Respondent Understanding of Data Linkage Consent

Survey Methods: Insights from the Field

Joseph W. Sakshaug | Alexandra Schmucker | Frauke Kreuter | Mick P. Couper | Leonie Holtmann

How to cite this article : Sakshaug, J.W., Schmucker, A., Kreuter F., Couper M.P. & Holtmann, L. (2021). Respondent Understanding of Data Linkage Consent Survey Methods: Insights from the Field. Retrieved from https://surveyinsights.org/?p=15407

DOI : 10.13094/SMIF-2021-00008

Copyright : © the authors 2021. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Abstract : Across survey organizations around the world, there is increasing pressure to augment survey data with administrative data. In many settings, obtaining informed consent from respondents is required before administrative data can be linked. A key question is whether respondents understand the linkage consent request and if consent is correlated with respondent understanding. In the present study, we investigate these issues in separate telephone and Web surveys, where respondents were presented with follow-up knowledge questions to assess their understanding of the linkage consent request. Overall, we find that understanding of the linkage request is relatively high among respondents who consent to linkage and rather poor among those who do not consent, with some variation in the understanding of specific aspects of the linkage request, including data protection. Additional correlates of understanding were also identified, including demographic characteristics, privacy attitudes, and the framing and placement of the linkage consent questions. Practical implications of these results are provided along with suggestions for future research.

Introduction

Many surveys in the social sciences conduct data linkages by merging their interview data to external data sources, including administrative records. Linking surveys with administrative data has several advantages, including enhancing research opportunities for social scientists and policy-makers, improving the cost-efficiency of data collection, and minimizing respondent burden. These advantages have prompted several high-ranking committees and organizations to endorse the use of data linkage for official statistics and evidence-based policymaking (US National Academies of Sciences, Engineering, and Medicine 2017; US Commission on Evidence-Based Policymaking 2017; Longitudinal Studies Strategic Review 2018). However, linking survey data to administrative data is not without challenges. One of the biggest challenges is obtaining consent from respondents. Particularly in Europe, obtaining explicit (i.e. opt-in) linkage consent is a requirement under the EU General Data Protection Regulation (GDPR; European Parliament and Council of European Union 2016), with some exceptions for surveys conducted by National Statistical Institutes. As such, many large-scale surveys, including the UK Household Longitudinal Study and the German Socio-Economic Panel carry out extensive procedures to inform respondents about the linkage prior to requesting their consent (Buck and McFall 2012; Goebel et al. 2019). For various reasons, however, not all respondents agree to the linkage request. This has potential consequences for the utility of the linked data as non-consent reduces the analytic sample size and may

introduce bias in linked-data analyses if consenting respondents are systematically different from respondents who do not consent (e.g. Sakshaug and Kreuter 2012; Sala et al. 2012; Sakshaug and Antoni 2017).

One of the primary reasons why respondents decline the linkage request is due to risk perceptions of privacy and data confidentiality (Sala, Knies, and Burton 2014), which have increased over time due to the popularity of social media and other information sharing websites (Tsay-Vogel, Shanahan, and Signorielli 2018). The introduction of the GDPR also seems not to have positively affected perceptions of trust in data collectors (Bauer et al. 2021). Although surveys often go to great lengths to address data privacy and confidentiality concerns by informing respondents of the safeguards put in place to protect their data, the extent to which respondents fully understand them is unclear. Respondent understanding of the data linkage request and its underlying procedures is crucial to ensuring they make a decision that is truly informed in the sense assumed by regulations requiring linkage consent (e.g. GDPR). Thus, research is needed to shed light on respondents' understanding of the linkage consent request and the extent to which linkage consent is truly informed. The sparse literature on this topic – both qualitative and quantitative – suggests that respondent understanding of the linkage request is limited. For instance, Bates (2005) conducted cognitive interviews with 20 participants and found that misunderstanding opt-in (or active) and opt-out (or passive) linkage requests was common, and more than half the participants failed to understand the purpose of the linkage request. A key point of confusion was that participants falsely assumed a multi-way linkage, in which their survey data would be shared with other government agencies.

Thornby et al. (2018) also found through qualitative interviews with 20 participants that comprehension of opt-in linkage consent requests was low. A common issue was that participants only skimmed the information materials and didn't understand some of the terminology used to describe the linkage. Interestingly, participants whose comprehension improved during the qualitative interview become more positive about the idea of linking their own data. Jäckle et al. (2018) conducted qualitative interviews with 25 members of the UKHLS Innovation Panel, who were presented with two opt-in consent questions used in the UKHLS mainstage survey. Participants who had an "accurate understanding" of the general linkage process seemed to understand the purposes of linkage and the one-way directional flow of personal information. Misunderstandings mainly stemmed from confusion over whether government departments or third-party organizations would have access to their survey responses and/or linked data and use them for purposes other than research (e.g. surveillance, telemarketing). Reviewing the information booklet tended to improve participant understanding of the general linkage process.

The only empirical study of respondent understanding of the linkage request comes from Das and Couper (2014), who conducted a follow-up Web survey with 745 members of the Dutch LISS (Longitudinal Internet Studies for the Social Sciences) online panel. Members were informed in advance through a letter or email (the mode was experimentally crossed) that their survey data would be linked with information from Statistics Netherlands unless they objected, i.e. opted out. The text contained information about the purpose of the linkage and included data confidentiality assurances. An extended version of the text was introduced experimentally, which included more details and examples about the linkage process and data storage, access, and uses. In the follow-up Web survey, respondents were presented with seven questions that tested their knowledge of the linkage details described in the advance letter and email. The majority of respondents answered correctly to only four of the seven statements, indicating that knowledge of data linkage aspects was not particularly high. In general, respondents were knowledgeable about aspects concerning data access and the inability to trace the study results back to an individual. However, lower levels of knowledge were most apparent for the use and storage of the linked data and personal identifiers (e.g. name, gender, date of birth). Respondents who opted out of the linkage (about 5 percent

of the sample) tended to give fewer correct answers, on average, compared to those who did not opt out, suggesting that non-consenters were less knowledgeable about the aspects of linkage compared to consenters. Respondents who received the shorter version of the advance text were also less knowledgeable regarding the linkage aspects than those who received the extended text.

The sparse evidence base suggests that respondent understanding of the linkage request is rather low. This raises the question of whether respondents are sufficiently equipped to make an informed decision, or whether additional measures may be needed to aid their understanding. As stated, the empirical literature on this topic is limited to an application of passive (i.e. opt-out) consent among panel members of a single survey (Das and Couper 2014). However, with the GDPR legislation opt-in or explicit consent has become more important. Beyond data protection regulations, respondent understanding may also have important implications for linkage consent rates, as the literature suggests a positive association between the two (Das and Couper 2014; Thornby et al. 2018). However, no empirical study of opt-in consent has explored this relationship. In addition, no study has attempted to assess the variability of understanding to the linkage request across respondent subgroups. Identifying subgroups that are likely to be less knowledgeable about the proposed linkage would be useful for employing targeted interventions or additional prompts that aid in improving their understanding.

The present study adds to the empirical literature by analyzing respondents' understanding of opt-in linkage consent using data collected in two separate cross-sectional Web and telephone surveys. Using these data, we investigate the extent to which respondents understand different aspects of the linkage request and whether understanding is associated with their likelihood of providing linkage consent. In addition, we investigate correlates of understanding, including respondent background characteristics, attitudes towards privacy and willingness to consent to hypothetical linkage requests, and design features (placement and framing) of the linkage consent question in order to identify subgroups that may vary in their level of understanding.

Data and Methods

Survey Data Collection

The present study, sponsored by the German Institute for Employment Research (German abbreviation: IAB), was carried out in two separate telephone and Web survey implementations, under the theme "Challenges in the German Labor Market 2014." Both surveys used simple random samples of named individuals drawn from register data of the Federal Employment Agency of Germany (German abbreviation: BA). The register includes complete coverage of individuals of all ages who are employed in a position that is liable to social security contributions or have registered for employment support services at the BA (vom Berge, Burghardt, and Trenkle 2013). Civil servants and self-employed individuals are generally not covered in the register. The total non-coverage of the register at the time the samples were drawn is estimated to be about 11 percent of the total (non-pension eligible) German labor force between the ages of 15 and 64 (Bundesagentur für Arbeit 2012; Statistisches Bundesamt 2015). All sample members were sent an advance letter which included information about the study and a data protection statement (discussed in more detail below).

The telephone sample (n = 7,001) was fielded between 9th October – 19th November 2014 and produced

677 interviews for an overall response rate of 9.7 percent (Response Rate 1; AAPOR 2016). The Web sample (n = 4,952) was fielded from 11th November 2014 – 12th February 2015 and yielded 651 completions, for a response rate of 13.2 percent (Response Rate 1; AAPOR 2016). Both response rates are low but comparable to those reported in other Web and telephone surveys (Manfreda, Bosnjak, Berzelak, Haas, and Vehovar 2008; Kennedy and Hartig 2019). Both survey questionnaires covered several substantive topics, including past employment and job seeking activities, social media usage, privacy attitudes, among others. Researchers may apply for access to these data by submitting an application to the IAB's Research Data Centre (https://fdz.iab.de/en/FDZ_Data_Access/FDZ_On-Site_Use.aspx).

Data Protection Statement, Consent Request, and Knowledge Assessment

Enclosed in the advance letter was a general data protection statement (see appendix) which listed several points describing how the collected survey data would be used and safeguarded. The key points were the following:

- Your answers will be saved without your name and without your address (i.e. in anonymous form) in the dataset.
- Names and addresses are kept separate from the survey data and only held until the end of the survey, after which they are deleted.
- All questionnaires will be analyzed (without name and address) to produce aggregate statistics (e.g. counts, percentages).
- The overall study results and results for subgroups (e.g. men, women) are published in tabular form. Details of individual persons are not recognizable.

The data protection statement did not specifically address the issue of data linkage. The linkage request was presented during the survey and included assurances of anonymity, aggregation of the study results, and that all data protection regulations would be adhered to (English translation):

> "We would like to include in the analysis data from the Federal Employment Agency, which are available at the Institute for Employment Research in Nuremberg. For example, this may include additional information about periods of employment, unemployment, or participation in unemployment provisions. For the linkage of this data to the interview data, we would like to ask for your consent. In the analysis, all data protection regulations will be strictly adhered to, which means that the results are always anonymous and no conclusions drawn about yourself will be permitted. Your consent is of course voluntary. You can also revoke it at any time.
>
> [Experimental manipulation of placement (beginning/end) and framing (gain loss)]: The information that you will give us (have already given us) in the course of the interview will be more (less) useful if you agree (disagree) to link with the data of the Federal Employment Agency.
>
> Do you consent to the linkage of this information?"

The placement and framing of the linkage consent request were both experimentally varied. The linkage request was administered either at the beginning or end of the survey and an introductory gain- or loss-framing sentence was presented immediately prior to the linkage consent question. Further details of the experiments are described in Sakshaug et al. (2019). Overall, about 82 and 77 percent of telephone and Web survey respondents consented to the linkage request, respectively (see Sakshaug et al. 2019).

At the very end of both surveys, respondents were presented with several knowledge questions which tested their grasp of the relevant points covered in the data protection and linkage consent statements. The questions were preceded by an introductory statement, which slightly differed depending on whether the respondent had previously consented to the linkage:

> "[FOR CONSENTERS] Previously, you gave us your consent to the linking of your answers from the survey to data of the Institute for Employment Research, or IAB for short. We would now be interested to know what you think actually happens to your answers from the survey, after you have given us your consent to the linkage."

> "[FOR NON-CONSENTERS] Previously, you did not give us your consent to the linking of your answers from the survey to the data of the Institute for Employment Research, or IAB for short. We would now be interested to know what you think actually happens to your answers from the survey, after you have not given us consent to the linkage."

A total of seven knowledge questions were asked, though the exact number varied by respondent as some questions were only relevant for those who consented or did not consent to the linkage. The Web survey contained one fewer knowledge question than the telephone survey for design reasons unrelated to the present study. The questions, presented in the Results section (Table 1), could be answered with a response of yes or no. The questions were phrased so that the true answer was not always "yes" or "no" in order to avoid acquiescence bias or straightlining threats. Respondents were also offered "don't know", "do not understand question", and "no answer" options.

Statistical Analysis

To assess respondents' levels of understanding for each item, we report the percentage of correct answers given to each knowledge question separately. The percentages are reported separately for respondents who consented and did not consent to data linkage in order to assess whether understanding differs between these two groups, as is suggested in the literature. To identify correlates of understanding, we fit separate logistic regression models of giving a correct answer for each knowledge question on respondent characteristics and experimental design features of the linkage consent question. These analyses are weighted to account for nonresponse in each survey. The weighting variables included sex, age, education, and employment status.

The regression covariates include background variables: sex, age, education, German born; an employment variable: current (un)employment status; private internet access (telephone only); privacy attitudes and social media use (Web only); consent behavior: consent to audio interview recording (telephone only), consent to up to five hypothetical linkage requests (telephone only), consent to actual linkage; reported receipt of advance letter (telephone only); and experimental design features of the consent question: framing (gain vs. loss) and placement (beginning vs. end). Descriptive statistics for each variable are provided in Appendix Tables A1 (telephone) and A2 (Web).

Some of these variables are hypothesized to be related to respondents' understanding of the linkage request. For instance, demographic characteristics are known predictors of linkage consent (Mostafa, 2016; Sakshaug et al. 2012; Sala et al. 2012; Knies and Burton 2014; Jenkins et al. 2006; Young et al. 2001; Haider and Solon 2000). Education, in particular, has been found to be positively related to subjects' knowledge and understanding of scientific research practices (Benson et al. 1988; Flory and

Emanuel 2004; Dunn and Jeste 2001; Cervo et al. 2013). Being (un)employed or receiving benefits connected to unemployment has also been shown to be related to linkage consent (Haider and Solon 2000; Jenkins et al. 2006; Sakshaug et al. 2012) and, in our study, may be an indicator of experience with/knowledge of the Federal Employment Agency, which is the administrative data holder and the agency responsible for registering one's unemployment status in Germany. Respondents with private internet access and those who use social media are expected to have more understanding of linkage requests, as they may be more familiar with how personal data are used and protected. In contrast, people without internet access and those who place a very high importance on privacy are expected to be less knowledgeable and perhaps more skeptical about how their personal data will be used and safeguarded. Consent to hypothetical and actual data linkage requests (and audio recording of the interview) are expected to be positively related to understanding based on the reviewed literature (Das and Couper 2014; Thornby et al. 2018).

Having reported not receiving the advance letter is expected to be negatively related to understanding, as these respondents are unlikely to have been exposed to the data protection statement. Administering the linkage consent question at the beginning of the survey (as opposed to the end) is expected to have a negative effect on understanding, as the knowledge questions were administered at the end of the survey, at which point the saliency of the linkage request and data protection assurances is likely to be lower. We do not have a specific hypothesis regarding the framing of the linkage consent question. The consent literature suggests that both gain framing and loss framing can have a positive effect on linkage consent (Sakshaug and Kreuter 2014; Sakshaug et al. 2019; Kreuter, Sakshaug, and Tourangeau 2016), but the potential relationship between framing and understanding of the linkage request is unclear.

Results

Level of Understanding

Table 1 shows the percentage of correct answers given to the seven knowledge questions for respondents in the telephone and Web surveys. Note that the percentages are based on all respondents, including those who provided a non-substantive answer (e.g. don't know, refusal). Non-substantive answers are treated as incorrect answers (i.e. lack of understanding) for a particular knowledge question. Percentages of non-substantive answers are shown in Appendix Table A3.

In both surveys, levels of understanding differed markedly by whether or not the respondent consented to the linkage request. Starting with the telephone survey, those who gave consent overwhelmingly understood correctly that their responses would be merged with IAB data (88.6%), but only a minority of non-consenters correctly thought that their responses would not be merged (31.8%). In other words, about 68% of non-consenters thought their responses would be linked to IAB data anyway, despite their wishes. Likewise, less than half of the non-consenters (43.4%) understood that their survey responses would be passed on to the IAB (i.e. the survey sponsor), compared to the majority of consenters (84.8%). Hence, non-consenters seem to have significantly less understanding about the linkage request and transmission of their survey data compared to the consenters. Regarding item missing data (Appendix Table A3), non-consenters have higher rates of "don't know", "did not understand question", and "no answer" responses compared to the consenters.

Table 1. Percentage (and Standard Errors) of Correct Answers Given to Knowledge Questions.

	Telephone Survey			Web Survey	
Knowledge Questions	Consenters	Non-consenters		Consenters	Non-consenters
Q1) Do you think that your answers will be passed on to the IAB? (Y)	84.8 (1.6)	43.4 (4.5)		N/A	N/A
Q2) Do you think that your answers will be merged with the data of the IAB? (Y, Consenters; N, Non-consenters)	88.6 (1.4)	31.8 (4.2)		81.6 (1.8)	13.4 (2.9)
Q3) When your answers are merged with the information available about you at the IAB…					
Q3a) …Do you think that your name and address are stored together with the linked data? (N)	63.5 (2.1)	N/A		54.9 (2.3)	N/A
Q3b) …Do you think that the results of the analyses could allow conclusions to be drawn about you as a person? (N)	60.0 (2.1)	N/A		48.8 (2.4)	N/A
Q3c) …Do you think that the linked data are kept exclusively in the IAB and only analyzed for research projects in anonymized form? (Y)	82.4 (1.6)	N/A		65.2 (2.2)	N/A
Q3d) …Do you think the linked data will be made public without anonymization? (N)	85.7 (1.5)	N/A		81.9 (1.8)	N/A
Q4) Do you think that your personal data at the IAB will be destroyed? (N)	N/A	52.8 (4.6)		N/A	40.6 (4.2)
N	553	124		503	148

Note: N/A means item was not asked. Results are weighted for nonresponse. Correct answers (Y/N) shown in parentheses.

Among the linkage consenters, most correctly understood that the linked data would not be made public without anonymization (85.7%), would be kept exclusively at the IAB and only analyzed for research projects in anonymized form (82.4%), would not be stored together with their name and address (63.5%), and would not be used to produce results that could be used to draw conclusions about them as a person (59.7%). Among the non-consenters, about half (52.4%) of them correctly understood that their personal data at the IAB would not be destroyed.

The Web survey results largely mirror those of the telephone survey, but with generally lower percentages of correct answers for consenters and non-consenters. A majority of consenters (81.6%) understood that their survey responses would be linked to IAB data, whereas only a small minority of non-consenters (13.4%) believed that their consent decision would be respected and their responses not linked to IAB data. Among the linkage consenters, most understood that the linked data would not be made available to the public without anonymization (81.9%), that the linked data would be kept exclusively at the IAB and used only for research purposes (65.2%), and that their name and address would not be stored together with the linked data (54.9%), but less than half thought the linked data would not produce results that could be used to draw conclusions about them as a person (48.8%). Also, less than half (40.6%) of the non-consenters correctly answered that their personal information at the IAB would not be destroyed. Similar to the telephone survey, non-consenters in the Web survey have higher rates of “don’t know”, “did not understand question”, and “no answer” responses compared to consenters (Appendix Table A3).

Correlates of Understanding

Regressions of giving a correct answer to each of the knowledge questions are presented in Table 2 (telephone survey) and Table 3 (Web survey). In both surveys, no variables are consistently associated with giving correct answers to all knowledge questions. In the telephone survey, males have lower levels of understanding about the linkage request than females for two knowledge questions (Q3a/Q3b) and

individuals born in 1977 or later have a higher level of understanding than those born earlier for one knowledge question (Q3b). German-born respondents have higher understanding than their non-German born counterparts (Q1). Respondents who are unemployed have lower levels of understanding for two knowledge questions (Q3a/Q3c). Having private internet access is positively associated with several indicators of understanding (Q1/Q3a/Q4). Rating privacy as being “very important” is negatively associated with one understanding indicator (Q3d). Giving consent to hypothetical linkage requests is positively associated with understanding (Q2) as is giving actual linkage consent in the survey (Q1/Q2). There is no relationship between understanding and education, whether the advance letter was reportedly received, consent to audio recording, and the framing or placement of the linkage consent question. Multilevel models fitted with respondents clustered within interviewers yielded similar results. Further analyses of interviewer-respondent interactions during the knowledge questions are presented later.

Table 2. Telephone Survey: Logistic Regression Coefficients of Correct Answer to Knowledge Questions.

	Q1 (N=677)	Q2 (N=677)	Q3a (N=554)	Q3b (N=554)	Q3c (N=554)	Q3d (N=554)	Q4 (N=123)
Intercept	-1.67**	-1.21	0.01	0.23	1.84*	1.32	-2.27
Sex							
Male	-0.18	-0.19	-0.41*	-0.40*	-0.43	0.15	-0.16
Female (REF)	--	--	--	--	--	--	--
Age (birth year)							
1938-1958 (REF)	--	--	--	--	--	--	--
1959-1965	-0.04	-0.32	0.20	-0.11	-0.34	0.56	1.10
1966-1976	-0.15	-0.42	-0.44	0.04	0.18	0.49	0.33
1977+	-0.02	-0.01	-0.16	0.54*	0.07	0.42	0.87
Education							
Low (REF)	--	--	--	--	--	--	--
Middle	-0.05	-0.01	0.07	-0.16	0.20	0.06	0.11
High	-0.20	0.15	0.49	-0.09	0.31	0.51	-0.39
German born							
Yes	0.63*	0.27	0.44	0.09	-0.16	0.04	0.22
No (REF)	--	--	--	--	--	--	--
Unemployed							
Yes	-0.06	-0.38	-0.64**	-0.36	-0.73*	-0.48	-0.14
No (REF)	--	--	--	--	--	--	--
Private internet access							
Yes	0.91**	0.63	0.74*	0.32	0.03	0.29	1.94**
No (REF)	--	--	--	--	--	--	--
Privacy very important							
Yes	0.30	0.02	0.05	0.03	0.05	-0.65*	-0.45
No (REF)	--	--	--	--	--	--	--
Hypothetical linkage consent (index; 1-5)	0.03	0.19**	-0.02	0.02	0.12	0.03	0.16
Received letter							
Yes (REF)	--	--	--	--	--	--	--
No	0.01	0.04	-0.06	0.15	0.56	-0.03	1.04
Consent recording							
Yes (REF)	--	--	--	--	--	--	--
No	0.11	-0.42	-0.23	-0.37	-0.56	-0.41	0.81
Actual linkage consent							
Yes	2.05***	2.50***	N/A	N/A	N/A	N/A	N/A
No (REF)	--	--					
Consent Q placement							
Front	-0.15	0.04	-0.28	-0.05	-0.13	-0.10	-0.83
Back (REF)	--	--	--	--	--	--	--

Consent Q framing							
Gain	0.28	-0.08	0.00	0.11	-0.18	0.17	0.53
Loss (REF)	--	--	--	--	--	--	--
Model fit statistics							
Cox & Snell's R^2	0.141	0.230	0.075	0.034	0.042	0.040	0.206
Nagelkerke's R^2	0.213	0.353	0.103	0.046	0.069	0.070	0.275
AIC	658.433	567.562	714.879	757.275	526.181	466.765	173.222

*Notes: *p<0.05; **p<0.01; ***p<0.001. Results are weighted for nonresponse. Descriptive statistics for each variable are provided in Appendix Table A1. Q1: Answers passed on to IAB; Q2: Answers merged with IAB data; Q3a: Name/address stored with linked data; Q3b: Conclusions drawn about person; Q3c: Linked data kept at IAB in anonymized form; Q3d: Linked data made public without anonymization; Q4: Personal data at IAB destroyed.*

In the Web survey, males have less understanding about the linkage request compared to females for one indicator (Q3c). Respondents born between 1966-1976 have more understanding than those born between 1937-1958 (Q4). Having "middle" or "high" education is positively associated with understanding for the majority of indicators (Q2/Q3a/Q3c/Q3d). Respondents who report regularly using social media are comparable to those who report not using it; however, those who did not answer this item have significantly less understanding than those who regularly use social media (Q2/Q3d/Q4). Consenting to data linkage is strongly and positively associated with understanding (Q2). Gain framing the linkage consent question (as opposed to loss framing) and positioning the consent question at the front of the questionnaire (as opposed to the back) are both associated with having less understanding for one indicator (Q4). There is no association between understanding and being born in Germany, (un)employment status, or the importance of privacy.

Table 3. Web Survey: Logistic Regression Coefficients of Correct Answer to Knowledge Questions.

	Q2 (N=651)	Q3a (N=503)	Q3b (N=503)	Q3c (N=503)	Q3d (N=503)	Q4 (N=148)
Intercept	-2.89***	-0.01	1.02	1.17	-0.16	-0.44
Sex						
Male	-0.12	-0.22	-0.29	-0.55**	-0.06	-0.04
Female (REF)	--	--	--	--	--	--
Age (birth year)						
1937-1958 (REF)	--	--	--	--	--	--
1959-1965	-0.05	0.14	-0.44	0.07	0.28	0.02
1966-1976	-0.18	-0.04	-0.50	-0.22	-0.15	1.34*
1977+	0.09	0.02	-0.26	-0.29	0.08	0.22
Education						
Low (REF)	--	--	--	--	--	--
Middle	1.31***	0.45	0.29	0.62*	1.18***	0.32
High	0.98**	1.01***	0.44	0.73*	1.36***	0.19
German born						
Yes	0.62	-0.09	-0.79	-0.51	0.96	0.33
No (REF)	--	--	--	--	--	--
Unemployed						
Yes	0.05	-0.12	-0.05	0.44	0.52	-1.00
No (REF)	--	--	--	--	--	--
Regularly use social media						
Yes (REF)	--	--	--	--	--	--
No	-0.31	0.19	-0.31	0.26	-0.41	-0.04
Missing	-1.32***	-0.44	-0.60	-0.47	-1.09**	-1.42*
Privacy very important						
Yes	-0.10	-0.25	0.08	-0.19	-0.27	0.75
No (REF)	--	--	--	--	--	--
Actual linkage consent						
Yes	3.36***	N/A	N/A	N/A	N/A	N/A
No (REF)	--					
Consent Q placement						
Front	0.03	-0.10	-0.22	-0.38	0.38	-0.93*
Back (REF)	--	--	--	--	--	--
Consent Q framing						
Gain	-0.15	-0.01	0.15	0.04	-0.15	-1.01*
Loss (REF)	--	--	--	--	--	--
Model fit statistics						
Cox & Snell's R^2	0.352	0.051	0.032	0.059	0.082	0.233
Nagelkerke's R^2	0.488	0.069	0.043	0.082	0.135	0.315
AIC	579.481	694.911	709.309	647.106	460.161	184.76

*Notes: *p<0.05; **p<0.01; ***p<0.001. Results are weighted for nonresponse. Descriptive statistics for each variable are provided in Appendix Table A2. Q2: Answers merged with IAB data; Q3a: Name/address stored with linked data; Q3b: Conclusions drawn about person; Q3c: Linked data kept at IAB in anonymized form; Q3d: Linked data made public without anonymization; Q4: Personal data at IAB destroyed.*

Interviewer-Respondent Behavior Coding in the Telephone Survey

In interviewer-administered surveys, the interviewer-respondent interaction may influence respondents' understanding of the linkage consent request. Administering the request to respondents requires interviewers to read a relatively long consent statement, which requires careful listening and patience on the part of the respondent. There is a risk that some interviewers might deviate from the scripted statement or add additional (unscripted) comments. Likewise, respondents may interrupt interviewers as they're reading the statement, express confusion or misunderstanding, or pose questions to the interviewer. All of these unscripted behaviors have the potential to influence respondent understanding or hinder communication of the relevant aspects of the linkage request, which could have, in turn, affected the answers given to the knowledge questions in the telephone survey. We explored this possibility by carrying out behavior coding on recorded interviews in the telephone survey. For the 573 (out of 677) telephone respondents who consented to audio recording, research assistants listened to the interviews and coded the following interviewer and respondent behaviors that occurred during the linkage consent request: interviewer deviated from reading the full consent statement, interviewer added unscripted comments to the statement, respondent asked questions, respondent expressed confusion or lack of understanding, and respondent interrupted the interviewer while reading the statement.

The results of the interviewer-respondent behavior coding are presented in Table 4. Overall, the percentage of unscripted behaviors was small. Interviewers deviated from the scripted consent statement in only 5.2% of interviews and added additional comments to the statement in 14.8% of interviews. The additional comments were overwhelmingly neutral – in only 7 interviews did the interviewer make a comment in favor or against linkage. About 8.9% of respondents asked at least one question during the linkage request, 11.3% of respondents expressed confusion or lack of understanding to the request, and 10.7% of respondents interrupted the interviewer as they were reading the statement. To assess whether these behaviors affected the likelihood of giving a correct answer to the knowledge questions, we fitted separate regression models for each knowledge question on the subset of telephone respondents with audio recorded interviews. The behavior coding variables were included in the models in addition to the aforementioned covariates used in the regression model from Table 2. Overall, we found that the behavior codes were not associated with correctly answering the knowledge questions (results not shown). Thus, the unscripted interviewer/respondent behaviors don't seem to be related to differential understanding.

Table 4. Distribution of Interviewer and Respondent Behaviors During the Linkage Consent Request.

Type of behavior	N	% (SE)
Reading of consent statement		
No deviation	543	94.8 (0.9)
Some deviation	30	5.2 (0.9)
Interviewer additional comments		
In favor of consent	3	0.5 (0.3)
Against consent	4	0.7 (0.4)
Neutral	78	13.6 (1.4)
No additional comments	488	85.2 (1.5)
Respondent asked questions		
Yes	51	8.9 (1.2)
No	522	91.1 (1.2)
Respondent expressed confusion or lack of understanding		
Yes	59	11.3 (1.3)
No	514	89.7 (1.3)
Respondent interrupted interviewer		
Yes	61	10.7 (1.3)
No	512	89.4 (1.3)

Notes: Results are based on persons who consented to telephone audio recording. Percentages are unweighted.

Discussion

Summary of Results

This study investigated the extent to which respondents understand various aspects of the data linkage consent request. The results yielded two key findings. First, we identified a strong positive relationship between understanding the data linkage request and consenting to data linkage in separate telephone and Web surveys. Overall, the percentages of correct answers given to several knowledge questions were generally high for consenting respondents indicating a good understanding of relevant aspects of the data linkage request, including anonymization of the linked data, restricting data access, and storage of personally identifiable information. In contrast, the percentages of correct answers given by non-consenting respondents were comparatively lower, indicating poorer levels of understanding among this group. In particular, less than one-third of non-consenting respondents correctly answered that their interview data would not be merged with administrative data; in other words, the majority of non-consenters thought their survey data would be linked anyway, against their wishes. The patterns of understanding were similar between the Web and telephone surveys, though the percentages of correct answers in the Web survey were between 4-18 percentage points lower compared to the telephone survey.

And second, we identified several correlates of understanding beyond giving linkage consent. Giving an incorrect answer to at least one of the knowledge questions was associated with sex (males), older birth cohorts, lower levels of education (Web only), non-German born (telephone only), being unemployed

(telephone only), not having private internet access (telephone only), rating privacy as being “very important” (telephone only), unwillingness to consent to hypothetical linkage requests (telephone only), and unwillingness to answer a social media usage item (Web only). Furthermore, two experimental design features of the linkage consent question were correlated with understanding, particularly in the Web survey. Specifically, placing the consent question at the beginning of the questionnaire (as opposed to the end) and framing the question in terms of gains (as opposed to losses) were both associated with higher levels of misunderstanding regarding the storage of personal data.

We acknowledge some limitations of the study. First, the population of inference consists of employed persons who are liable to social security contributions or have registered for employment support services. Thus, one should be cautious about generalizing the results to the full (German) population. Second, the response rates of both surveys were relatively low which may also affect inference. And, third, we analyzed individual items, not scales, that differed between consenters and non-consenters. Further work is needed here to develop scales that both consenters and non-consenters can answer.

Implications for Practice

These results indicate a strong connection between understanding data linkage consent and respondents’ likelihood of actually giving consent, which is consistent with findings from the qualitative and passive (opt-out) consent literature (Das and Couper 2014; Thornby et al. 2018). On the one hand, this is an encouraging finding for survey research as it suggests that people who give consent are likely making an informed decision to do so. On the other hand, it raises the question of whether respondents who do not consent are making an informed decision when they decline the linkage request, or simply do not trust the survey sponsor to comply with their data protection assurances. If low levels of respondent understanding of the linkage request are an impediment to obtaining consent, then additional efforts to improve respondent understanding may pay off in terms of increasing linkage consent rates and more closely adhering to the principles of informed consent as defined by data protection regulations, such as those described in the GDPR. Aspects of respondent understanding that are ripe for improvement seem to be those related to ensuring that the wishes of non-consenters will be respected, that personal identifiers will be removed from the linked data, and that the study results cannot be traced back to any one individual.

One approach for improving these aspects would be to target key subgroups that are less likely to understand the linkage request with additional explanation or informational material. Our analysis points to several subgroups that may be more susceptible to misunderstanding of the linkage request. For instance, respondents with item missing data, those who rate privacy as being a very important issue, and those who are unwilling to consent to hypothetical linkages may benefit from additional information about the data linkage process to improve their understanding and address possible points of concern. Another approach would be to employ tailored messages by asking the consent understanding questions first, then doing a targeted intervention to address any misunderstandings prior to administering the linkage request. However, identifying key subgroups and applying targeted interventions prior to the linkage request would require that linkage consent be asked later in the questionnaire, and not directly at the beginning. While many experimental studies (including the present data source; see Sakshaug et al. 2019) have found that asking for linkage consent at the front yields higher consent rates than at the back (Sakshaug and Vicari 2018; Sakshaug, Tutz, and Kreuter 2013), our results showed that this approach may lead to more misunderstandings in a Web survey. Although these results could be due to the placement of the knowledge questions at the end of the survey, it may be beneficial to identify the relevant subgroups and apply any interventions as early as possible in the survey, followed by the linkage

consent request, to maximize both the consent rate and consent understanding.

Conclusions and Future Research

In conclusion, this is the first study to empirically examine respondent understanding of opt-in data linkage requests in a survey. Overall, it is apparent that respondents vary in their understanding of various aspects of the linkage request and misunderstandings seem to be highly correlated with not giving consent. Respondent understanding of the linkage request could be improved, particularly for certain subgroups. Efforts to improve understanding are important, not only for ensuring that respondents make a more informed decision, but also such efforts may be useful for improving linkage consent rates and minimizing the risk of non-linkage error. However, more research is needed to shed further light on this potential relationship, which may aid in developing a theory of understanding. Future research could also test the causal relationship between understanding and consent (i.e. does increasing understanding increase consent?) or what drives the relationship between understanding and consent (e.g. if the respondent is not going to consent, will they invest the energy and effort into informing themselves about the process?), as well as the factors (e.g. trust) that drive both knowledge and consent.

Appendix

Data Protection Statement Included in Advance Letter

References

1. The American Association for Public Opinion Research (AAPOR). (2016*). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 9th edition. AAPOR.
2. Bates, N. (2005). Development and Testing of Informed Consent Questions to Link Survey Data with Administrative Records. *Proceedings of the Survey Research Methods Section* (pp. 3786-3793). American Statistical Association.
3. Bauer, P.C., Gerdon, F., Keusch, F., Kreuter, F., and Vannette, D. (2021). Did the GDPR Increase Trust in Data Collectors? Evidence from Observational and Experimental Data. *Information, Communication & Society*. Advance online access.
4. Benson, P. R., Roth, L. H., Appelbaum, P. S., Lidz, C. W., and Winslade, W. J. (1988). Information Disclosure, Subject Understanding, and Informed Consent in Psychiatric Research. *Law and Human Behavior*, 12(4), 455-475.
5. Buck, N. and McFall, S. (2012). Understanding Society: Design Overview. *Longitudinal and Life Course Studies*, 3(1), 5–17.
6. Bundesagentur für Arbeit. (2012). Statistik: Dokumentation “Bezugsgröße 2012”. Retrieved from https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Definitionen/Berechnung-Arbeitslosenquote/Dokumentation/Generische-Publikationen/Dokumentation-der-Bezugsgroesse-2012.pdf?__blob=publicationFile&v=3.
7. Cervo, S., Rovina, J., Talamini, R., Perin, T., Canzonieri, V., De Paoli, P., and Steffan, A. (2013). An Effective Multisource Informed Consent Procedure for Research and Clinical Practice: An Observational Study of Patient Understanding and Awareness of Their Roles as Research Stakeholders in a Cancer Biobank. *BMC Medical Ethics*, 14(30).
8. Das, M., and Couper, M. P. (2014). Optimizing Opt-Out Consent for Record Linkage. *Journal of Official Statistics*, 30(3), 479-497.

9. Dunn, L., and Jeste, D. (2001). Enhancing Informed Consent for Research and Treatment. *Neuropsychopharmacoly*, 24(6), 595–607.
10. Flory, J., and Emanuel, E. (2004). Interventions to Improve Research Participants' Understanding in Informed Consent for Research: A Systematic Review. *Journal of the American Medical Association*, 292(13), 1593-1601.
11. Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel (SOEP), *Jahrbücher für Nationalökonomie und Statistik*, 239(2), 345-360.
12. Haider, S., and Solon, G. (2000). Non Random Selection in the HRS Social Security Earnings Sample. Santa Monica, CA: RAND Corporation.
13. Jäckle, A., Beninger, K., Burton, J., and Couper, M. P. (2018). Understanding Data Linkage Consent in Longitudinal Surveys. *Understanding Society Working Paper Series 2018-07*, Understanding Society at the Institute for Social and Economic Research.
14. Jenkins, S. P., Cappellari, L., Lynn, P., Jäckle, A., and Sala, E. (2006). Patterns of Consent: Evidence from a General Household Survey. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(4), 701–722.
15. Kennedy, C., and Hartig, H. (2019). Response Rates in Telephone Surveys Have Resumed Their Decline. Technical Report. Pew Research Center. Retrieved from https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/.
16. Kreuter, F., Sakshaug, J. W., and Tourangeau, R. (2016). The Framing of the Record Linkage Consent Question. *International Journal of Public Opinion Research*, 28(1), 142-152.
17. Longitudinal Studies Strategic Review. (2018). 2017 Report to the Economic and Social Research Council. Retrieved from https://esrc.ukri.org/files/news- eventsand – publications/ publications/longitudinal – studies – strategic-review-2017/.
18. Manfreda, K.L., Bosnjak, M., Berzelak, J., Haas, I., and Vehovar, V. (2008). Web Surveys Versus Other Survey Modes: A Meta-Analysis Comparing Response Rates. *International Journal of Market Research*, 50(1), 79-104.
19. Mostafa, T. (2016). Variation Within Households in Consent to Link Survey Data to Administrative Records: Evidence from the UK Millennium Cohort Study. *International Journal of Social Research Methodology*, 19(3), 355-375.
20. Sakshaug, J.W., and Antoni, M. (2017). Errors in Linking Survey and Administrative Data. In P.P. Biemer, E. Leeuw, S. Eckman, B. Edwards, F. Kreuter, L.E. Lyberg, N.C. Tucker, and B.T. West (Eds.), *Total Survey Error in Practice* (pp. 557-573), John Wiley & Sons.
21. Sakshaug, J. W., Couper, M. P., Ofstedal, M. B., and Weir, D. R. (2012). Linking Survey and Administrative Records: Mechanisms of Consent. *Sociological Methods & Research*, 41(4), 535-569.
22. Sakshaug, J. W., and Kreuter, F. (2012). Assessing the Magnitude of Non-Consent Biases in Linked Survey and Administrative Data. *Survey Research Methods*, 6(2), 113-122.
23. Sakshaug, J.W., and Kreuter, F. (2014). The Effect of Benefit Wording on Consent to Link Survey and Administrative Records in a Web Survey. *Public Opinion Quarterly*, 78(1), 166-176.
24. Sakshaug, J.W., Tutz, V., and Kreuter, F. (2013). Placement, Wording, and Interviewers: Identifying Correlates of Consent to Link Survey and Administrative Data. *Survey Research Methods*, 7(2), 133-144.
25. Sakshaug, J. W., Schmucker, A., Kreuter, F., Couper, M. P., and Singer, E. (2019). The Effect of Framing and Placement on Linkage Consent. *Public Opinion Quarterly*, 83(S1), 289-308.
26. Sakshaug, J.W., and Vicari, B. (2018). Obtaining Record Linkage Consent from Establishments: The Impact of Question Placement on Consent Rates and Bias. *Journal of Survey Statistics and Methodology*, 6(1), 46-71.
27. Sala, E., Burton, J., and Knies, G. (2012). Correlates of Obtaining Informed Consent to Data Linkage: Respondent, Interview, and Interviewer Characteristics. *Sociological Methods & Research*, 41(3),

414-439.
28. Sala, E., Knies, G., and Burton, J. (2014). Propensity to Consent to Data Linkage: Experimental Evidence on the Role of Three Survey Design Features in a UK Longitudinal Panel. *International Journal of Social Research Methodology*, 17(5), 455-473.
29. Statistisches Bundesamt. (2015). Bevölkerung: Deutschland, Stichtag, Altersjahre. Wiesbaden.
30. Thornby, M., Calderwood, L., Kotecha, M., Beninger, K., and 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.
31. Tsay-Vogel, M., Shanahan, J., and Signorielli, N. (2018). Social Media Cultivating Perceptions of Privacy: A 5-Year Analysis of Privacy Attitudes and Self-Disclosure Behaviors Among Facebook Users. *New Media & Society*, 20(1), 141-161.US Commission on Evidence-Based Policymaking. (2017). The Promise of Evidence-Based Policy-Making: Report of the Commission on Evidence-Based Policymaking. Retrieved from https://journals.sagepub.com/doi/10.1177/1461444816660731.
32. US National Academies of Sciences, Engineering, and Medicine. (2017). Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps. Washington, DC: The National Academies Press.
33. Vom Berge, P., Burghardt, A., and Trenkle, S. (2013). Sample of Integrated Labour Market Biographies: Regional File 1975-2010 (SIAB-R 7510). (FDZ-Datenreport, 09/2013 (en)), Nürnberg.
34. Young, A. F., Dobson, A. J., and Byles, J. E. (2001). Health Services Research using Linked Records: Who Consents and What is the Gain? *Australian and New Zealand Journal of Public Health*, 25(5), 417–420.