Deutsche Rentenversicherung Bund

DRV Discussion Paper Series

No. 04/2025

The "If" and "How" of Record Linkage Consent – Collection of Informed Consent to Combine Survey and Administrative Data

Ulrich Brandt
Christin Czaplicki
Dina Frommert
Thorsten Heien
Marvin Krämer

Dezember 2025

https://doi.org/10.57823/en8w-5749

The DRV Discussion Paper Series aims to encourage discussion of current topics in pension policy and research. The views and conclusions expressed in the papers are those of the authors and do not necessarily represent the position of Deutsche Rentenversicherung Bund.

Deutsche Rentenversicherung Bund, Abteilung Forschung und Entwicklung
Ruhrstraße 2, 10709 Berlin, Germany

DRV Discussion Paper No. 04/2025

The "If" and "How" of Record Linkage Consent – Collection of Informed Consent to Combine Survey and Administrative Data

Ulrich Brandt **Christin Czaplicki**[a] **Dina Frommert**
Thorsten Heien **Marvin Krämer**

Abstract
The combination of data from different sources offers many benefits. However, new sources of error can be introduced, e.g., bias inherent to the new or alternative data source, bias introduced by the technical process of the data linkage, or, if the active consent of (survey) respondents is needed, another form of nonresponse error. For the German study "Life courses and pension provisions" (LeA), survey data on life courses and pension provisions are linked with administrative data from the respondents' individual pension accounts. Because of national data protection rules, respondents' explicit consent is required to extract and link the administrative data. Therefore, the respondents need to sign a consent form during the interview. In addition to a paper consent form used in most surveys, LeA provided the respondents with an electronic form to reduce the burden of signing and sending off the form. The paper reports the consent rates and shows the extent to which the forms were used. In addition, we estimate the determinants of consent rate and consent mode by means of logit models. The results demonstrate that the amount of added selectivity is small, even though consent rate and consent mode vary between different groups of the population.

Zusammenfassung
Die Kombination von Daten aus verschiedenen Quellen bietet zahlreiche Vorteile und erweitert die analytischen Möglichkeiten. Gleichzeitig können Verzerrungen entstehen, etwa solche, die in den einzelnen Datenquellen angelegt sind oder aus dem technischen Prozess der Verknüpfung resultieren. Zudem prägen organisatorische und rechtliche Vorgaben die Umsetzung der Verknüpfung und können die Datenqualität beeinflussen. In der deutschen Studie „Lebensverläufe und Altersvorsorge" (LeA) aus dem Jahr 2016 wurden Befragungsdaten mit administrativen Informationen aus den individuellen Rentenkonten der Teilnehmenden verknüpft. Da hierfür die ausdrückliche Zustimmung der Befragten erforderlich ist, wurde während des Interviews ein Einwilligungsformular unterzeichnet. Neben dem Papierformular wurde auch ein elektronisches Formular bereitgestellt, um den Aufwand für die Teilnehmenden zu verringern. Der Beitrag beschreibt die Einwilligungsquoten und die Nutzung beider Formulare. Zudem werden Faktoren analysiert, die die Zustimmung und die Wahl der Einwilligungsform beeinflussen. Die Ergebnisse zeigen, dass zusätzliche Verzerrungen insgesamt begrenzt bleiben, auch wenn Einwilligungsquote und bevorzugte Einwilligungsform zwischen Bevölkerungsgruppen variieren.

JEL No. C18, C83

[a]Corresponding author: christin.czaplicki@drv-bund.de

1 Introduction

The individual linking of social science survey data with administrative data from the health and social services sector ("record linkage") has become very popular in the recent past. It is currently practiced in many large-scale surveys, including the "German Socio-Economic Panel" (GSOEP), the "Health and Retirement Study" (HRS), the "Panel Study of Income Dynamics" (PSID), the "Survey of Health, Ageing and Retirement in Europe" (SHARE) or "Understanding Society – The U.K. Household Longitudinal Study" (USoc) (Börsch-Supan et al. 2020, Jäckle et al. 2021, 2022, Korbmacher & Schroeder 2013, Lüthen et al. 2022, Sakshaug & Antoni 2017, Sakshaug et al. 2017). The main motivation for record linkage is that information can be obtained which could not be gained at all in a survey or at least not in the same level of quality or detail, and possibly not at the same costs, and without putting too much of a burden on the respondents (Sakshaug 2018, Sakshaug & Antoni 2017, Schnell 2015, 2016).

The key to linking data is the so-called "informed consent". While the latter can be interpreted differently depending on national legislation or the field of study, it will however include asking for permission to link the datasets – as tick in a survey question or by filling in and signing a consent form. Apart from the bias inherent to the datasets which are linked, the informed consent introduces additional selection bias, as the willingness to consent might not be equally distributed in the population of interest. To grade the success of a record linkage project we must thus look at two dimensions: How high is the consent or linkage rate? And how skewed was the willingness to consent?

Research on record linkage has provided many insights on how to increase consent rates and who the potential non-consenters are. In Germany, other issues have been the practicalities of collecting informed consent. Usually, the respondents had to fill in a paper form, sign it and post it to a record linkage agency, a process with many potential sources of error: from spelling mistakes and missing signatures to forms lost in piles or by the post. For

the project “Life courses and pension provisions” (Lebensverläufe und Altersvorsorge; LeA) we used an electronic version of the form which was integrated in the CAPI instrument to streamline the process of collecting consent and keep more people on board. On the one hand, respondents would not have to change mode during the interview, and on the other hand, the consent data could be checked for obvious errors during the interview process already.

In this paper we aim to take stock and analyze LeA record linkage consent rates and their determinants as well as share our experiences on consent mode rates and their determinants. The following section provides an overview of previous research and elaborates on determinants of successful record linkage consent. The next section introduces the data and presents details of the LeA record linkage process. The results section contains bivariate and multivariate analyses of consent rates as well as consent mode rates and the determinants of both. Finally, we discuss our results and offer some concluding remarks.

2 Research on record linkage consent

2.1 Consent to record linkage

For methodological as well as cost reasons, a high and unbiased consent is essential for the success of a record linkage project (Hartmann & Krug 2009, Sakshaug & Antoni 2017, Sakshaug & Kreuter 2012). Therefore, it is important to know the determinants of record linkage consent, which can be further differentiated as follows (Jäckle et al. 2017, 2021, 2022, Korbmacher & Schroeder 2013, Kreuter et al. 2016, Sakshaug 2018, Sakshaug & Kreuter 2012, Sakshaug et al. 2013, 2017, Thornby et al. 2018):

- survey population and characteristics of respondents,
- topic and content of survey and administrative data,
- method of data collection and, if applicable, characteristics and training of interviewers,

- request of consent: positioning, wording.

Aspects like the survey population, the topic, and the content of the data as well as the method of data collection are fixed for most surveys (especially in a longitudinal design) and don't allow for variation, but they may be important determinants of the overall level of consent. Previous research shows that especially males, younger people, single people, wealthier people, or migrants (Al Baghal et al. 2014, Korbmacher & Schroeder 2013) are less willing to consent. However, empirical results with respect to sociodemographic characteristics of consenters and non-consenters are not always consistent across studies with some research finding no relationship or even an opposite relationship of these variables with the likelihood of consent (Lüthen et al. 2022, Sakshaug et al. 2017). This may be explained by different populations or by differing study parameters accompanied with other mechanisms of consent (Sakshaug et al. 2012). Not surprisingly, respondents with privacy concerns who are unwilling to disclose financial information are less likely to consent to record linkage (Korbmacher & Schroeder 2013, Sakshaug et al. 2012). In any case, trainings – if applicable – should prepare interviewers for contacting and convincing different groups of potential consenters.

Although there's mixed evidence whether certain topics or contents of survey and administrative data correlate with lower or higher consent rates (Al Baghal et al. 2014, Jäckle et al. 2021, Sakshaug et al. 2012, Sakshaug & Kreuter 2012, Thornby et al. 2018), a high accordance of both like in LeA should be helpful since respondents may regard the administrative data as a plausible addition to the survey data. Consequently, respondents should be informed in detail about the linked administrative data not only because of existing national and EU data protection regulations, but also to maximize consent (Jäckle et al. 2022, Korbmacher & Schroeder 2013, Thornby et al. 2018).

The choice of data collection method can have a substantial effect on record linkage consent. In a review of more than 20 surveys, Fulton (2012) found higher consent rates in face-to-face surveys compared to telephone and mail surveys. Furthermore, in an employee survey, where the method of data collection was randomly assigned, Sakshaug et al. (2017) observed a much higher consent rate of 94 % in the interviewer-administered mode (face-to-face) than in the self-administered modes (mail/web) with 54 % (cf. Jäckle et al. 2021, 2022, Thornby et al. 2018). The bias of consent was larger in the self-administered mode group, too, albeit with a much smaller difference of less than six percentage points. The higher consent rate in the interviewer-administered mode comes as no surprise since "the concept of linkage is likely to raise questions (and possible concerns) among respondents. Interviewers can address these issues directly and ensure the necessary information is collected to certify the consent (e.g., signature, unique ID key)" (Sakshaug et al. 2017: 172). The positive effect of interviewers on consent can be improved by specific training measures covering all aspects of the record linkage process. With regard to characteristics of the interviewers themselves, it is striking that the older they are, the higher the approval of the respondents, while the interviewers' work experience can have a negative effect (Korbmacher & Schroeder 2013).

With respect to the way of asking for record linkage consent, it has been shown that an early request for consent can have a positive effect (Sakshaug et al. 2013). Nevertheless, in the case of a personal interview, the interviewer must use the time before the consent question to establish trust. The reference to a concrete benefit of consent ("gain frame") makes no or at least only a small difference – at least for interviewer-based surveys (Pascale 2011, Sakshaug et al. 2013, Sakshaug et al. 2019) – while emphasizing the possible negative consequences of non-consent ("loss frame") has positive effects (Kreuter et al. 2016).

2.2 Mode of consent to record linkage

Another aspect of how to ask for consent is the mode of consent, at least if the gathering of consent involves more than including a question in the survey. In addition to the paper form used in in most surveys, a new electronic form asking for the same information but integrated in the (CAPI-)questionnaire was used in the LeA study. The intention was to maximize consent by offering a form which is more consistent and comfortable to use for the respondents (Garrett 2010). However, from a behavioral economics perspective (Thaler & Sunstein 2008) it could also be argued that the "burden" to decide between two alternative forms might result in lower valid consent rates overall.

According to our knowledge, so far there is no empirical evidence with respect to consent modes, neither on whether additional consent modes (and especially a digital option) improve consent rates nor on who is likely to use which consent mode. Next to our assumption that the digital option in LeA would help to maximize record linkage consent, we hypothesize that especially older people tend to use the paper form in contrast to younger "digital natives". Another assumption is that people who are reluctant to disclose personal information are more likely to choose the paper form.

3 Data

3.1 The survey on "Life courses and old-age-provision" (LeA)

The LeA study was run by Kantar Public for the Federal German Pension Insurance (Deutsche Rentenversicherung Bund; DRV Bund) and the Federal Ministry of Labour Affairs (Czaplicki et al. 2020, Heien et al. 2021). It is concerned with old age provisions and therefore collected comprehensive and detailed information on the socio-economic situation, the life courses and the public, occupational and private old-age-provisions of respondents. The survey was based

on a multi-stage probability sample of 36,000 people aged 40 to 59 years from the population registers of 360 municipalities in Germany (Table 1).

Table 1
Results of fieldwork and record linkage (LeA target persons only)

	n	Consent rate	Consent mode
Gross sample	36,000		
- Non-respondents	28,366		
Net sample (survey)	7,634		
- Not analyzed	887		
Net sample (analyses)	6,747	100.0%	
- No valid consent	759	11.2%	
Valid consent	5,988	88.8%	100.0%
- Electronic form	5,606		93.6%
- Paper form	382		6.4%

Source: LeA, own calculations (no weights applied).

In addition to the survey, respondents were asked (using a loss frame) to give their informed consent to link the survey data to administrative data from their individual pension accounts at the German Pension Insurance Institutions. The combined information enables policy researchers to discuss and evaluate pension issues based on valid and comprehensive data (Czaplicki et al. 2020, Frommert 2020, Frommert & Rieckhoff 2020, Heien et al. 2021, Roßbach 2019, Scherger & Wagner 2020).

3.2 Survey fieldwork and record linkage process

The LeA interviews took place from May to December 2016. Since retirement decisions are often taken in a partner context, not only respondents from the initial sample (or target persons) were interviewed but also (if applicable) their partners. The interviews were conducted as personal interviews in CAPI mode, resulting in a net sample of 10,544 interviews: 7,634 interviews were interviews with target persons ("net sample (survey)" in Table 1) and 2,910 interviews were partner interviews (not shown in Table 1). In the following, we will only

consider interviews with target persons, as a target person's and their partner's probability to consent might be associated.

In the middle of the LeA interview and before the lengthy life history calendar, the respondents were informed about the aims, background, and procedure of the record linkage and were asked for their consent to individually link their survey data with the administrative data from their pension insurance accounts. All 7,634 target persons were asked for their consent to record linkage, but only 6,747 of these were identified as people who should have a pension insurance account. The following analysis is limited to these respondents ("net sample (analysis)" in Table 1).

The consent to the record linkage was gathered in two distinct steps. If the respondents signaled consent during the interview, they were asked to fill in and sign the consent form (electronic or paper). The forms asked for the Social Security Number (SSN), personal data which was used to check or replace the SSN (date, name, and place of birth) and the respondents' signature. As the record linkage was an integral part of the LeA study, the interview was shortened for the respondents who withheld consent. This is unfortunate, as it implies that many variables of interest, mainly the ones concerning the topic of the survey (life courses and old-age provisions) were only gathered for consenters and cannot be used for comparing the two groups.

Consent during the interview and then filling in the form is not the same as valid consent (Czaplicki et al 2017). In line with the different steps needed to gather the informed consent, several sources of error can be identified: non-consent during the interview, no form received by the record linkage agency, no valid signature or misspelling on the form, and later withdrawal of consent. By far the largest group of non-consenters in LeA are respondents who did not consent during the interview (695 respondents, additional information not shown in Table 1). A further 64 respondents who had agreed to the record linkage during the interview

were lost subsequently. Respondents would e.g. write “no” in the field intended for the signature, or paper forms got lost on the way. Invalid consent was found for 31 respondents who opted for the electronic form (0.5 % of all electronic forms) and for 33 respondents who used the paper form (8 % of all paper forms). In total, 759 respondents (11.2 %) did not consent to the record linkage. For our analysis we use the measure of valid consent (5,988 respondents), as it is the precondition to extract the administrative data.

The electronic form was much more popular with the respondents: 5,606 of the respondents who gave valid consent (93.6 %) opted for the electronic version, whilst only 382 of them (6.4 %) decided to use the paper form. Staying with the electronic medium of the questionnaire seems to have been the more natural choice for most respondents. However, since the survey was administered by interviewers, it might to some extent have been the interviewers’ preferences.

In the final LeA dataset, there are fewer initial respondents with linked data than the valid consent measure presented here. This is due to the possible loss of cases due to false or unidentifiable SSN, when extracting records from the administrative database, or when linking survey data and administrative records (Brandt et al. 2020, Frommert et al. 2019, 2021, Sakshaug 2019). For the LeA study these processes were optimized, resulting in very few cases lost in these steps of the record linkage process (Czaplicki et a. 2017).

4 Results

4.1 Consent to record linkage

The consent rates we find in LeA of approx. 90 % (Table 2) are higher than for other recent surveys in Germany which link survey data with data from the administrative pension records. Lately, the GSOEP was able to realize a consent rate of 54 % (Lüthen et al. 2022: 295). For SHARE, linkage rates vary from 61 % in Wave 3 in 2008 and 2009 (the first-time respondents

were asked for their consent to the record linkage) to 80 % in Wave 7 in 2017 (Börsch-Supan et al. 2020, FDZ-RV & SHARE-ERIC 2024). They however report linkage rates and not consent rates. The consent rates should be somewhat higher due to invalid information on the forms or dropout processes during the processes of extracting the administrative data and the actual linking of the data. But the rates reported also include respondents who were asked for their consent repeatedly and the rates for earlier waves are reported at the time of publication and not restricted to the initial consent.

We attribute the high consent rates in LeA to five main factors:

- Training: There was a strong focus on record linkage and the need for informed consent in the (personal) interviewer training. The importance of this part of the study was reinforced by the presence of representatives of the Federal German Pension Insurance at each of the one-day training events.
- Content: The focus of the study was closely related to the administrative data the respondents were asked to give access to. The LeA study was not a multi-topic survey like GSOEP or SHARE but was specifically concerned with life courses and old age provisions. This should help respondents to understand why their administrative records from the pension insurance were needed for the study and might encourage them to consent.
- Trust: One of the initiators of the study was the German Pension Insurance itself. So, the institution asking for the consent was a main player in the field of interest to the study and the same one that held the administrative data. This might have caused respondents to consent more readily, since they might have perceived the request as particularly legitimate.
- Parsimony: The cross-sectional LeA survey was a single occasion and not a panel survey like GSOEP or SHARE. In a panel survey, respondents might be more likely to respond

negatively to a record linkage request, since they are interviewed repeatedly and give so much information already (Jäckle et al. 2021).

- Design: The framing of the question and the position in the questionnaire were given a lot of attention in LeA. In the pretest, different positions and wordings were tested. While the results on the position in the questionnaire were unambiguous, the framing of the question had less effect. Thus, the framing was oriented at the evidence from previous research, which found that a negative framing was more successful (Kreuter et al. 2016).

Additionally, the electronic form might have reduced the burden for the respondents, since they did not have to change mode in the interview situation and/or later remember to post the paper form. We'll come back to this in the final discussion.

Next to the total consent rate of 88.9 %, the rates for different sociodemographic groups are shown in Table 2. Note that the total value differs slightly from the one presented in Table 1 since it is based on weighted data (design weights to compensate for over proportional sampling of migrants and East Germans). The same applies to the consent mode rate presented in Table 4. The results presented in Table 2 are not exhaustive but represent the most interesting effects. Other variables such as marital status or the contribution history will be controlled for in the logit model but did not show much variance on the bivariate level.

Table 2
Consent rate by socio-demographic indicators (in %)

		Yes	**No**
Gender*Region	Men West	88.6	11.4
	Women West	88.5	11.5
	Men East	89.5	10.5
	Women East	91.2	8.8
Migration history	Germans born in Germany	89.6	10.4
	Germans born abroad	85.3	14.7
	People from other nations	79.3	20.7
Age (in 2016)	40 - 44	90.0	10.0
	45 - 49	89.1	10.9
	50 - 54	88.2	11.8
	55 - 59	88.7	11.3
Education level	Low	86.7	13.3
	Medium	90.4	9.6
	High	89.0	11.0
	No formal certificate	85.6	14.4
Number of children	No children	86.3	13.7
	1 child	89.1	10.9
	2 children	88.7	11.3
	3 and more children	92.0	8.0
Personal net income*	Q 1	89.4	10.6
(quartiles)	Q 2	89.7	10.3
	Q 3	91.4	8.6
	Q 4	89.6	10.4
Readiness to disclose	Information provided	90.0	10.0
personal net income	Refusal	68.3	31.7
	Don't know	91.4	(8.6)
Total		**88.9**	**11.1**

Source: LeA, own calculations (design weights applied).
Note: n = 6,747; * n = 6,226; / n < 10; ( ) 9 < n < 30.

There is a slight difference between men and women in West and East Germany. The rates are higher in East Germany and women in East Germany show the highest consent rate of 91.2 %. We combine gender and region in our analysis since the life courses and old age provisions of men and women in East and West Germany are still substantially different and we expected the consent measure to reflect these differences to some extent. For migration history we find more pronounced differences. Respondents who were not born in Germany are

less willing to consent. This is evident for Germans who were born abroad, and is even more pronounced for respondents with other nationalities, of whom only 79.3 % consented to the record linkage. If we look at the education level, which refers to schooling in this paper, we find that respondents with a medium level of education show the highest consent rate of 90.4 %. Respondents with a high level of education rank second, and respondents with a low level of education rank third with a consent rate of 86.7 %. Only 85.6 % of respondents with no formal certificate agreed to have their data linked. In contrast, there are only very small differences with respect to age with consent rates ranging from 88.2 % (50 to 54 years) to 90.0 % (40 to 44 years) although greater differences may result when surveying younger and/or older cohorts than the LeA population. There are differences relating to the number of children: Respondents without children were least likely to consent (86.3 %). Respondents with one (89.1 %) or two children (88.7 %) were more likely to consent and the highest consent rate of 92.0 % was found for respondents with three or more children. A differentiation according to personal net income at the time of the survey does not show substantial differences. Across the four income groups the consent rate is roughly 90 %, only respondents in the third quartile (91.4 %) consented slightly more often to the record linkage than the other ones. More important for the decision to consent is the willingness to disclose income information. Both items represent personal information related to the financial situation, which is a sensitive topic for respondents (Juster & Smith 1997). If income information was provided, 90.0 % of the respondents also agreed to the record linkage. Respondents who refused to disclose their income also opposed the record linkage. Only 68.3 % of these respondents consented, which is by far the lowest rate among the different socio-demographic groups.

To confirm the bivariate results, we ran a logit model predicting the binary outcome consent (1) or refusal (0). The resulting odds ratios (OR) and their 95% confidence interval (CI) are presented in Table 3.

Table 3
Logit model for consent rate (consent=1)

		OR	95 % CI	
Gender*Region	Men West	0.906 ns	0.643	1.278
(ref: Women East)	Women West	0.899 ns	0.646	1.252
	Men East	0.879 ns	0.582	1.327
Migration history	Germans born in Germany	2.407 ***	1.720	3.366
(ref: Other nationality)	Germans born abroad	1.510 +	0.996	2.291
Age (in years)		0.989 ns	0.972	1.006
Education level	Medium	1.365 **	1.100	1.694
(ref: Low)	High	1.341 *	1.066	1.687
	No formal certificate	1.104 ns	0.642	1.900
Marital status	Single	0.875 ns	0.684	1.120
(ref: Married)	Divorced	0.967 ns	0.761	1.229
	Widowed	0.810 ns	0.446	1.472
Number of children		1.145 **	1.056	1.242
Contribution period (in years)		1.020 ns	0.971	1.072
Personal net income (in €)		1.000 ns	1.000	1.000
Constant		5.167 **	2.020	13.214
Nagelkerke R^2		0.020		

Source: LeA, own calculations (design weights applied).
Note: n (unweighted) = 6,226; *** p < 0.001; ** p < 0.01; *p < 0.05; + p < 0.1.

As could be expected from the bivariate analysis, there are very few significant differences. The only significant factors are the migration history, the education level, and the number of children. Most notably, the odds to consent to the record linkage are about 140 % higher for Germans who were born in Germany (OR = 2.407) than for respondents with other nationalities. Generally, migrants might be less likely to have been in contact with the German Pension Insurance than Germans, but here, we only consider respondents with pension accounts. We therefore assume this is a substantial finding. Neither age (OR = 0.989) nor the length of the contribution period (OR = 1.020) show a significant effect. But we would still like to point out that they are not as closely associated as one might think and both have their place in the model. The length of the contribution period can be viewed as a proxy for labor market integration and might be longer with age, but life courses often contain episodes which

are not covered by the pension insurance and the two variables are thus not perfectly correlated (the correlation coefficient r varies around 0.4, depending on the selected data and the method of calculation).

4.2 Consent mode

This section looks at respondents who consented to the record linkage. The measure presented is valid consent in contrast to stating consent in the interview only. The difference might be small, it is however an important point when comparing the results internationally, where the informed consent might be collected in less formal ways. Table 4 shows the proportion of respondents who opted for the electronic and the paper form. It is striking, how many respondents preferred the electronic form (94.4 %; paper: 5.6 %), especially when considering that the respondents were between 40 and 59 years old.

As expected, we find that older respondents have a slight tendency to prefer the paper form (55 to 59 years: 6.8 %), however, the differences between the age groups are very small. In contrast to the consent measure discussed above, we do find a more pronounced difference between East and West Germany: Slightly more than ten percent of the East German respondents preferred the paper form (men: 10.5 %; women: 10.6 %). Compared to West Germans (men: 4.2 %; women: 4.6 %), this share is more than twice as large. For the other groups presented in Table 4 there are no differences, the share of respondents choosing paper forms is always between four and six percent. So East Germans do stick out as the ones with most reservations against digital methods of data collection. We assumed that people who are reluctant to disclose personal information would rather choose the paper form instead of the digital option. Interestingly, and in contrast to the consent measure, we find no effect on the consent mode (information provided: 5.6 %; refusal: 6.0 %).

Table 4
Consent mode by socio-demographic indicators (in %)

		Electronic	**Paper**
Gender*Region	Men West	95.8	4.2
	Women West	95.4	4.6
	Men East	89.5	10.5
	Women East	89.4	10.6
Migration history	Germans born in Germany	94.3	5.7
	Germans born abroad	94.9	(5.1)
	People from other nations	96.5	(3.5)
Age (in 2016)	40 - 44	95.5	4.5
	45 - 49	94.7	5.3
	50 - 54	95.1	4.9
	55 - 59	93.2	6.8
Education level	Low	94.7	5.3
	Medium	93.9	6.1
	High	94.8	5.2
	No formal certificate	95.8	/
Number of children	No children	94.6	5.4
	1 child	94.7	5.3
	2 children	94.1	5.9
	3 and more children	94.6	5.4
Personal net income*	Q 1	94.7	5.3
(quartiles)	Q 2	93.9	6.1
	Q 3	93.6	6.4
	Q 4	95.5	4.5
Readiness to disclose	Information provided	94.4	5.6
personal net income	Refusal	94.0	(6.0)
	Don't know	96.4	/
Total		**94.4**	**5.6**

Source: LeA, own calculations (design weights applied).
Note: n = 5,988, * n = 5,596; / n < 10; ( ) 9 < n < 30.

As in the section above concerning the consent to record linkage, we ran a logit model on the preferences for electronic (1) over paper (0) forms. The resulting odds ratios (OR) and their 95% confidence interval (CI) are presented in Table 5.

Table 5
Logit model for consent mode (electronic=1)

		OR	95 % CI	
Gender*Region	Men West	2.726 ***	1.868	3.980
(ref: Women East)	Women West	2.323 ***	1.639	3.291
	Men East	1.044 ns	0.699	1.560
Migration history	Germans born in Germany	0.718 ns	0.333	1.550
(ref: Other nationality)	Germans born abroad	0.739 ns	0.302	1.808
Age (in years)		0.982 ns	0.959	1.006
Education level	Medium	1.160 ns	0.852	1.581
(ref: Low)	High	1.105 ns	0.794	1.540
	No formal certificate	1.362 ns	0.521	3.562
Marital status	Single	0.768 ns	0.547	1.078
(ref: Married)	Divorced	0.864 ns	0.628	1.189
	Widowed	1.940 ns	0.606	6.210
Number of children		0.981 ns	0.882	1.091
Contribution period (in years)		1.000 ns	0.933	1.072
Personal net income (in €)		1.000 ns	1.000	1.000
Constant		32.030 ***	7.720	132.892
Nagelkerke R^2		0.028		

Source: LeA, own calculations (design weights applied).
Note: n (unweighted) = 5,596; *** p < 0.001; ** p < 0.01; *p < 0.05; + p < 0.1.

The results in Table 5 are very much in line with the bivariate results: The only significant difference is the one between East and West German respondents. The odds to choose the electronic form are about 130 % higher for West German women (OR = 2.323) than for East German women and even about 170 % higher for West German men (OR = 2.726). In contrast to the slight differences concerning bivariate relationships shown above, there is no significant age effect. Older respondents are as likely to opt for the electronic form as younger ones (OR = 0.982). However, we only considered ages 40 to 59, and it might be different if older age groups are included in the survey.

Our results concerning the consent to record linkage indicate that digital methods are accepted by a large proportion of the population and might be used to reduce the burden of

consenting. Still, it is necessary to offer a paper alternative. At least for East Germany a substantial subgroup would otherwise not be represented well in the sample.

5 Discussion and conclusion

Our findings based on the LeA study show a high overall level of record linkage consent of 89 %. This is attributed to five different aspects of the study in general and the record linkage process in particular: training, content, trust, parsimony, and design. Furthermore, our results highlight that the precondition of informed written consent for record linkage does not necessarily introduce much additional sample selectivity. Also, the sub-groups who were less likely to consent are well known from previous research into sample selectivity and non-response error: Namely migrants and people who are generally wary of disclosing personal or income information.

With respect to the mode of consent to record linkage, almost all respondents (94 %) opted for the electronic form. Only East Germans chose the electronic form at a significantly lower rate, but still a large majority of 90 % made use of it. Contrary to our expectations, we found no evidence that older people and people reluctant to disclose personal information are more likely to use the paper form. Our results thus provide strong evidence for the fact that digital methods are widely accepted.

Whether digital methods can increase consent rates significantly by streamlining the process of consenting remains an open question. While we do believe that our results point towards a reduction of the burden on the respondents through digital methods and a higher rate of valid consent, we did not systematically control for all factors of influence. For example, the choice of medium might have been driven by the interviewers' rather than the respondents' preferences. An experimental assignment to both modes could shed more light on the effects of offering alternative ways of written consent. Digital methods have additional advantages: Less respondents are lost who consented in the interview but then failed to provide the signed

form, either because they changed their mind, forgot to post the form or it was lost on the way. The electronic data can be checked during the interview process and the risks of spelling errors or missing data are considerably reduced. Subsequently, less cases are lost in the following steps and a larger proportion of the sample retained. Some sub-groups might be uneasy with the electronic form and are better reached with a paper alternative. It might be less convenient for the interview process to offer several consent modes, but a mixed mode approach can result in higher consent rates with less selectivity.

The main limitations of our analysis are the narrow scope of possible predictors and the absence of a controlled experiment to test the effect of the alternative consent modes. The record linkage was an important part of the LeA study, so if respondents had a pension account and did not consent to the record linkage, the interview was shortened out of considerations for the overall costs of the study. This means that many predictors which are more closely related to the theme of the study were collected for consenters only and cannot be used to compare consenters with non-consenters. Also, the study was conducted in a non-academical setting where the inclusion of methodological experiments is still not common practice. Future record linkage projects could target these methodological issues by collecting a broader set of comparable indicators or randomly assigning the consent mode at least to a part of the sample.

Linking different data sources is gaining importance in several scientific fields. However, new data protection regulation in the EU demands that explicit consent is given by the individual respondents for any record linkage. The legal interpretation of explicit consent may vary in different countries, but still the need for methods which reliably provide high, unbiased consent rates is apparent. To facilitate the comparison of methods and consent rates across studies and countries, a standard should be established on how consent and linkage rates should be reported.

References

Al Baghal, T., Knies, G., & Burton, J. (2014). Linking administrative records to surveys: Differences in the correlates to consent decisions. *Understanding Society Working Paper Series*, 2014-09, Essex: Institute for Social and Economic Research.

Börsch-Supan, A., Czaplicki, C., Friedel, S., Herold, I., Korbmacher, J., & Mika, T. (2020). SHARE-RV: Linked Data to Study Aging in Germany. *Journal of Economics and Statistics, 240* (1), pp. 121–132.

Brandt, U., Frommert, D., Czaplicki, C., Heien, T., & Krämer, M. (2020): *How to combine survey data with administrative data successfully? Lessons from the German study on "Life courses and pension provisions"*. [Poster presentation] 2nd Conference "Big Data Meets Survey Science" (BigSurv20), Online.

Czaplicki, C., Frommert, D., Langelüddeke, A., & Zanker, D. (2017). Save as many as possible: How to reduce selectivity in the record linkage process. [Paper presentation] 7th Conference of the European Survey Research Association (ESRA), Lisbon.

Czaplicki, C., Frommert, D., & Zanker, D. (2019). Die Studie Lebensverläufe und Altersvorsorge (LeA) im Überblick. *Deutsche Rentenversicherung, 74* (1), pp. 25-50.

Czaplicki, C., Frommert, D., & Zanker, D. (2020). Individuelle Ausgestaltung von selbstständigen und abhängigen Erwerbstätigkeiten im Lebensverlauf. Frerichs, F. & Fachinger, U. (Ed.), *Selbstständige Erwerbstätigkeit und Erwerbskarrieren in späteren Lebensphasen. Potenziale, Risiken und Wechselverhältnisse*, Berlin: Springer VS, pp. 19-44.

FDZ-RV, SHARE-ERIC (2024). *SHARE-RV. Release version: 9.0.0.* SHARE-ERIC. Data set. DOI: 10.6103/SHARE.SHARE-RV.900

Frommert, D. (2020). Soziale Elemente im Mehr-Säulen-System und ihre gleichstellungspolitische Relevanz. Bundesministerium für Arbeit und Soziales (Ed.), *Bericht der Kommission Verlässlicher Generationenvertrag – Band II – Materialien.* Berlin: BMAS, pp. 134-144 (https://www.bmas.de/DE/Soziales/Rente-und-Altersvorsorge/Kommission-Verlaesslicher-Generationenvertrag/Bericht-der-Kommission/bericht-der-kommission.html; 13.11.2024).

Frommert, D., Czaplicki, C., Heien, T., & Krämer, M. (2019). *Selectivity in the record linkage process – A big issue?* [Paper presentation] 8th Conference of the European Survey Research Association (ESRA), Zagreb.

Frommert, D. & Rieckhoff, C. (2020). Riester-Rente: Beteiligung und Anwartschaften in der zweiten Hälfte des Erwerbslebens. *RVaktuell, 67* (1), pp. 13-24.

Frommert, D., Czaplicki, C., Heien, T., & Krämer, M. (2021): Digital Vs. Analogue Ways of Gaining Record Linkage Consent. [Paper presentation] International Sociological Association (ISA), 4th ISA Forum of Sociology "Challenges of the 21st Century: Democracy, Environment, Inequalities, Intersectionality", Online.

Fulton, J. A. (2012). *Respondent consent to use administrative data.* University of Maryland, PhD dissertation: College Park, MD.

Garrett, J. J. (2010). *The Elements of User Experience: User-Centered Design for the Web and Beyond* (2nd edition), Berkeley: New Riders.

Hartmann, J. & Krug, G. (2009). Verknüpfung von personenbezogenen Prozess- und Befragungsdaten – Selektivität durch fehlende Zustimmung der Befragten? *Zeitschrift für ArbeitsmarktForschung, 42* (2), pp. 121-139.

Heien, T. & Frommert, D. (2016). *Migration and pensions: How to collect valid data on pension entitlements of immigrants* [Paper presentation] 2nd International Conference on Survey Methods in Multinational, Multiregional, and Multicultural Contexts (3MC), Chicago.

Heien, T. & Krämer, M. (2018a). *Lebensverläufe und Altersvorsorge der Personen der Geburtsjahrgänge 1957 bis 1976 und ihrer Partner.* DRV-Schriften 115 and BMAS-Forschungsbericht F519, Berlin (www.lea-studie.de; 13.11.2024).

Heien, T. & Krämer, M. (2018b). *Lebensverläufe und Altersvorsorge (LeA) – Methodenbericht.* München: Kantar – Public Division (www.lea-studie.de; 13.11.2024).

Heien, T., Krämer, M., & Frommert, D. (2021), The gender pension gap: addressing income inequality in retirement. PUBLIC, Nr. 2 (https://kantar.turtl.co/story/public-journal-02/; 13.11.2024).

Jäckle, A., Gaia, A., & Benzeval, M. (2017). Mixing modes and measurement methods in longitudinal studies. London: UCL Institute of Education.

Jäckle, A., Beninger, K., Burton, J., & Couper, M.P. (2021), Understanding data linkage consent in longitudinal surveys. Lynn, P. (Ed.) Advances in Longitudinal Survey Methodology. Hoboken, NJ: Wiley, pp.122–150.

Jäckle, A., Burton, J., Couper, M. P., Crossley, T. F., & Walzenbach, S. (2022), How and Why Does the Mode of Data Collection Affect Consent to Data Linkage? *Survey Research Methods, 16* (3), 387–408.

Juster, T. F. & Smith, J. P. (1997). Improving the Quality of Economic Data: Lessons from the HRS and AHEAD. *Journal of the American Statistical Association, 92* (440), pp. 1268-1278.

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

Kreuter, F., Sakshaug, J. W., & Tourangeau, R. (2016). The Framing of the Record Linkage Consent. *International Journal of Public Opinion Research, 28* (1), pp. 142-152.

Lüthen, H., Schröder, C., Grabka, M., Goebel, J., Mika, T., Brüggmann, D., Ellert, S. & Penz, H. (2022). SOEP-RV: Linking German Socio-Economic Panel Data to Pension Records. *Journal of Economics and Statistics, 242* (2), pp. 291-307.

Pascale, J. (2011). *Requesting Consent to Link Survey Data to Administrative Records: Results from a Split-Ballot Experiment in the Survey of Health Insurance and Program Participation (SHIPP).* Washington: U.S. Census Bureau.

Roßbach, G. (2019). Die Relevanz von guten Daten – Zur Studie „Lebensverläufe und Altersvorsorge“ (LeA). *RVaktuell, 66* (2), pp. 38-42.

Sakshaug, J. W. (2018). Methods of Linking Survey Data to Official Records. Vannette, D. L. & Krosnick, J. A. (Ed.), *The Palgrave Handbook of Survey Research*, Cham: Palgrave, pp. 257-266.

Sakshaug, J. W. (2019). *Advantages and Challenges of Linking Surveys with Big Data* [Paper presentation] Conference „Qualität bei zusammengeführten Daten – Befragungsdaten, Administrative Daten, Neue digitale Daten: Miteinander besser?“, Wiesbaden: Federal Statistical Office.

Sakshaug, J. W. & Antoni, M. (2017). Errors in linking survey and administrative data. Biemer, P. P., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L. E., Tucker, N. C., & West, B. T. (Ed.), *Total survey error in practice*, Hoboken: Wiley, pp. 557-573.

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

Sakshaug, J. W., Hülle, S., Schmucker, A., & Liebig, S. (2017). Exploring the Effects of Interviewer- and Self-Administered Survey Modes on Record Linkage Consent Rates and Bias. *Survey Research Methods, 11* (2), pp. 171-188.

Sakshaug, J. W. & Kreuter, F. (2012). Assessing the Magnitude of Non-Consent Biases in Linked Survey and Administrative Data. *Survey Research Methods, 6* (2), pp. 113-122.

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

Sakshaug, J.W., Stegmaier, J., Trappmann, M., & Kreuter, F. (2019). Does benefit framing improve record linkage consent rates? A survey experiment. *Survey Research Methods, 13* (3), pp. 289-304.

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

Scherger, S. & Wagner, G. G. (2020). Soziale Ungleichheiten in Lebensläufen, Alterseinkommen und Lebenserwartung. Bundesministerium für Arbeit und Soziales (Ed.), *Bericht der Kommission Verlässlicher Generationenvertrag – Band II – Materialien.* Berlin: BMAS, pp. 19-96 (https://www.bmas.de/DE/Soziales/Rente-und-Altersvorsorge/Kommission-Verlaesslicher-Generationenvertrag/Bericht-der-Kommission/bericht-der-kommission.html; 13.11.2024).

Schnell, R. (2015). Linking Surveys and Administrative Data. Engel, U., Lynn, P., Jann, B., Scherpenzeel, A., & Sturgis, P. (Ed.), *Improving Survey Methods – Lessons from Recent Research*, New York: Routledge, pp. 273-287.

Schnell, R. (2016). Record Linkage. Wolf, C., Joye, D., Smith, T. W., & Fu, Y. (Ed.), *The SAGE Handbook of Survey Methodology*, London: SAGE, pp. 662-669.

Thaler, R. D. & Sunstein, C.R. (2008). *Nudge – Improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.

Thornby, M., Calderwood L., Kotecha M., Beninger K. & Gaia A. (2018). Collecting Multiple Data Linkage Consents in a Mixed-mode Survey: Evidence from a large-scale longitudinal study in the UK Survey Methods: Insights from the Field. Retrieved from https://surveyinsights.org/?p=9734.

Zanker, D. & Czaplicki, C. (2019). Die Erhebung von Erwerbsverläufen und Rentenanwartschaften im In- und Ausland – Ein Vergleich von Personen mit und ohne Migrationserfahrung. *Sozialer Fortschritt, 68* (11), pp. 867-886.