, SHARE collects subjective and objective health measurements, such as self-perceived health status, grip strength, walking speed, and lung power, since the first wave of data collection (for an overview see Sakshaug et al., 2014; Schaan, 2013; Hank et al., 2009). In 2011, SHARE implemented a pilot study of collecting new biomeasures in the fourth wave within the German subsample (Schaan, 2013). This pilot study tested the feasibility of collecting additional biomeasures with non-medical interviewers at the respondents' homes. This new module consists of four measurements: 1) height, 2) waist circumference, 3) blood pressure, and 4) the collection of blood spots. For these four measurements, the respondents' written consent is required, which is collected by the interviewer on a separate paper form. Respondents have to tick a box for each measurement they agree to, so that agreeing to one of these measurements is separate from agreeing to any of the other three measurements. This paper will only focus on the collection of dried blood spots, as this is a new and very sensitive procedure in social surveys. Neither German survey agencies nor German interviewers have had any experience applying this technique.

The Role of the Interviewer and Potential Interviewer Effects

The tasks for the interviewers when collecting dried blood spots in SHARE are manifold, with a high potential for interviewer effects. They not only have to ask for the respondent's consent, but also have to conduct the measurement and administer the process.

All these tasks are assumed to influence the respondent's willingness to participate.

- **Asking for consent and answering all the respondents' questions**. Obtaining the respondent's informed consent is necessary when collecting dried blood spots (Sakshaug, 2013). This includes informing respondents about how the procedure works, the potential risks involved, which parameters will be analyzed from the blood, how consent can be withdrawn, and so on. In addition, respondents can restrict the parameters to be analyzed. As a result, the consent form is very long, consisting of four pages with very detailed information. Compared to other survey questions, this task requires much more skills on the part of the interviewer, as this request can hardly be completely scripted. This is of particular importance if respondents have questions or doubts, as the interviewers have to react spontaneously. One could assume that interviewers differ in their reactions, so that this request is quite prone to interviewer effects.

- **Conducting the measurements**. The collection of dried blood spots, that is, letting an interviewer prick a small needle into one's finger, probably requires that the respondent place more trust in the interviewer than does answering survey questions or participating in other physical measurements. The interviewer–respondent relationship during the interview thus far could play an important role in the decision of the respondent to participate.
  Interviewers might differ in how successful they are in building a trustful situation with the respondent, which could then affect the respondent's willigness to consent.

- **Administration of the process.** This task includes the handling of the materials in preparation of the blood collection. The blood is collected on special filter cards, which have to be prepared with a unique barcode sticker and sent to the laboratory in a special envelope. The number on the sticker has to be entered into the computer system so that the results can be linked to the correct respondent. This last step is also assumed to affect the respondent's decision to consent, as the way interviewers handle the materials prior to the consent question can affect the respondent's assessment of the interviewer's experience in that measurement.

This brief overview shows that the role of the interviewer in collecting dried blood spots in SHARE is much more prominent than for other questions or measurements in a survey. Interviewers who feel uncomfortable with the measurement and the whole procedure are assumed to be not as successful in 'selling' this request to respondents than are interviewers who do not have any concerns or fears. These systematic differences would then result in interviewer effects in the consent question.

3 Previous Research

The number of studies analyzing consent requests in general is increasing but most studies focus on respondents' characteristics as determinants of consent. Only a small number of studies take the effect of the interviewer into account. Four recent studies on consent to record linkage show that interviewers are important as they have an influence on the respondent's decision to consent (see: Korbmacher and Schröder, 2013; Sakshaug et al., 2013; Sala et al., 2012; Sakshaug et al., 2012). All studies analyzed the interviewers' demographics and only Korbmacher and Schröder (2013) found a significant effect of age. All four studies analyzed the effect of the experience of an interviewer, but the results are not clear. In addition, the studies differ in which aspect of experience they measure. Korbmacher and Schröder (2013) analyzed the experience within the current wave of data collection and found a statistically significant negative effect which is also found in Sakshaug et al. (2012) but not statistically significant. Sakshaug et al. (2013) included overall job experience as a dummy to compare interviewers working 37 months as an interviewer and those reporting a shorter period, and also found a statistically negative effect. Sala et al. (2012) included both job experience in years and the number of previous interviews, and found a positive effect of both but only the effect of the number of previous interviews is statistically significant. Sala et al. (2012) and Sakshaug et al. (2013) used additional data on the interviewers coming from an interviewer survey to analyze interviewer effects on consent to record linkage. Beside experience, they controlled for additional characteristics, such as attitudes and personality traits, interviewers' income, hypothetical own-consent to a different consent request, membership in social networks, and the expected consent rate. Only the interviewer's own-willingness to consent to a series of consent requests showed a significant (and positive) effect (Sakshaug et al., 2013).

Even less is known about interviewers' influence on consent to the collection of dried blood spots. Previous studies show that interviewers vary a lot in the consent rates they obtain and that these consent rates also vary between different biomeasures (Sakshaug, 2013; McFall et al., 2014; Jaszczak et al., 2009). To my knowledge there is only one study which systematically analyzes the consent to the collection of dried blood spots taking the interviewer into account.[2] Sakshaug and colleagues (Sakshaug et al., 2010) used the 2006 wave of the HRS and analyzed the differences between consenters and non-consenters to the collection of dried blood spots. In addition to the respondents' demographics and widespread measures of the respondents' health status, they included a set of variables measuring general survey resistance indicators coming from paradata

[2]The authors analyzed consent to a set of three biomeasures, including dried blood spots, together as the dependent variable. Separating the regressions by biomeasure did not change the results, so I will use the phrase "consent to dried blood spots" in the remainder of the paper.

as well as information on the interviewer. They controlled for the interviewer's age, gender, race, educational level, Hispanicity, and experience being an HRS interviewer. At the interviewer level, only the interviewer's race shows a significant effect on the respondents' consent. A significant interviewer variance term suggests that interviewer characteristics (other than demographics) have an influence on consent. This variance term shows that more information about the interviewers is needed to get the full picture. The implementation of an interviewer survey allows filling that gap by collecting that specific information on the interviewer which is assumed to have an influence.

4 Research Question

As interviewer effects on different consent questions have been proven to exist in several surveys (Korbmacher and Schröder, 2013; Sakshaug et al., 2013; Sala et al., 2012; Sakshaug et al., 2012, 2010), I hypothesize that interviewer effects also occur in SHARE when asking for consent to the collection of dried blood spots. Therefore, the first step of this analysis is to test that assumption and quantify the effect of the interviewer.
If this assumption can be confirmed, the next step will be to analyze the effect of interviewer characteristics, focusing on experience and expectations, as these are two characteristics of the interviewer which can be manipulated via selection and/or training of the interviewer.

Interviewers' Experience

The interviewers' experience seems to matter in their success at getting respondents' consent to record linkage (Korbmacher and Schröder, 2013; Sakshaug et al., 2013; Sala et al., 2012; Sakshaug et al., 2012). I hypothesize that interviewers' experience also influences the consent to the collection of dried blood spots. Three different aspects of an interviewer's experience will be distinguished: job experience, experience in collecting dried blood spots, and experience in measuring blood sugar.

- *Job experience*: This is measured as the number of years working as an interviewer. In contrast to Sala et al. (2012), I do not expect the effect of job experience to be linear. I hypothesize that an increase in experience is mainly effective at the very beginning of the career. In addition, being on the job for a very long time also implies that the interviewer's job and therewith the required tasks has changed substantially. I suspect that interviewers who started working as an interviewer a long time ago are less successful than interviewers who started more recently.

- *Experience in collecting dried blood spots in the actual wave of data collection* is not related to the first experience measurement. As SHARE is the first survey in Germany to collect dried blood spots by interviewers, they all start without any experience in asking respondents for consent to that measurement. But one could assume that interviewers learn how to persuade respondents from interview to interview. I hypothesize that interviewers are less successful in getting consent at their very first interviews.

- *Prior experience in the technique of collecting blood spots*: The procedure is almost identical to measuring blood sugar levels for people who have diabetes. I hypothesize that interviewers who are experienced in that measurement (independently from their job as interviewers) are better at getting the respondent's consent, as they are less fearful about the procedure.

Interviewers' Expectations of the Consent Rate They Will Reach

Even if the effect of an interviewer's expectations when asking for consent to record linkage was not significant in the work of Sakshaug (2013), I hypothesize that expectations are important in this specific consent request. It will be tested whether interviewers who expect to achieve a higher consent rate also reach higher consent rates. The theoretical assumption behind this is the theory of self-fulfilling prophecies, which should affect all interviewers in the same way: expectations influence the behavior of the interviewer and therewith the respondents' reactions to the request.

5 Data

SHARE: Survey Data

This paper is based on release 1.0.0 of the German Wave 4 panel sample, in which the collection of dried blood spots was implemented for the first time.[3] The target population of SHARE consists of persons aged 50 or older at the time the sample was drawn, including partners living in the same household regardless of their age (Börsch-Supan et al., 2013). A total of 1,570 respondents were asked for consent to the collection of dried blood spots during their personal interview.

Interviewer Survey

Since the information about the interviewers delivered by the survey agency is limited to a few demographical characteristics, a separate interviewer survey was conducted with the interviewers working for the fourth wave of SHARE in Germany. The questionnaire asked for information about the interviewers' experiences and expectations related to different features of the fourth wave of SHARE-Germany, including the collection of dried blood spots (see Blom and Korbmacher (2013) for further details on the interviewer

[3]For a detailed overview of SHARE's cooperation rates, see Kneip (2013).

survey). The interviewers were asked to complete the survey at the end of the training session, ensuring that their attitudes were measured independently from their first experiences in the field. Out of 197 interviewers attending the training, 165 completed that survey (a response rate of 83.8%).

Combining Both: Linking the SHARE Survey with the Interviewer Survey Data

The data of the two surveys could be linked via the interviewer ID which was requested in both surveys: in the CAPI instrument at the end of each completed interview and at the beginning of the interviewer survey. Despite the high response rate of the interviewer survey, several causes limited the number of cases for which respondent survey data could be linked to the survey data of the interviewer who conducted the interview.

1. Not all interviewers who had been trained for SHARE decided to complete their job as a SHARE interviewer: 40 of them quit before the fieldwork started.

2. Additional interviewers were hired during the fieldwork and were trained at separate training sessions. As the interviewer survey was implemented at the regular training session, these new interviewers were not asked to participate in the survey. These interviewers were mainly deployed for the refreshment sample, which is not included in the following analysis. Only three new interviewers worked for the panel sample and conducted 27 interviews.[4]

3. Unit- and item-nonresponse in the interviewer survey are responsible for an additional reduction of the sample size. The question regarding the interviewer ID in the interviewer survey suffers from item-nonresponse, so that these data could not be linked to the survey data. The survey data of 26 interviewers who conducted 555 (36%) interviews suffer from unit- or item-nonresponse.

The survey data of 988 respondents could be linked successfully with the data of the interviewer survey. This corresponds to 63% of the completed panel sample ($N = 1,570$). As the selection into the final sample is not random but depends on the interviewer, one cannot rule out that the sample is selective. However, a t-test of the interviewer sociodemographics of those interviewers who are included and those who are excluded from the final sample shows no significant differences in the characteristics which are available for all interviewers as they are provided by the agency (see Table 1). In addition, differences in the interviewer specific consent rate and the total number of interviews were tested and also show no significant differences. With respect to the seven variables which are available for all interviewers, the sample is not selective.

[4]This corresponds to 1.7% of the sample.

Table 1: Comparison of Interviewers who are Excluded with the Final Sample

	Excluded		Final sample	
Age	57.2	(1.96)	58.5	(1.24)
Men	48.3%	(0.94)	53.4%	(0.07)
Experience	5.4	(0.90)	5.2	(0.50)
Years of education	11.3	(0.39)	11.9	(0.21)
Having SHARE experience	41.4%	(0.09)	37.9%	(0.06)
Consent rate	50.5%	(4.95)	54.2%	(3.9)
Total no. of interviews	20.1	(2.96)	17.1	(1.9)
Number of interviewers	29		58	

Notes: *, **, *** mark significance on the 10, 5, 1 percent level, respectively
Standard errors in parentheses

5.1 Methods and Models

To analyze the effect of the interviewer on the respondent's consent requires a multilevel model to take the hierarchical data structure into account, as the respondents (first level) are nested within the interviewers (second level). The dependent variable in this model is the consent to the collection of dried blood spots. The consent form was handed to the respondent after the interviewer explained the procedure. This form collects consent for all four biomeasures separately. Only if the respondent signs this form is the interviewer allowed to conduct the measurements. At the end of the biomeasure module, the interviewer answered a question in the CAPI instrument indicating which measurements were completed. That final result is the dependent variable which is coded as a dummy, being 1 if the interviewer states that he or she conducted the measurement and 0 if not.

Intercept-Only Model

As a first step, an intercept-only model is calculated which does not include any explanatory variables. The outcome Y^*_{ij} for respondent i interviewed by interviewer j is explained as the regression intercept α, the residual at the interviewer level u_j, and the respondent level residual ϵ_{ij} (see Hox, 2010).

$$Y^*_{ij} = \alpha + u_j + \epsilon_{ij} \tag{1}$$

$$Y = \mathbb{1}(Y^*_{ij} \geq 0) \tag{2}$$

As the dependent variable is a dummy variable which can either be 1 or 0, Y equals

1 if the latent variable Y^*_{ij} is greater than or equal to zero. This intercept-only model provides an estimate of the intraclass correlation (ICC) ρ, which is the proportion of variance at the highest level compared to the overall variance. The ICC is calculated as the variance of the residuals at the interviewer level $\sigma^2_{u_j}$ divided by the total variance $(\sigma^2_{u_j} + \sigma^2_{\epsilon_{ij}})$. Given that the respondent level variance $\sigma^2_{\epsilon_{ij}}$ is not distributed normally but logistically, this term is fixed at $\frac{\pi^2}{3}$.

$$\rho = \frac{\sigma^2_{u_j}}{\sigma^2_{u_j} + \frac{\pi^2}{3}} \tag{3}$$

Full Model

The interviewer–respondent assignment in SHARE Germany is not random (no interpenetrated sample) but by region, which implies that all respondents interviewed by the same interviewer also live in the same region. If the respondents in one region differ systematically in some characteristics that also influence their consent, this would result in a high ICC. In such a case, the interpretation of the ICC as interviewer effects would be misleading, as these are in fact area effects. To take into account such potential area effects, the respondents' basic demographics and some health related parameters which showed significant influences on consent in other studies are controlled for. In the next step, the characteristics of the interviewer as well as of the respondent will be included in the model, where the X_{pij} are the p explanatory variables at the respondent level and the Z_{qj} are the q explanatory variables at the interviewer level. The slopes of the X_{pij} are assumed not to vary at the interviewer level (fixed slope model) in the final model. Part of the interviewer level variance of Model (1) is assumed to be explained by Z_{qj}, with u_{1j} being the remaining interviewer level residual. Simultaneously, part of the respondent level variance of Model (1) is assumed to be explained by X_{pij} with ϵ_{1ij} being the remaining respondent level residual.

$$Y^*_{ij} = \alpha + \gamma_p X_{pij} + \gamma_q Z_{qj} + u_{1j} + \epsilon_{1ij} \tag{4}$$

In both models (1) and (4) the term u_j is included, which is the residual at the interviewer level (random intercept). In multilevel models these residuals are assumed to be normally distributed (Hox, 2010; Snijders and Bosker, 2012; Rabe-Hesketh and Skrondal, 2008). For further discussion see Appendix C.1.

Explanatory Variables

Interviewers' Experience and Expectations

Three different measures of the interviewers' experience are included: The experience of working as an interviewer was measured in years[5] and is included as a continuous variable. In addition, the quadratic term of years of experience is included to test the assumption of an inversely u-shaped effect.

Experience in collecting dried blood spots in the actual wave of data collection is included to take learning effects into account. The dummy variable is 1 if the actual interview is within the first five interviews of that interviewer. As this variable is not a fixed interviewer characteristic which is stable over all respondents interviewed by the same interviewer, this variable should (from a multilevel point of view) be categorized as a respondent level characteristic. As I am interested in the learning effect of the interviewer, I will interpret this variable at the interviewer level. In addition, we asked interviewers about their experience in that measurement[6] and include that measure of experience as a dummy variable (1= familiarity, 0= otherwise).

Interviewers expectations with regard to the collection of biomeasures are asked in the interviewer questionnaire separately for each of the four new biomeasures.[7] These expectations are included as a continuous variable. The theoretical assumption behind the expected effect of the interviewer's expectations is the theory of self-fulfilling prophecies, which should affect all interviewers in the same way: the expectations influence the behavior and thereby the respondent's reaction to the request. Another explanation of a potential correlation between an interviewer's expectations and the respondents' willingness to consent could be based on the interviewer's experience. Experienced interviewers could be assumed to be more realistic in the assessment of their own abilities, meaning that they are more realistic in their expectations. If this mechanism is the one driving the effect, interviewers' expectations should be more important for experienced interviewers. To test whether the effect of interviewers' expectations differs by the interviewer's experience, an interaction of the experience and the expectations is included in the model[8] to differentiate between the two potential mechanisms.

---

[5] Q1: How long in total have you been working as an interviewer?

[6] Q23: Do you personally have experience with measuring blood sugar levels, either because you or someone you know has diabetes?

[7] Q21: What percentage of your respondents do you think will consent to [...] the collection of small blood spots?

[8] Both variables are centered around their mean.

Control Variables at the Interviewer Level

In addition to these explanatory variables, some control variables at the interviewer level are included. Interviewers *age* in years (included as a continuous variable), *gender* (1=male, 0=female), and *education* (as three dummies for low, medium, and high educational levels) are controlled for. We asked interviewers whether, if they themselves were SHARE respondents, would have consented to the collection of dried blood spots. This variable is included as a dummy. The variable *social networks* is a dummy variable taking the value of 1 if the interviewer is active in online social networks like Facebook, Myspace, or Twitter, and 0 otherwise. It is used as an indirect indicator of how open-minded an interviewer is about new technologies and the disclosure of personal information.

There are different reasons why people decide to work as interviewers. The measurement of their motivation is more complex, as the question includes a battery of different reasons for working as an interviewer, which should be rated in their importance to the interviewer separately.[9] Two reasons are included in the model: first, to "be involved in research that serves society," expecting these people to be intrinsically motivated, and second, "to have the opportunity to interact with other people," assuming that this reason does not reflect an intrinsic motivation. The two variables included in the model do not reflect the value of the rating (1–7) but they are coded as a dummy which is 1 if the corresponding aspect was rated more highly than the other aspects. Table 2 summarizes the distribution of these interviewer characteristics.

Control Variables at the Respondent Level

To control for systematic differences between respondents living in the same region, the respondents' demographics and characteristics which showed an effect in a preliminary unpublished analysis (see Weiss, 2013) of differences between SHARE respondents who consent and those who refuse are included in the model. As at the interviewer level, the respondents' gender, age, and education are included. Respondents who grew up in the former German Democratic Republic seem to be more willing to consent to record linkage (Korbmacher and Schröder, 2013; Lamla and Coppola, 2013) and to the collection of dried blood spots in SHARE (Weiss, 2013) than those who had not.[10] As respondents' health status could be influenced by their residential area and also affect their willingness to consent to the collection of dried blood spots (Weiss, 2013; Sakshaug et al., 2010), the three health measurements Weiss (2013) used in her study are included. These are whether respondents had been diagnosed with high blood cholesterol or diabetes (both

[9] There are different reasons for working as an interviewer. How important are the following aspects to you? 1=not at all, 7=very important.

[10] This effect could not be found when taking the interviewer into account.

Table 2: Sample Statistics

Interviewer characteristics	Mean/%		Min	Max
Years of job experience	7.7	(9.1)	0	41.5
Experience in measuring blood sugar	36.4			
Expected consent rate	59.6	(18.5)	4	90
Hypothetical own consent	78.2	-	-	-
Age	57.2	(9.4)	36	76
Male	54.6	-	-	-
Low educational level	7.2	-	-	-
Medium educational level	56.4	-	-	-
High educational level	36.4	-	-	-
Motivation “socialize”	27.3	-	-	-
Motivation “research”	47.3	-	-	-
Member of social networks	36.4	-	-	-
Number of interviewers	55			

Notes: Standard deviation of means in parentheses

have a significant positive effect) and the number of difficulties with everyday activities due to health problems (which influences consent negatively).[11] We know from previous research on consent to record linkage that the willingness to provide income information is a strong predictor of the probability of consenting to record linkage (Korbmacher and Schröder, 2013; Lamla and Coppola, 2013; Sala et al., 2012; Beste, 2011), and consenting to the collection of dried blood spots (Weiss, 2013). Therefore a dummy variable is included which is 1 if income was not reported and 0 otherwise. The variable “urban” is a characteristic of the area the respondent lives in and is 1 if the interviewer coded the area as a big city, suburbs or outskirts of a big city or a large town, and 0 otherwise.

6 Results

To answer the first research question, whether interviewer effects occur, we turn to an explorative approach. Figure 1 displays the interviewer specific consent rate, where each circle represents one interviewer and the size of the circle corresponds to the number of interviews that interviewer conducted.[12] This pattern shows that there is a large variation between interviewers in how successful they are in getting respondents’ consent to the collection of dried blood spots, and is a first hint of the existence of interviewer

[11]Included are activities such as dressing, preparing a meal, eating, getting in or out of bed, and so on.

[12]Included are only those interviewers who are in the final sample. The graph which refers to all interviewers looks very similar.

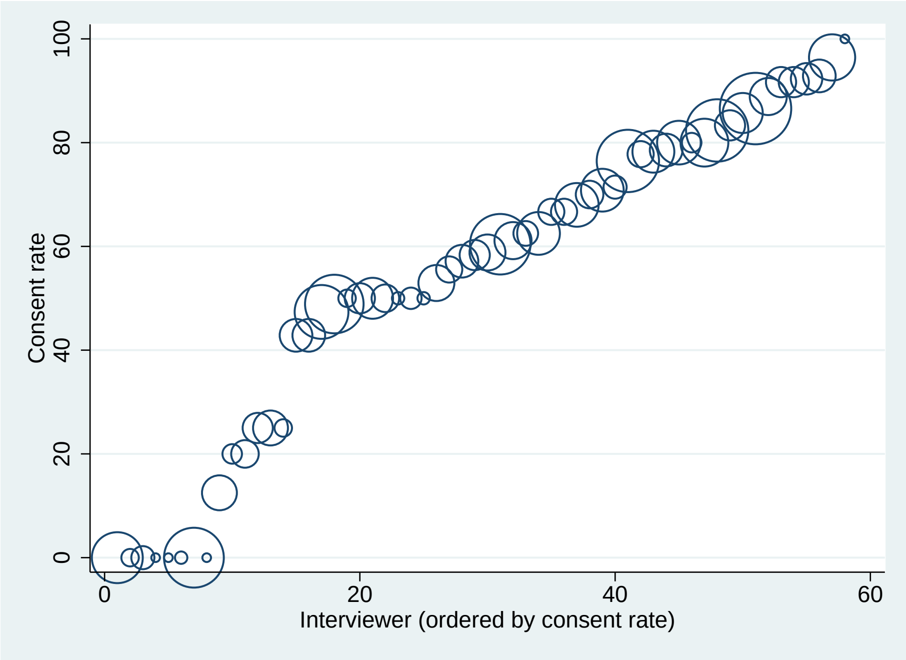

Figure 1: Interviewer Specific Consent Rate

effects. The intercept-only model is used to confirm the interpretation of Figure 1: there are sizeable interviewer effects on respondents' decisions to consent to the collection of dried blood spots. The ICC of the null model is 0.36 and is statistically significant, meaning that 36% of the total variance is at the interviewer level (see Table 3, Model 1).

The next step is to analyze the effect of the interviewer's experience and expectations. Model 2 of Table 3 shows the results of the interviewer characteristics from the full model. As the respondents' characteristics are only control variables, the results are not discussed here but displayed in Appendix A, Tabel 4. Two of the three experience measures show significant effects on the respondents' willingness to consent: the experience within the actual wave of SHARE and the years of job experience. There is a positive learning effect within the field period, as the significantly negative effect of being within the first five interviews shows. Respondents who are not one of the first being interviewed by that interviewer are more likely to consent to the collection of dried blood spots. The coefficient of the job experience (years working as an interviewer) as well as the coefficient of the quadratic term are both negative and statistically significant, indicating that the effect of job experience is not linear. As the interpretation of the effect gets rather complex when transformations of a variable and interactions with other variables are included, a graphical display of the relationship helps to understand the effect of experience on the dependent variable.

Table 3: Multilevel Estimation: Consent to the Collection of Dried Blood Spots

	Model 1	Model 2	
Age		1.03**	(0.02)
Male		0.62	(0.19)
Low educational level		0.16***	(0.10)
Medium educational level		1.30	(0.42)
Member of social networks		1.48	(0.50)
Hypothetical own consent to dried bs		1.16	(0.45)
Motivation: "socialize"		0.51*	(0.19)
Motivation: "research"		0.96	(0.34)
Experience in measuring blood sugar		0.83	(0.27)
1–5. interview		0.56***	(0.11)
Years of experience		0.96*	(0.02)
Years of experience^2		0.99***	(0.00)
Expected consent rate		1.04***	(0.01)
Years*Expectations		1.01***	(0.00)
ICC	0.36	0.09	
Number of interviewers	55	55	
Number of cases	843	843	
χ^2 against logistic regression	174.79***	5.93 ***	
χ^2 of LR test against previous model		99.08***	
(degrees of freedom; p-value of LR test)		(27; 0.000)	

Notes: *,**,*** mark significance on the 10, 5, 1 percent level, respectively
Dependent variable in all models is the dichotomous variable "consent to dbs collection"
All models are estimated in a multilevel logistic regression with Stata's xtlogit command with a random intercept on the interviewer level. Coefficients are odds ratios.
χ^2 are the respective test statistics; Model also controls for respondent characteristics.
Standard errors in parentheses.

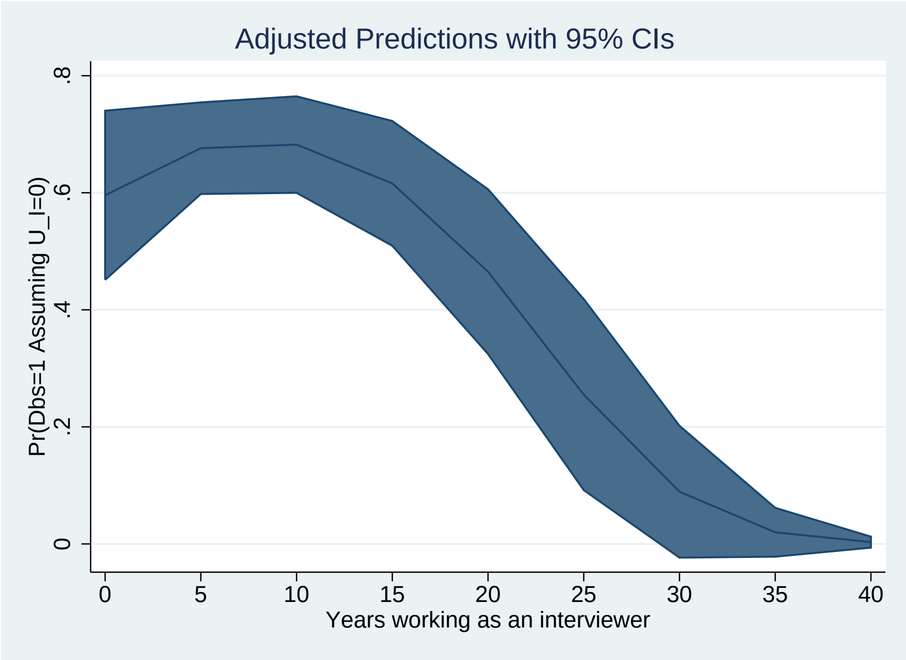

Figure 2: Predicted Probabilities of Consenting by Experience

Figure 2 shows the predicted probabilities for different levels of experience. All other variables are fixed at their mean and the random intercept is fixed at zero, meaning that the predicted probability refers to the average interviewer. The curve of the predicted probabilities shows that experience has a positive effect at the beginning of a career, reaching its peak at 7 years. After 7 years, the effect of experience is negative, meaning that the predicted probability of consenting decreases with each additional year of being an interviewer. But as the confidence interval shows, the increase in the first years is not statistically significant. The third experience measure, whether interviewers have experience in measuring blood sugar levels, shows a negative effect on consent, but this effect is statistically not significant.

Interviewers' expectations regarding the consent rate that they will reach show a positive effect on respondents' consent, as does the interaction term of expectations with experience. Similar to the results of experience, the effect of an interviewer's expectations will be discussed with the help of the predicted probabilities. As the interaction term is significant, the predicted probabilities will be presented for three different levels of experience: the lowest 10%, the average level of experience, and the highest 10%. Figure 3 shows that interpreting the positive coefficient of expectations independently from experience would be misleading. The curves of the predicted probabilities for expectations differ substantially depending on the interviewer's experience. In contrast to inexperienced interviewers, who show a negative but statistically non-significant effect of expectations on the predicted probability of consenting, the effect is positive for interviewers with about 10 years of experience. Very experienced interviewers show a different pattern, as the increase in the predicted probabilities starts later with a larger slope. The share of unexplained variance at the interviewer level after controlling for interviewer characteristics decreased to 9%: a reduction of 27 percentage points.[13]

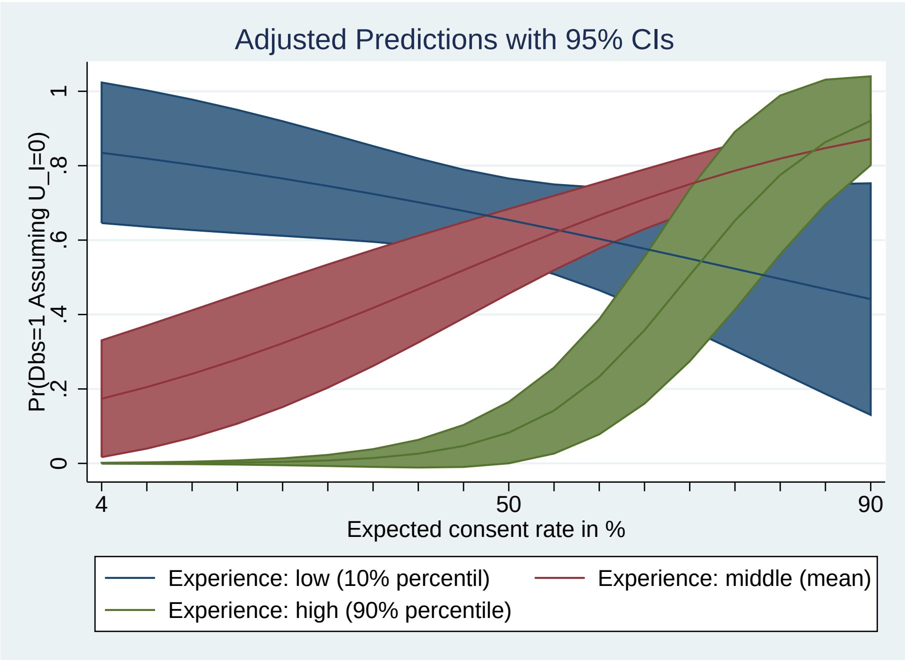

Figure 3: Predicted Probabilities of Consenting: Expectations by Experience

Regarding interviewers' demographics, both education and age have significant effects on respondents' consent. Less educated interviewers are significantly less successful than more educated interviewers in obtaining respondents' consent. Age is also positively correlated with consent. Of the two reasons for working as an interviewer, only the motivation "socialize" shows a significant effect: interviewers who report that the opportunity to interact with other people is most important are less successful in obtaining consent than are interviewers who rated other motivations higher.

A replication of the analysis after excluding two conspicuous interviewers who both interviewed a high number of respondents who all refused to consent, shows that the effect

[13]As the scale of the outcome variable changes when variables are added to the model in logistic regressions, the comparison of regression coefficients and variance components is difficult (Blom et al., 2011). I rescaled the variance components of the full model to the metric of the empty model as described in Hox (2010) to calculate the ICC, but this does not change the results. For details, see Appendix C.2.

of interviewers' experience is not stable, which could violate the external validity of the results (see Appendix B). It is questionable whether these interviewers should be included in the sample or not. Following a suggestion in Matschinger et al. (2005), I decided to not exclude them for two different reasons: First, it is hard to decide which interviewers should be excluded as there is no criterion available (Matschinger et al., 2005). So the decision is in some way arbitrary. And second, by removing interviewers based on certain characteristics, the assumption of a random set of interviewers (which should be representative of the whole sample) is systematically violated.

7 Summary and Discussion

The goal of this paper was to examine whether interviewers have an effect on respondents' decision to consent to the collection of dried blood spots and whether the interviewer effects can be explained by characteristics of the interviewer. There are three main findings: First, interviewers have a large effect on respondents' consent decision, as the ICC of 0.36 in the empty multilevel model shows. Second, information on interviewers collected in an additional interviewer survey is very useful in explaining interviewer effects. The interviewer variance in the full model was decreased substantially to an ICC of 0.09. Third, interviewers' experience in the job and with the measurement and interviewers' expectations regarding their success prior to the first interview shows significant effects, as does the interaction term of job experience and expectations.

This study shows that the effect of experience is not linear, but curvilinear, being positive at the beginning and turning negative after seven years of experience. Comparing the effect of interviewers' experience over different studies reveals ambiguous results, making it hard to come to a final conclusion about how experience affects survey outcomes. One possible explanation of this contradictory and unclear result is that different aspects overlap when controlling for the number of years a person has worked as an interviewer. Durrant et al. (2010) demonstrate that the positive effect of experience on cooperation changed after controlling for interviewers' pay grade, which reflects skill level in this study, but also increases over time. In addition, the authors suggest that self-selection processes can complicate the interpretation of experience effects. The authors assume that interviewers stay longer in a job if they are successful, whereas interviewers with a lower level of performance quit earlier. It is debatable whether these two mechanisms can be generalized over surveys and countries, as the organization of survey agencies and the arrangement of how interviewers are paid varies a lot between countries. As interviewer experience is an important aspect in the process of interviewer recruitment prior to a survey, learning more about this effect is very important. This is an issue where survey agencies and survey researchers should better cooperate.

Another aspect discussed above is related to the changing demands of the interviewer's job over the last decades. This affects particularly those interviewers who have been on this job for a very long time. Due to the implementation of computer-assisted personal interviews, the abilities needed to conduct an interview have changed substantially. So the question is whether all interviewers are able to acquire the new skills at the same speed as the demands change. Further research is needed to disentangle these different aspects and learn more about the effect of experience.

The effect of interviewers' expectations regarding their own success at getting consent seems to be positive at first glance. But the results change when taking interviewers' experience into account. For interviewers with an average level of experience and for those with much experience, a positive trend is observable. The relationship of experience and predicted success is the reverse for inexperienced interviewers. As discussed above, the effect of the interaction term could be a hint that self-fulfilling prophecies are not the underlying mechanism that explains the correlation of experience and outcome. It rather seems to be the case that experienced interviewers are better at assessing their own abilities, and therefore they perform better at forecasting their consent rates. Two limitations of this study make it difficult to differentiate between "self-fulfilling prophecies" and the ability to forecast consent rates. First, due to the problems noted in Section 5, the number of cases which are included in the calculation of the predicted probabilities is very low, not allowing more differentiated analyses. In addition, expectations are only measured at one point in time (after the interviewer training) but could be assumed to change during fieldwork as interviewers obtain their initial experiences. Therefore, the effect of expectations prior to the first interview could affect the first interviews differently. The low number of cases does not allow limiting the analysis to the first interviews. Further research is needed to analyze the effect of expectations. The collection of dried blood spots will be repeated in the sixth wave of SHARE with a new refreshment sample, which will quadruple the sample size and thereby also increase the number of interviewers.

8 Bibliography

Beste, J. (2011). Selektivitätsprozesse bei der Verknüpfung von Befragungs- mit Prozessdaten. Record Linkage mit Daten des Panels „Arbeitsmarkt und soziale Sicherung” und administrativen Daten der Bundesagentur für Arbeit. *FDZ-Methodenreport 09/2011*, 1–28. Nuremberg: Institut für Arbeitsmarkt- und Berufsforschung.

Blom, A. G., E. D. de Leeuw, and J. J. Hox (2011). Interviewer Effects on Nonresponse in the European Social Survey. *Journal of the Royal Statistical Society: Series B 47*(2), 203–210.

Blom, A. G. and J. M. Korbmacher (2013). Measuring Interviewer Characteristics Pertinent to Social Surveys: A Conceptual Framework. *Survey Methods: Insights from the Field Retrieved from http://surveyinsights.org/?p=817*, 1–16.

Börsch-Supan, A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck, and S. Zuber (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology 43*(1), 1–10.

Durrant, G. B., R. M. Groves, L. Staetsky, and F. Steele (2010). Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys. *Public Opinion Quarterly 74*(1), 1–36.

Fowler, Floyd J., J. (2009). *Survey Research Methods: 1 (Applied Social Research Methods Series)*, Chapter Survey Interviewing, pp. 127–144. SAGE Publications.

Groves, R. M. and M. P. Couper (1998). *Nonresponse in Household Interview Surveys.* New York: Wiley.

Groves, R. M., F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau (2009). *Survey Methodology.* Hoboken, New Jersey: Wiley Series in Survey Methodology.

Hank, K., H. Jürges, and B. Schaan (2009). Die Erhebung biometrischer Daten im Survey of Health, Ageing and Retirement in Europe. *Methoden Daten Analysen: Zeitschrift für Empirische Sozialforschung 3(1)*, 97–108.

Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications* (2 ed.). New York: Routledge Academic.

Jaszczak, A., K. Lundeen, and S. Smith (2009). Using Nonmedically Trained Interviewers to Collect Biomeasures in a National In-home Survey. *Field Methods 21*(1), 26–48.

Kneip, T. (2013). Survey Participation in the Fourth Wave of SHARE. In F. Malter and A. Börsch-Supan (Eds.), *SHARE Wave 4: Innovations & Methodology*, pp. 140–155. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Korbmacher, J. M. and M. Schröder (2013). Consent when Linking Survey Data with Administrative Records: The Role of the Interviewer. *Survey Research Methods 7(2)*, 115–131.

Lamla, B. and M. Coppola (2013). Please Sign Here: Asking for Consent in Self-Administered Surveys. Talk at the 5th Conference of the European Survey Research Association, July 2013 Slovenia.

Matschinger, H., S. Bernert, and M. C. Angermeyer (2005). An Analysis of Interviewer Effects on Screening Questions in a Computer Assisted Personal Mental Health Interview. *Journal of Official Statistics 21*(1), 657–674.

McFall, S., C. Booker, J. Burton, and A. Conolly (2012). Implementing the Biosocial Component of Understanding Society- Nurse Collection of Biomeasures. *Understanding Society Working Paper Series 04*.

McFall, S., A. Conolly, and J. Burton (2014). Collecting Biomarkers Using Trained Interviewers. Lessons Learned from a Pilot Study. *Survey Research Methods 08*(1), 57–66.

McKelvey, R. D. and W. Zavoinab (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology 4*(1), 103–120. doi: 10.1080/0022250X.1975.9989847.

O'Muircheartaigh, C., S. Eckman, and S. Smith (2009). Statistical Design and Estimation for the National Social Life, Health, and Aging Project. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 64B*, i12–i19.

Rabe-Hesketh, S. and A. Skrondal (2008). *Multilevel and Longitudinal Modeling Using Stata* (2 ed.). Texas: Stata Press.

Sakshaug, J. W. (2013). Using Paradata to Study Response to Within-Survey Requests. In F. Kreuter (Ed.), *Improving Surveys with Paradata. Analytic Uses of Process Information*, pp. 169–186. Hoboken, New Jersey: John Wiley & Sons.

Sakshaug, J. W., M. P. Couper, and M. B. Ofstedal (2010). Characteristics of Physical Measurement Consent in a Population-Based Survey of Older Adults. *Medical Care 48*(1), 64–71.

Sakshaug, J. W., M. P. Couper, M. B. Ofstedal, and D. R. Weir (2012). Linking Survey and Administrative Records: Mechanisms of Consent. *Sociological Methods & Research 41*(4), 535–569.

Sakshaug, J. W., M. B. Ofstedal, H. Guyer, and T. J. Beebe (2014). The Collection of Biospecimens in Health Surveys. In T. P. Johnson (Ed.), *Handbook of Health Survey Methods*. Wiley. forthcoming.

Sakshaug, J. W., V. Tutz, and F. Kreuter (2013). Placement, Wording, and Interviewers: Identifying Correlates of Consent to Link Survey and Administrative Data. *Survey Research Methods 7(2)*, 133–144.

Sala, E., J. Burton, and G. Knies (2012). Correlates of Obtaining Informed Consent to Data Linkage: Respondent, Interview, and Interviewer Characteristics. *Sociological Methods & Research 41*(3), 414–439.

Schaan, B. (2013). Collection of Biomarkers in the Survey of Health, Ageing and Retirement in Europe (SHARE). In F. Malter and A. Börsch-Supan (Eds.), *SHARE Wave 4: Innovations & Methodology*, pp. 38–46. MEA, Max Planck Institute for Social Law and Social Policy.

Schaeffer, N. C., J. Dykema, and D. W. Maynard (2010). Interviewers and Interviewing. In P. V. Mardsen and J. D. Wright (Eds.), *Handbook of Survey Research*, pp. 437–470. Binley, UK: Emerald Group Publishing.

Schnell, R. (2009). Biologische Variablen in Sozialwissenschaftlichen Surveys. *RatSWD Working Paper Series 107*, 1–4.

Snijders, T. A. and R. J. Bosker (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2 ed.). London: Sage Publishers.

Tutz, G. and M.-R. Oelker (2014). Modeling Clustered Heterogeneity: Fixed Effects, Random Effects and Mixtures. Technical Report 156, Department of Statistic University of Munich.

Weir, D. (2008). Elastic powers: The integration of biomarkers into the health and retirement study. In Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Surveys. Maxine Weinstein (Ed.), *Biosocial Surveys*, pp. 78–95. Washington, DC: The National Academies Press.

Weiss, L. M. (2013, July). True Blood? Validation Approaches for Dried Blood Spots Collection and Analyses. Talk at the weekly Seminar of the Max Planck Institute for Social Law and Social Policy, Munich.

9 Appendix

A Full Model also Including Respondent Level Variables

Table 4 displays the full model including all variables at both the respondent and the interviewer levels. The respondents' standard demographics do not show a significant effect on their willingness to consent to the collection of dried blood spots. The respondent's health status shows an ambiguous effect: on the one hand, respondents who are diabetic are more willing to consent than are those who do not have this disease. On the other hand, the more reported limitations in their daily living activities, the lower their likelihood of consenting. These results are in line with the results of Weiss (2013) and are not that surprising, considering that diabetics are used to the technique. Respondents who did not report their income are also less likely to consent to the biomeasure. In comparison to the results of Weiss (2013), who used to entire German Wave 4 sample, only one coefficient differs: the positive effect of having high cholesterol is significant in Weiss (2013) but not significant here. All other variables at the respondent level have the same sign and level of significance even if the sample used by Weiss (2013) is twice the size. The robustness of the results at the respondent level in this sub-sample supports the assumption that the sample used here is not selective.

Table 4: Multilevel Logistic Regression: Full model

	Full Model	
Respondent chrarcteristics:		
Male	1.07	(0.19)
Age: <=59	0.89	(0.26)
Age: 60–64	1.12	(0.31)
Age: 65–69	0.84	(0.25)
Age: 70–75	1.20	(0.33)
Low educational level	0.96	(0.29)
Medium educational level	1.05	(0.21)
DDR	0.98	(0.33)
High cholesterol	1.44	(0.33)
Diabetic	1.69*	(0.46)
Difficulties with activities	0.80***	(0.04)
Income missing	0.45***	(0.12)
Living in urban area	0.98	(0.23)
Interviewer chrarcteristics:		
Age	1.03**	(0.02)
Male	0.62	(0.19)
Low educational level	0.16***	(0.10)
Medium educational level	1.30	(0.42)
Member of social networks	1.48	(0.50)
Hypothetical own consent to DBS	1.16	(0.45)
Motivation: “socialize”	0.51*	(0.19)
Motivation: “research”	0.96	(0.34)
Experience in measuring blood sugar	0.83	(0.27)
1–5. interview	0.56***	(0.11)
Years of experience	0.96*	(0.02)
Years of experience^2	0.99***	(0.00)
Expected consent rate	1.04***	(0.01)
Years*Expectations	1.01***	(0.00)
ICC	0.09	
Number of interviewers	55	
Number of cases	843	
χ^2 against logistic regression	5.93 ***	
(degrees of freedom; *p*-value of LR test)	(27; 0.000)	

Notes: *,**,*** mark significance on the 10, 5, 1 percent level, respectively
Exponentiated coefficients; Standard errors in parentheses
Dependent variable in all models is the dichotomous variable “consent to dbs collection”
All models are estimated in a multilevel logistic regression with Stata’s xtlogit command with a random intercept on the interviewer level. Coefficients are odds ratios.
χ^2 are the respective test statistics.

B Excluding Interviewers

The results demonstrate that interviewers' experience seems to be an important determinant in predicting interviewers' success in getting respondents' consent. One limitation of that study is the low number of interviewers. Of those 55 interviewers, two are conspicuous because they differ a lot from all other interviewers in two aspects: they both interviewed a high number of respondents (in sum about 10% of the sample), and have a 0% consent rate (see Fig. 1, page 15). In addition, they are very experienced (with 15 and 40 years of experience). As these two interviewers could be assumed to be very influential, the analyses were repeated by excluding these two interviewers from the sample. Due to the exclusion of these interviewers, the sample is reduced by two level-two units (interviewers) and 78 level-one units (respondents). A comparision of the two models can be found in Table 5. Two main changes can be reported: first, the effect of an interviewer's experience disappears, and second, the curves of the predicted probabilities are much closer and smoother. Fig. 4 displays the predicted probabilities for the reduced

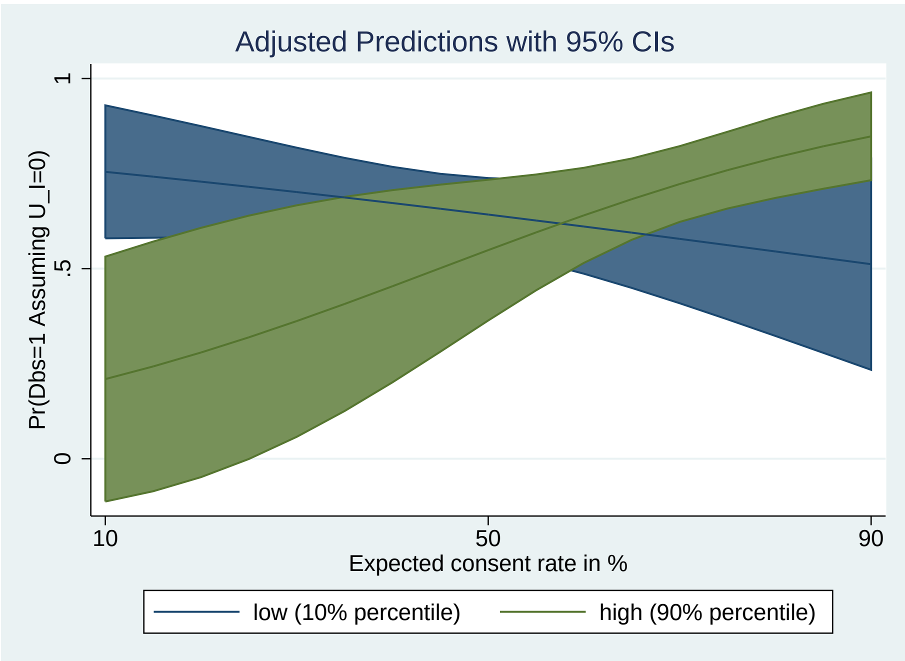

Figure 4: Predicted probabilities of consenting to the collection of dried blood spots: Reduced sample

sample of interviewers (excluding two interviewers) for less experienced and for highly experienced interviewers. [14]

[14]The predicted probabilities for interviewers with average experience are not included in the graph as these are very close to the blue curve, which would make it very hard to read the graph.

Table 5: Multilevel Logistic Regression: Comparison of Reduced and Full Sample

	Model 1 Reduced Sample		Model 2 Full Sample	
Respondent characteristics:				
Male	1.08	(0.19)	1.07	(0.19)
Age: <=59	0.94	(0.27)	0.89	(0.26)
Age: 60–64	1.21	(0.33)	1.12	(0.31)
Age: 65–69	0.94	(0.28)	0.84	(0.25)
Age: 70–75	1.23	(0.33)	1.20	(0.33)
Low educational level	0.91	(0.27)	0.96	(0.29)
Medium educational level	1.02	(0.20)	1.05	(0.21)
DDR	1.13	(0.31)	0.98	(0.33)
High cholesterol	1.42	(0.32)	1.44	(0.33)
Diabetic	1.79**	(0.49)	1.69*	(0.46)
Difficulties with activities	0.80***	(0.04)	0.80***	(0.04)
Income missing	0.43***	(0.11)	0.45***	(0.12)
Living in urban area	0.98	(0.20)	0.98	(0.23)
Interviewer characteristics:				
Age	1.04***	(0.01)	1.03**	(0.02)
Male	0.79	(0.21)	0.62	(0.19)
Low educational level	0.29**	(0.15)	0.16***	(0.10)
Medium educational level	1.72**	(0.46)	1.30	(0.42)
Member of social networks	1.50	(0.39)	1.48	(0.50)
Hypothetical own consent to DBS	0.71	(0.24)	1.16	(0.45)
Motivation: "socialize"	0.54**	(0.17)	0.51*	(0.19)
Motivation: "research"	1.04	(0.30)	0.96	(0.34)
Experience in measuring blood sugar	1.12	(0.31)	0.83	(0.27)
1–5. interview	0.53***	(0.10)	0.56***	(0.11)
Years of experience	1.02	(0.02)	0.96*	(0.02)
Years of experience^2	1.00	(0.00)	0.99***	(0.00)
Expected consent rate	1.01	(0.01)	1.04***	(0.01)
Years*Expectations	1.00**	(0.00)	1.01***	(0.00)
ICC	0.02		0.09	
Number of interviewers	53		55	
Number of cases	765		843	

Notes: *,**,*** mark significance on the 10, 5, 1 percent level, respectively
Exponentiated coefficients; Standard errors in parentheses
Dependent variable in all models is the dichotomous variable "consent to dbs collection"
All models are estimated in a multilevel logistic regression with Stata's xtlogit command with a random intercept on the interviewer level. Coefficients are odds ratios.

C Multilevel Modeling

C.1 Assumptions of the Model

Even of the random intercept model seems to be a valid way to model the heterogeneity among interviewers which allows to estimate the parameter of interest: the share of variance on the interviewer level there are some drawbacks which are the assumptions of the model. The random effect is assumed to be normally distributed (Rabe-Hesketh and Skrondal, 2008; Tutz and Oelker, 2014) an assumption which can hardly be tested. Using the 'xtmelogit' allows to predict the random intercepts but they should not be used for model diagnostic within the logistic regression (Rabe-Hesketh and Skrondal, 2008). The assumption of normally distributed random intercepts implicitly assumes that all interviewers differ in their intercept (Tutz and Oelker, 2014). The second assumption refers to the independence of the random effects and the covariates (Tutz and Oelker, 2014). An alternative model which is recommended by Tutz and Oelker (2014) is the fixed effects model as this overcomes the assumptions of the random effects model. But the disadvantage of this model is that one cannot include group-specific explanatory variables. As this is the main interest of this work, fixed effect models cannot be used.

C.2 Scale Correction

In logistic regressions, the scale of the unobserved latent variable is standardized to the same distribution in each model. By adding explanatory variables, one would expect to find smaller variance components in the full model. But in logistic regressions, the latent variable is rescaled so that the lowest level residual variance is again $\frac{\pi^2}{3}$. As a consequence, regression coefficients and variance components cannot be compared across models (see (Hox, 2010)). To correct for that fact, Hox (2010) follows the approach of McKelvey and Zavoinab (1975) by calculating a scale correction factor. The scale correction factor for the variance components is the ratio of the total variance of the null model $\sigma_0^2 = \sigma_{u0}^2 + \frac{\pi^2}{3}$ to the total variance of the model which includes only level-one characteristics, $\sigma_m^2 = \sigma_F^2 + \sigma_{u0}^2 + \frac{\pi^2}{3}$, where σ_F^2 is the variance of the linear predictor from the fixed part of the model. Before calculating the intra class correlation of the final model, all variance components have to be multiplied by this factor. The scale correction factor is 0.93, this does not change the reported ICC.