

Willingness to Consent to Data Linkage in Austria - Results of a Pilot Study on Hypothetical Willingness for Different Domains

Survey Methods: Insights from the Field, Volume 1, issue 2 (2023)

Johann Bacher, Institute of Sociology, Johannes Kepler University Linz, Austria

How to cite this article : Bacher, J. (2023). Willingness to Consent to Data Linkage in Austria - Results of a Pilot Study on Hypothetical Willingness for Different Domains. *Survey Methods: Insights from the Field*. Retrieved from <https://surveyinsights.org/?p=18071>

DOI : 10.13094/SMIF-2023-00012

Copyright : © the authors 2023. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Abstract : In surveys, attempts are increasingly made to link survey data with register, geospatial and/or social media data on an individual level. Usually, this requires informed consent to the data linkage. Respondents must agree to the linking of their survey answers to other datasets, and researchers are obligated to inform them sufficiently. In contrast to other countries, few attempts at obtaining informed consent for record linkage are available for Austria. Therefore, this pilot study investigated hypothetical willingness to consent to data linkage in the Austrian case. Respondents were asked whether they would agree to four different data linkage requests: two addressed less-sensitive domains and two more-sensitive ones. The results reveal an average willingness of 66% to consent to linkage for the less-sensitive domains and 42% for the more-sensitive ones. Furthermore, willingness to consent depends on gender, income and trust in institutions. These dependencies result in a larger record linkage consent bias if data are linked across all domains rather than just the less-sensitive ones.

1 Introduction

In surveys, attempts are increasingly made to link survey data with register, geospatial and/or social media data on an individual level. Usually, this requires obtaining informed consent for data linkage. Respondents must agree to the linking of their survey answers with other datasets, and researchers are obligated to inform them sufficiently.

Experiences in other countries show that data linkage consent rates vary widely (Eisnecker, Erhardt, Kroh, & Trübswetter, 2017; Sakshaug & Kreuter, 2012) and depend on several factors, which explains the large variation. One important factor is the request content (Beuthner, Weiß, Silber, Keusch, & Schröder, 2023; Walzenbach, Burton, Couper, Crossley, & Jäckle, 2022). As anticipated by the tailored-design approach of Dillman, Smyth, and Christian (2015), the willingness to consent decreases for more-sensitive domains (Beuthner et al., 2023; Walzenbach et al., 2022). Survey design factors (Beuthner et al., 2023; Kreuter, Sakshaug, & Tourangeau, 2016; Sakshaug & Kreuter, 2014), such as the order of the requests, the wording or incentives, also influence consent rates. However, their influence seems weaker than that of

the sensitivity of the domain. Respondents' characteristics represent a further group of influencing factors that covers socio-demographic variables on the one hand and more individual (personal) variables, such as value orientations, attitudes, personality etc., on the other.

In contrast to other countries, there are few reports of studies in Austria which ask for data linkage consent.^[1] In a recent publication, Hadler, Klösch, Reiter-Haas, and Lex (2022) examine consent to linkage with social media data in a web survey in German-speaking countries, including Austria. Gerich, Moosbrugger, and Heigl (2022) report a study where survey data was linked with registered health service data. Elsewhere, Statistics Austria implemented informed consent for a COVID-19 prevalence study (Paškvan et al., 2021) and the Institute for Advanced Studies (IHS) conducted a survey as part of the Student Social Survey 2019 in two universities (Binder, 2022a, 2022b).

This paucity of existing research motivated a web-based pilot study in Austria that surveyed the willingness of respondents to agree to (hypothetical) informed consent to data linkage. It required hypothetical multiple consent to data linkage (Walzenbach et al., 2022), including less-sensitive domains (e.g. information about education and occupational status) and more-sensitive domains (e.g. information about income and health). The respondents were asked in the survey whether they would agree, disagree or don't know. A relatively long introduction was provided to increase understanding and the willingness to consent (Sakshaug, Schmucker, Kreuter Frauke, Couper, & Holtmann, 2021). The pilot study was the first to study these domains for Austria. It captures hypothetical willingness in a general form, without specifying the data source by name. We decided on this approach in order to obtain an initial overall picture for Austria. In the following, when we report on 'willingness' regarding data linkage in our pilot study, we always mean 'hypothetical willingness', even if this attribute is not always explicitly mentioned and we refer only to 'willingness'.

This paper presents the results of the pilot study concentrating on the following research questions: (1.) How high is the average willingness to consent to data linkage? (2.) How large are the differences in the willingness to consent to data linkage for different domains? (3.) Are there differences in the willingness to consent to data linkage by respondents' characteristics? (4.) How big is the consent bias if only respondents who agree to data linkage are included in analysis?

The paper is organised in the following way. Section 2 provides a brief background and formulates expectations and, where possible, hypotheses. Section 3 introduces the data, section 4 describes the analysis plan, and section 5 reports the results. Finally, section 6 discusses the findings and draws initial conclusions.

2 Background and Expectations

As already mentioned, rates of consent to data linkage differ widely and exhibit a wide range. In their review of ten studies, Sakshaug and Kreuter (2012) reported consent rates of 24%–98% for administrative data, which covered different fields with health as a focus. More recently, Eisnecker et al. (2017) reported consent rates ranging between 32% and 92% (median = 75.5%). Their report includes six studies from Germany, whereas no German study is included in the older review of Sakshaug and Kreuter (2012). Consent rates vary between 60% and 92% (median = 78.0%) in the quoted German studies. These studies link survey data with administrative data on employment and different social benefits. In their own study (Eisnecker et al., 2017), the consent rate was 57.7% for the IAB-SOEP Migration Sample, in which respondents with a migration background were asked in the SOEP-survey to agree to link their survey responses with administrative data on their employment history from IAB (Institute for Employment

Research, Germany).

Hadler et al. (2022) cover German-speaking countries (Austria, Germany and German-speaking Switzerland). In a web survey, the respondents were asked to consent to linking their survey data with their Facebook and Twitter accounts (35% and 30%, respectively, consented to linkage). Differences^[2] by country were small and statistically insignificant (Facebook: Germany = 34%, Austria = 38%, Switzerland = 34%; Twitter: Germany = 28%, Austria = 32%, Switzerland = 33%). In another study, Gerich et al. (2022) ask 3,500 individuals for consent to link their survey data with registered data on health services in a mail survey in Upper Austria; of these, 509 (14.5%) agreed and returned the questionnaire with their social security number. No differentiation between the response rate and consent rate is possible with this study; therefore, a pure consent rate cannot be estimated. During the COVID-19 pandemic, Statistics Austria ran prevalence studies asking survey respondents to agree to be tested for SARS-CoV-2 and to use the test results in their analysis (Paškvan et al., 2021). Consent rates of more than 95% were achieved.^[3] The IHS study (Binder, 2022a, 2022b) also achieved a high consent rate of 84%^[4]. Students were asked to link their survey data with their administrative data from their university, and were promised that health data would not be linked.

With regard to research question 1, it is difficult to make a prediction about the average willingness to consent to data linkage for our pilot study because consent rates vary widely and little related research is available for Austria.

A further reason is that we asked for hypothetical consent in a general form (see below), whereas the studies referenced above asked for actual consent and specified the data source, which would be linked to the survey responses. In contrast, we specified four domains/content areas but without specifying the data source.

According to the tailored-design approach of Dillman et al. (2015), consent rates depend on perceived rewards, perceived costs and trust. Sensitive questions increase perceived costs; therefore, consent rates for more-sensitive domains will be lower than for less-sensitive requests. Existing studies and reviews (Beuthner et al., 2023; Walzenbach et al., 2022) support this conclusion. For example, Walzenbach et al. (2022) report higher consent rates for education than for receiving social benefits, state of health, energy use, or tax-related information.

Therefore, with regard to research question 2, we expect a lower willingness to consent to data linkage for more-sensitive requests than for less-sensitive requests in our pilot study (Hypothesis 1).

Whether respondents regard a topic as sensitive or not depends on several factors, such as which topic is analysed, who is asked, which survey mode is applied and in which part of the questionnaire the topic is placed (Lensvelt-Mulders, 2008; McNeeley, 2012; Tourangeau & Yan, 2007). For consent to data linkage, the order of requests seems to have an influence (Beuthner et al., 2023; Walzenbach et al., 2022), such that the consent rate may be higher for more-sensitive items if they are placed first than for less-sensitive items which are positioned last.

Whereas the first two research questions focus on consent rates, the third picks up another important methodological aspect of evaluating consent to data linkage, namely data linkage consent bias (Sakshaug & Kreuter, 2012). This bias occurs if consent to data linkage differs by respondents' characteristics. For example, if fewer older people agree to link data, they would be underrepresented in the linked file. Several studies found differences in consent rates by socio-demographic variables of respondents

(Eisnecker et al., 2017; Lüthen et al., 2022; Sakshaug & Kreuter, 2012). However, the findings are inconsistent. Age seems to be an exception; older respondents tend to decline requests more frequently (Lüthen et al., 2022). However, there are studies where no age effect or a positive age effect is identified (Eisnecker et al., 2017). Individual characteristics, such as value orientations, attitudes, personality etc., seem to have a greater effect. For example, Hadler et al. (2022) report a strong, statistically significant effect of attitudes towards pro-COVID measures on a person's willingness to link their data with their social media data, whereas they found nearly no effects of socio-demographic variables. In the context of the tailored-design approach (Dillman et al., 2015), this finding can be explained with the influence of trust on consent rates. The mentioned individual characteristics are more strongly associated with trust than socio-demographic variables.

The tailored design approach assumes that trust influences the perception and evaluation of rewards and costs. Respondents who trust that the promised rewards will be provided will participate in a survey and complete it. This kind of trust is related to the 'survey climate'. In a recent study, Silber et al. (2022) found that institutional trust influences the perception and evaluation of a survey and, subsequently, participation in it. People with a lower level of trust in institutions perceive surveys more negatively and therefore decide more frequently not to participate. In our opinion, this finding can be applied to the willingness to consent to data linkage. Respondents with a low level of trust in institutions might evaluate the responsible persons/organisation of the survey more negatively, might be less convinced of the benefits of the linkage and could fear more disadvantages arising from it. Therefore, we develop the following hypothesis:

Respondents with a higher level of institutional trust will agree more frequently to data linkage than will respondents with a lower level of institutional trust (Hypothesis 2).

In contrast to a positive effect of institutional trust on the willingness to consent to data linkage, we assume a negative effect on the willingness to consent to data linkage for those who voted for the Freedom Party in Austria because there are many dissatisfied persons in this group (e.g. Wineroither, 2021), and dissatisfaction reduces willingness to consent similarly to distrust (Silber et al., 2022). Therefore, we advance the following hypothesis:

Respondents who report voting for the Freedom Party in the last national election in Austria will agree less frequently to data linkage than will respondents who report voting for another party (Hypothesis 3).

We also expect a lower level of willingness among those respondents who do not answer questions about voting (item non-response). For them, the protection of privacy and thus anonymity are important (e.g. Cohen & Cassell, 2023), and they do not answer questions about voting. Therefore, the following can be assumed:

Respondents who do not answer question about voting in the last national election in Austria will agree less often to data linkage than will respondents who answered questions on voting (Hypothesis 4).

As advised by Sakshaug and Kreuter (2012), the evaluation of consent to data linkage should not only concentrate on consent rates but also on possible consent bias. Consent rates may be low but the bias small, and to the contrary, consent rates may be higher but the bias larger. Therefore, research question 4 addresses consent bias.

3 Data

The data were collected between December 2022 and January 2023 within the pilot study of the Austrian Social Survey (ASS, <https://ausdda.at/sozialer-survey-oesterreich/>). Similar to the General Social Survey (GSS, <https://gss.norc.org/>) in the United States or the ALLBUS (<https://www.gesis.org/en/allbus/allbus-home>) in Germany, the ASS is a general social survey covering different fields of life that asks questions about a wide range of topics, value orientations, attitudes towards different political topics, characteristics of social status and behavioural aspects. It also includes modules of the International Social Survey Programme (ISSP, <https://issp.org/>). It started 1986, and at the time of writing, six waves are available.

The pilot study was conducted for the seventh wave and applied a quota sampling strategy within an online access panel. Gender, age and education were used as marginal quotas. The pilot study was carried out as a web-based survey. People 18 years old or above were included. The main focus of the pilot study was to test items of a new ISSP-module on digital societies, which is part of the seventh wave of ASS and of ISSP-Programme 2024 (<https://issp.org/data-download/by-topic/>). This module covers approx. 80% of all items (99 items). The average response time was 25–30 minutes.

The fieldwork was conducted by marketagent (<https://b2b.marketagent.com>), a private market agency with ISO certificate 20252. It conducts surveys for private companies and scientific institutes and has an open-access panel of more than 2.7 million members, who are recruited in different ways. For the pilot study, 5,100 panellists were invited and 397 started the questionnaire, 394 persons agreed to data collection and processing according to General Data Protection Regulation (GDPR) and, finally, 300 answered the questions about willingness to consent to data linkage, which were placed at the end of the questionnaire. The other respondents ($n = 94$, 23.9% of 394) dropped out earlier. The implementation of the pilot study was led by the Austrian team of ASS and ISSP at the University of Graz (<https://centrum-sozialforschung.uni-graz.at/de/csr/mitarbeiterinnen/>). The data analysis is based on the 300 cases.

The question regarding willingness to agree to (hypothetical) informed consent to data linkage (see Table 1) used a relatively long introduction for two reasons. Firstly, we wanted to provide sufficient information so that respondents could understand the consent request and provide informed consent. According to Sakshaug et al. (2021), understanding the data linkage request increases the probability of agreement. Secondly, we provided information connected to perceived rewards, perceived costs and trust in order to promote consent further because Sakshaug and Kreuter (2014) found, in an experimental study, that referring to benefits increased consent rates. Time savings and more interesting questions were used as perceived benefits of agreement. The promise of confidentiality and anonymity, as well as of only scientists using the data, should increase trust.

The respondents were asked four different requests. Request for data linkage on education and on occupation were placed first. Income and health follow. Our decision for this order was based on Walzenbach et al. (2022) and our assumption about the sensitivity of the four domains. The study of Walzenbach et al (2022) reveals that higher consent rates can be obtained if the request starts with less sensitive domains. We assumed that education and occupation are less sensitive topics than income and health on average in a representative sample. For income, all reports on sensitive topics (Krumpal, 2013; Lensvelt-Mulders, 2008; McNeeley, 2012; Tourangeau & Yan, 2007) we have reviewed confirm this assumption. Health is also mentioned as sensitive topic in the literature (Lensvelt-Mulders 2008; McNeeley 2012). However, it should be noted that Walzenbach et al. (2022) regard them as less sensitive

than income related items. In Tourangeau and Yan (2007), health related behavioral items (mainly sexual behavior) are less sensitive than income, but more sensitive than education if item nonresponse is used as indicator of sensitivity. Data on income and occupation are not reported in their paper.

Table 1: Willingness to Agree to Hypothetical Informed Consent

The duration of a survey is limited. Not all important information which scientists need for their research projects can be collected within a survey. To deal with this issue, survey data can be linked to already existing data. The linkage would take place under strict conditions in order to guarantee confidentiality and anonymity. No one can draw any conclusions about your identity. Only scientists will use the linked data. The advantages of this kind of data-linking would include more time remaining for interesting topics and a shorter survey for you to complete. Given these conditions (absolute confidentiality, anonymity, academic use only), for which of the following information would you agree to linking and for which you would disagree?			
I would agree / disagree to the linking of the following data	agree	disagree	don't know
IC1: data about my education (such as schools attended, grades or achievements)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
IC2: data about my occupation (such as occupations, working hours or start and end dates)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
IC3: data about my income (such as earned income or pension)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
IC4: data about my health (such as doctor visits or sick leave in the last year)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Source: Questionnaire ASS-ISSP2023-pilot-study

Note: English translation by the author. German version in Appendix A.

4 Analysis Plan

In order to answer the first two research questions (average willingness to consent, different levels of willingness by domain; see section 1), we computed relative frequencies for the four domains. *T*-tests for the dependent sample were used to test for the significance of differences between the requests. The items were dichotomised for statistical testing.

For the third research question, a stepwise procedure was applied. Inspired by the results of the first two research questions, we analysed the responses with a latent class model (LCA) in the first step. LCA detected four clusters (see Appendix C). The first cluster represents respondents who agreed to all four consent requests. The second cluster contains those who agreed to requests regarding the two less-sensitive domains and disagreed to requests regarding the two more-sensitive ones. The third cluster covers respondents who disagreed to all requests, and the fourth cluster contains those who did not have an opinion regarding all requests. For further analysis, we built two dependent dichotomous variables. The first variable represents the willingness to agree to all domains (1 = yes, 0 = no). The second variable captures the willingness to agree to the less-sensitive domains (1 = yes, 0 = no).

In the next step, we conducted bivariate and multivariate analyses of the influence of socio-demographic and individual variables on the willingness to consent to data linkage using logistic regression models. Table 2 provides an overview of the socio-demographic and individual variables, which were included in the analysis. Besides available socio-demographic variables, we included institutional trust variables and

reported voting behaviour in the last election.

Table 2: Descriptive Statistics for Independent Variables

Variable							Missing Values
Independent variable for multivariate analysis and for computing consent bias							
Gender		Male 50%	Female 50%				0
Age in years	18–34 12%	35–44 14%	45–59 31%	60 or older 43%			0
Higher education entrance qualification		no 62%	yes 38%				0
Living in urban area		no 25%	yes 75%				2
Income in € ^{a)}		1 ^{a)} 24%	2 ^{a)} 55%	3 ^{a)} 21%			34
Trust in institutions ^{b)}	range 0–10	median 5	mean 4.7	std. dev. 2.5			8
Voted for in last election ^{c)}	ÖVP 24%	SPÖ 16%	FPÖ 14%	Grüne 8%	other party 19%	no answer 18%	0
Additional variables for computing consent bias							
employed		no 55%	yes 45%				14
working hours per week	1-14 h 15%	15-34 22%	35-44 51%	45 or more 12%			144 ^{d)}

Source: Questionnaire ASS-ISSP2023-pilot-study, $n = 300$

Notes: a) Net household income per month, 1 = €1–1.000, 2 = €1.001–4.000, 3 = €4.011 and higher.

b) Mean of two items: trust in parliament and trust in court, Cronbach's alpha = 0.77.

c) ÖVP = Austrian People's Party; SPÖ = Social Democratic Party of Austria; FPÖ = Freedom Party of Austria, Grüne = the Greens, other party = other party inclusive of non-voters, no answer = don't want to answer question about voting. With reference to hypotheses 3 and 4, the variable was recoded into the following categories for the analysis: 1 = no answer/item non-response, 2 = FPÖ, 3 = other parties including non-voters.

d) Including 130 unemployed respondents.

Finally, we calculate the consent bias for data linkage by computing the difference in relative frequencies or means between respondents who consented (to all or only to the less-sensitive request) and the total sample:

$$BIAS_{ij,consent} = f_{ij,consent} - f_{ij,TOTAL}$$

In the case of the metric variable i , the relative frequency f_{ij} is replaced by the mean \bar{x}_i . In addition to the variables, which are part of the multivariate analysis, employment status and working hours were analysed. We were thus able to calculate the consent bias for variables of at least three domains.

However, we also included the other variables in Table 2, as the exclusion of respondents without consent to data linkage can also lead to bias in these variables, as well as in other variables not studied in this

paper.

To interpret the results of statistical tests, we use a threshold of $p = .05$. Results with a p -value less than or equal to the threshold are regarded as statistically significant. In addition, effect sizes are reported for pairwise t -tests.

5 Results

5.1 Willingness to consent

About 66% (see Table 3) agreed to the two less-sensitive requests (education, occupation) and about 42% to the more-sensitive requests (income, health). Conversely, 44% disagreed on the more-sensitive items, in contrast to 24% on the less-sensitive items. The percentage of *don't know* varies between 10% and 16%. Research question 1 about the average consent rate can be answered as follows: *the average willingness to consent to data linkage is 54%*.

Table 3: Willingness to Consent to Data Linkage

I would agree / disagree to the linking of the following data	I would agree / disagree		
	agree	disagree	don't know
IC1: data about my education (such as schools attended, grades, achievements)	65.3	23.3	11.3
IC2: data about my occupation (such as occupations, working hours, start and end dates)	66.0	24.3	9.7
IC3: data about income (such as earned income, pension)	41.7	44.3	14.0
IC4: data about my health (such as doctor visits, sick leave in the last year)	41.3	42.7	15.7
Mean—less-sensitive domains	65.7	23.8	10.5
Mean—more-sensitive domains	41.5	43.5	14.9
Mean—total	53.6	33.7	12.7

Source: ASS-ISSP2023-pilot-study, English translation; number of cases for all items $n = 300$.

Note: The numbers represent row percentages for each domain.

However, it does not make much sense to compute this average because there are large and statistically significant differences in the willingness to consent to data linkage between the less- and more-sensitive requests, whereas there are no significant differences within the two less-sensitive requests or within the two more-sensitive requests (see Table 4). *In response to research question 2 about differences in consent rates by domain, this result reveals differences according to the assumed sensitivity of the domains and confirms hypothesis 1.* However, differences may also be caused by the order of requests (see section 6).

Table 4: Paired Samples Test for Willingness to Consent

Pairs	Paired Differences Mean	SE Mean	Effect-size ^{a)}	Significance ^{b)}	
				One-Sided <i>p</i>	Two-Sided <i>p</i>
Pair 1 xic1a - xic2a	-0.007	0.019	-0.020	>0.999	>0.999
Pair 2 xic1a - xic3a	0.237	0.028	0.496	< .001	< .001
Pair 3 xic1a - xic4a	0.237	0.031	0.446	< .001	< .001
Pair 4 xic2a - xic3a	0.243	0.028	0.506	< .001	< .001
Pair 5 xic2a - xic4a	0.243	0.029	0.492	< .001	< .001
Pair 6 xic3a - xic4a	0.000	0.027	0,000	>0.999	>0.999

Source: ASS-ISSP2023-pilot-study, number of cases $n = 300$.

Note: xic1a to xic2a dichotomised items with 1 = agree and 0 = others. a) Cohen's *d*. b) Bonferroni's correction was used to adjust for multiple comparison.

5.2 Differences in the willingness to consent by respondents' characteristics

Table 5 summarises the results of bivariate and multivariate analyses in order to answer research question 3. In the bivariate case, the willingness to consent to the linking of all domains depends statistically significantly ($p < 0.05$) on gender, institutional trust, income and no answer in questions on voting behaviour. Females, persons of low income, persons with lower institutional trust and/or those who refuse to answer questions about their voting behaviour statistically significantly agree less frequently to a linkage to all domains. In the multivariate case, gender and trust in institutions remains statistically significant.

For consent to only less-sensitive domains, institutional trust and age are statistically significant in the bivariate case ($p < 0.05$). In the multivariate analysis, only trust in institutions maintains statistical significance.

Table 5: Results of Bivariate and Multivariate Analyses

Independent variables	Willingness to consent to all requests				Willingness to consent to less-sensitive requests			
	bivariate		multivariate		bivariate		multivariate	
	b	p	b	p	b	p	b	p
Female	-0.692	0.007	-0.725	0.011	-0.374	0.195	-0.240	0.429
Higher education	0.461	0.075	0.314	0.304	0.522	0.085	0.410	0.226
Urban area	0.099	0.739	-0.082	0.801	-0.214	0.537	-0.466	0.202
Trust in institutions	0.149	0.003	0.138	0.013	0.219	<.001	0.194	0.002
Household income								
low	-1.139	0.005	-0.700	0.109	-0.812	0.056	-0.394	0.393
middle	-0.109	0.732	0.135	0.703	-0.079	0.838	0.231	0.579
high	Ref.	.	Ref.	.	Ref.	.	Ref.	.
Age in years								
18–34	0.782	0.085	0.921	0.059	0.532	0.366	0.447	0.434
35–44	0.028	0.943	0.231	0.593	-0.816	0.049	-0.859	0.052
45–59	-0.419	0.164	-0.272	0.401	-0.203	0.545	-0.082	0.811
60 or older	Ref.	.	Ref.	.	Ref.	.	Ref.	.
Voted for in last election								
no answer	-0.704	0.032	-0.677	0.051	-0.404	0.152	-0.049	0.453
FPÖ	-0.142	0.346	0.149	0.359	-0.523	0.088	-0.102	0.404
others	Ref.	.	Ref.	.	Ref.	.	Ref.	.
Intercept			-0.460	0.396			0.572	0.335

Source: ASS-ISSP2023-pilot-study, number of cases $n = 250$.

Notes: The independent variables are described in Table 2 in section 4. Ref. = reference category

With regard to research question 3, hypothesis 2, which assumes a lower level of willingness to consent among respondents with a lower level of trust in institutions, is confirmed, whereas hypotheses 3 and 4 (except in the bivariate case for consent to all domains) are not confirmed. Among the analysed socio-demographic variables, few have an influence, with one exception (gender, consent to all domains) only in the bivariate case.

5.3 Consent Bias

Differences in consent rates to data linkage by respondents' characteristics reveal a consent bias towards the linked data (see Table 6). If we use only the data for respondents who agreed to all domains, the bias varies for the socio-demographic variables between 1 and 10 percentage points (pp). It is highest for the significant variables in bivariate analysis, gender and income. The percentage of females would be underestimated by 10 pp and the percentage of respondents with low income by 9 pp in the linked data. If we examine only those variables, which the request about data linkage addresses, the bias is between 1 pp (employment status, working hours of 1–14 h. and 35–44 h.) and 9 pp (low income).

For reported voting behaviour, the linked data underestimates by 7 pp the percentage of respondents who do not name the political party for which they voted. For trust in institution, the bias is 0.47 scale points. With reference to a scale from 0 to 10 points, this bias is small.

The bias for all variables decreases if we use only the less-sensitive domains for linkage. The maximal difference for the socio-demographic variables and for voting behaviour is 4 pp. The bias for trust in institutions reduces to 0.34 scale values.

With regard to research question 4, consent bias is present but may be smaller than other biases in the survey, such as bias due to unit- or item-non-response and/or measurement errors.

Table 6: Consent Bias in Linked Data

Variable		Bias for consent to all domains	Bias for consent to less-sensitive domains	Variable		Bias for consent to all domains	Bias for consent to less-sensitive domains
Gender	male	10 pp	3 pp	trust in institutions		0.47	0.34
	female	-10 pp	-3 pp				
Higher education	no	-7 pp	-4 pp	reported voting behavior	no answer	-7 pp	-4 pp
	yes	7 pp	4 pp		SPÖ	3 pp	3 pp
Urban area	no	-3 pp	-1 pp		FPÖ	0 pp	-1 pp
	yes	3 pp	1 pp		Grüne	4 pp	0 pp
Household income	low	-9 pp	-4 pp		others	1 pp	0 pp
	medium	5 pp	3 pp	ÖVP	1 pp	1 pp	
	high	4 pp	2 pp	employed	no	-1 pp	-1 pp
Age in years	18–34	3 pp	0 pp	yes	yes	1 pp	1 pp
	35–44	1 pp	-2 pp	working hours	1–14 h.	-1 pp	1 pp
	45–59	-6 pp	0 pp		15–34 h.	8 pp	2 pp
	60 or older	2 pp	3 pp		35–44 h.	1 pp	-3 pp
					45 h. or more	-8 pp	0 pp

Source: ASS-ISSP2023-pilot-study, number of cases for all items n = 252–300.

6 Discussion and Conclusion

In our Austrian pilot study, the average willingness to consent to data linkage of 54% is below the median consent rate of German studies (Eisnecker et al., 2017) and of the studies of IHS (Binder, 2022a, 2022b) and Statistics Austria (Paškvan et al., 2021), but it is higher than the consent rates for data linkage of Hadler et al. (2022). However, a comparison with other studies is difficult, because we study only hypothetical willingness to consent, whereas the other studies applied actual consent.

Furthermore, averaging is problematic, because the domain has an influence on consenting to data linkage. According to the literature on topic-sensitive issues in surveys (see section 3), more-sensitive domains yield lower consent rates. However, as already mentioned, the order of the domains plays a role. Domains that are placed first receive a higher rate of consent than do those that are placed later. For

example, Beuthner et al. (2023) examines seven data domains (administrative data, data from apps, bank data etc.) of consent in an experimental study. The average difference in the consent rate between the domain in position 1 and that in position 7 is 40 percentage points, regardless of the content of the domain. If we had placed the sensitive topics at the beginning, we would probably have received a higher level of consent. The fact that Beuthner et al. (2023) already report a clear drop between position 1 and position 2, which does not occur in our case, speaks against an explanation based solely on the order of the domains. The results of Walzenbach et al. (2022), who found smaller differences according to order in their experiment, also contradict this explanation by order. Nonetheless, this topic needs further research.

Like other studies (see section 2), our pilot study detects significant and non-significant effects of socio-demographic variables on consent. Gender is significant if consent to all domains is analysed. Females decline more frequently to consent to all requests than men. This gender effect cannot be explained by other independent variables in our analysis because gender is not associated with variables that have an influence on consent to data linkage. Gender is only associated statistically with age in our study, but age has no effect on data linkage consent. In contrast, low household income and institutional trust influence the willingness to consent but are uncorrelated with gender. The domain seems important because the effect of gender is insignificant for less-sensitive topics. However, other studies found no gender effect; therefore, further research is necessary that includes psychological and other variables which can explain gender differences.

Besides gender, household income influences the willingness to consent to data linkage. In the bivariate case, respondents with a low household income refuse statistically significantly more frequently to consent to data linkage than respondents with a high household income. In contrast to gender, trust in institutions can explain this bivariate association. This is also the case for no answer for reported voting behaviour, which is bivariately significant when consent is asked for all domains.

In accordance with the literature (see section 2), trust in institutions results in significant effects on the willingness to consent to data linkage for both constellations (i.e. willingness to consent to link all domains and consent to link less-sensitive domains).

Our study has several limitations. One is the small sample size. This might be one reason that some independent variables do not have a significant effect. For example, no answer for reported voting behaviour only just failed to reach significance, with a consent to all domains in the multivariate case. With a larger sample size, this effect would have been significant. Another reason for this lack of significance might be that variables that exert an impact on the willingness to consent to data linkage are not available or are measured only indirectly. For example, we have no measurements of whether participants understood our information about consent and whether they evaluated this information positively, neutrally or negatively. Institutional trust was measured by trust in the parliament and in the court system, other institutions, like educational system or science, were not asked. In addition, variables related to perceived costs, perceived rewards and trust in academic surveys, as well as variables on survey climate, would have played an important role in providing recommendations for research practice; however, they are missing. The data source for data linkage was also not specified. A major limitation is that this study lacks an experimental character. Therefore, we could not, for example, separate the effects of the position and sensitivity of domains. Thus, the experimental variation of some survey design factors might have been helpful. However, the inclusion of the above-mentioned variables and the implementation of experiments were not possible in the pilot study.

A further limitation results from the fact that we were unaware of some papers when we planned the pilot study. For example, Kreuter et al. (2016) show that referring to a loss in the introductory text could increase the consent rate. Therefore, it would have been beneficial to have formulated the importance of data linkage something like this: 'Without this linkage, Austrian scientists will be unable to analyse their research questions and will lose international competitiveness.' Beuthner et al. (2023) found that the wording has no influence; incentives also seem to have no effect on consent rate. This would suggest, for example, including a short introductory text. However, research ethics, the General Data Protection Regulation (GDPR) and the results of other research (reported in summary form in section 2) speak against this conclusion and suggest to use a detailed introduction.

The fact that we ask for hypothetical consent to data linkage in a general form without defining the data source is a further limitation that restricts the generalisability of the results. Generalisability is additionally reduced by the sampling compositions. The pilot study is based on a non-probability access panel, and, despite the quotas, younger respondents are underrepresented in the analysed sample. Finally, the estimation of consent to data linkage bias is based on the assumption that the values of all respondents are unbiased. This may not be the case, and it would be advantageous if 'true' values for the population were available, since it cannot be ruled out that the data set, which contains only persons with consent has smaller deviations.

Nevertheless, we think that the pilot study provides some initial insights into the (hypothetical) willingness to agree to data linkage in Austria. This research and that of Hadler et al. (2022) represent the only studies that investigate willingness to consent to data linkage in Austria, and our results confirm the findings of related studies. In the light of this research, we can recommend to attempt data linkage, whereby the domain should be well selected and one should concentrate on less-sensitive ones, as the bias may be smaller than expected. In addition, it might be useful to include questions on the survey climate in the survey, as well as on perceived costs, perceived rewards and trust in academic surveys, in order to be able to deduce proposals for targeting specific groups or for future surveys. As advised by Sakshaug and Kreuter (2012), the evaluation of consent to data linkage should not only concentrate on consent rates but also on possible consent bias. This implies collecting data with known population data in the survey. Obviously, further research on this topic is needed. One fruitful research question might be whether and how it is possible, within the framework of an academic survey, to strengthen trust in research and to reduce its dependencies on trust in other institutions, like politics. In any case, we recommended integrating experiments into future studies in order to deepen and expand current knowledge about consent rates and the factors influencing them.

Appendix A: Question on Willingness to Consent in German Language

Appendix B: Questions on Trust and Reported Voting Behaviour

Appendix C: Results of LCA

Endnotes

[1] Results are provided in the next section.

[2] Information provided by Markus Hadler (University of Graz).

[3] Information provided by Matea Paškvan (Statistics Austria). The consent rates for the prevalence

studies were as follows: 95.7%–97.0% (study 1), 95.0-97.9% (study 2) and 97.5-98.8% (study 3). The extensive information provided to the respondents is seen as one reason for this high level of agreement to data linkage reported by Statistics Austria. In our opinion, the reputation of Statistics Austria also contributed to this high level, as did the incentives (free COVID-19 test and free antibody test). However, the fact that presumably only people with a positive attitude toward COVID-19 measures accepted the invitation to the web-based survey and answered the first pages with the consent questions also plays a role and explains the high rate, since people with a negative attitude did not accept the invitation and did not open the first page.

[4] Information provided by David Binder (IHS).

[5] For the necessary information for consent, see Krügel (2019). In the case of an actual data linkage, persons must also be informed about the relevant aspects of the GDPR for data linkage and must agree to them. In addition, ethical aspects should be discussed and taken into account.

References

1. Bacher, J., Pöge, A., & Wenzig, K. (2008). *Clusteranalyse: Anwendungsorientierte Einführung* (3rd edition). München: Oldenbourg R.
2. Bacher, J., Pöge, A., & Wenzig, K. (2021). Unsupervised methods. In U. Engel, A. Quan-Haase, S. X. Liu, & L. Lyberg (Eds.), *Handbook of Computational Social Science, Volume 2* (pp. 334–351). London: Routledge. <https://doi.org/10.4324/9781003025245-23>
3. Beuthner, C., Weiß, B., Silber, H., Keusch, F., & Schröder, J. (2023). Consent to data linkage for different data domains – the role of question order, question wording, and incentives. *International Journal of Social Research Methodology*, 1–14. <https://doi.org/10.1080/13645579.2023.2173847>
4. Binder, D. (2022a). *Einflussfaktoren auf die Prüfungsaktivität von Studierenden der TU Graz. Ergebnisse der Pilotverknüpfung im Rahmen der Studierenden-Sozialerhebung 2019*. Wien: IHS.
5. Binder, D. (2022b). *Einflussfaktoren auf die Prüfungsaktivität von Studierenden der Universität Graz. Ergebnisse der Pilotverknüpfung im Rahmen der Studierenden-Sozialerhebung 2019*. Wien: IHS.
6. Cohen, M. J., & Cassell, K. J. (2023). Reducing Item Nonresponse to Vote-Choice Questions: Evidence from a Survey Experiment in Mexico. *Public Opinion Quarterly*. Advance online publication. <https://doi.org/10.1093/poq/nfad002>
7. Dillman, D. A. [Don A.], Smyth, J. D., & Christian, L. M. (2015). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4. ed.). Hoboken, New Jersey: Wiley.
8. Eisnecker, P. S., Erhardt, K., Kroh, M., & Trübswetter, P. (2017). *The Request for Record Linkage in the IAB-SOEP Migration Sample: SOEP Survey Papers 291: Series C*. Berlin: DIW/SOEP.
9. Gerich, J., Moosbrugger, R., & Heigl, C. (2022). Health literacy and age-related health-care utilisation: a multi-dimensional approach. *Ageing and Society*, 42(7), 1538–1559. <https://doi.org/10.1017/S0144686X20001609>
10. Hadler, M., Klösch, B., Reiter-Haas, M., & Lex, E. (2022). Combining Survey and Social Media Data: Respondents' Opinions on COVID-19 Measures and Their Willingness to Provide Their Social Media Account Information. *Frontiers in Sociology*, 7, 885784. <https://doi.org/10.3389/fsoc.2022.885784>
11. Kreuter, F., Sakshaug, J. W., & Tourangeau, R. (2016). The Framing of the Record Linkage Consent Question. *International Journal of Public Opinion Research*, 28(1), 142–152. <https://doi.org/10.1093/ijpor/edv006>
12. Krügel, S. (2019). *The informed consent as legal and ethical basis of research data production*. Swiss Centre of Expertise in the Social Sciences FORS. <https://doi.org/10.24449/FG-2019-00005>

13. Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, 47(4), 2025–2047. <https://doi.org/10.1007/s11135-011-9640-9>
14. Lensvelt-Mulders, G. (2008). Surveying sensitive topics. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *EAM book series. International handbook of survey methodology* (pp. 479–499). New York, NY: Psychology Press.
15. Lüthen, H., Schröder, C., Grabka, M. M., Goebel, J., Mika, T., Brüggmann, D., . . . Penz, H. (2022). SOEP-RV: Linking German Socio-Economic Panel Data to Pension Records. *Jahrbücher Für Nationalökonomie Und Statistik*, 242(2), 291–307. <https://doi.org/10.1515/jbnst-2021-0020>
16. McNeeley, S. (2012). Sensitive Issues in Surveys: Reducing Refusals While Increasing Reliability and Quality of Responses to Sensitive Survey Items. In L. Gideon (Ed.), *Handbook of Survey Methodology for the Social Sciences* (pp. 377–396). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-3876-2_22
17. Paškvan, M., Kowarik, A., Schrittwieser, K., Till, M., Weinauer, M., Göllner, T., . . . Kytir, J. (2021). *COVID-19 Prevalence November 2020 (SUF edition)*. Wien: AUSSDA. Retrieved from <https://doi.org/10.11587/G3C2CS>, AUSSDA, V1, UNF:6:l28SQd08cHeRAAxMjb9FTg== [fileUNF]
18. Sakshaug, J. W., & Kreuter, F. (2012). Assessing the Magnitude of Non-Consent Biases in Linked Survey and Administrative Data. Advance online publication. <https://doi.org/10.18148/srm/2012.v6i2.5094>
19. Sakshaug, J. W., & 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. <https://doi.org/10.1093/poq/nfu001>
20. Sakshaug, J. W., Schmucker, A., Kreuter Frauke, Couper, M. P., & Holtmann, L. (2021). Respondent Understanding of Data Linkage Consent. *Survey Methods: Insights from the Field*. Advance online publication. <https://doi.org/10.13094/SMIF-2021-00008>
21. Silber, H., Moy, P., Johnson, T. P., Neumann, R., Stadtmüller, S., & Repke, L. (2022). Survey participation as a function of democratic engagement, trust in institutions, and perceptions of surveys. *Social Science Quarterly*, 103(7), 1619–1632. <https://doi.org/10.1111/ssqu.13218>
22. Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859–883. <https://doi.org/10.1037/0033-2909.133.5.859>
23. Walzenbach, S., Burton, J., Couper, M. P., Crossley, T. F., & Jäckle, A. (2022). Experiments On Multiple Requests For Consent to Data Linkage in Surveys. *Journal of Survey Statistics and Methodology*. Advance online publication. <https://doi.org/10.1093/jssam/smab053>
24. Wineroither, D. M. (2021). Die Freiheitliche Partei Österreichs (FPÖ) – Trendsetterin mit Hang zur Macht. In W. Muno & C. Pfeiffer (Eds.), *Vergleichende Politikwissenschaft. Populismus an der Macht* (pp. 271–293). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-33263-1_10