

Optimizing Opt-Out Consent for Record Linkage

Marcel Das¹ and Mick P. Couper²

This article reports on a study testing the effects of different ways of administering an opt-out consent for record linkage in a probability-based Internet panel. First, we conducted cognitive interviews to explore reactions to a draft version of the opt-out consent text. Second, we conducted a two-factor experiment to test the effects of content manipulations and mode. The results indicate that the way in which respondents were informed did not have much effect on opting out. Results from a follow-up survey on attitudes regarding privacy, confidentiality, and trust, along with knowledge questions about the process of linking, showed no evidence that presenting the opt-out consent statement makes respondents more concerned about privacy. Knowledge about the aspects of record linkage is generally not high. When looking at long-term effects of sending an opt-out consent statement, we found no evidence that this leads to higher attrition or lower participation rates.

Key words: Informed consent; administrative data; probability-based Internet panel.

1. Introduction

There is growing interest in linking survey data to administrative records, whether to enhance the quality and quantity of data available on sample respondents, to reduce the response burden, to compensate for missing survey data, or for other reasons (see, e.g., [Calderwood and Lessof 2009](#)). A key question in such record linkage is whether consent must be obtained from respondents and, if so, how best to do so in order to minimize refusals and any consent bias that may result.

In the Netherlands, Statistics Netherlands (SN) makes microdata available for statistical research. Customized administrative datasets can be prepared by SN so that they can be linked to the large number of available survey datasets. Legally authorized institutes, including universities and policy-oriented institutes, can be given the relevant authorization to access these data for analysis under the Statistics Netherlands Act.

CentERdata, a research institute housed on the campus of Tilburg University (the Netherlands), administers the Longitudinal Internet Studies for the Social sciences (LISS) panel, an online panel consisting of about 5,000 households, comprising 8,000 individuals.

¹ CentERdata and Tilburg School of Economics and Management, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Email: das@uvt.nl

² Institute for Social Research, University of Michigan, P.O. Box 1248, Ann Arbor, MI 48109, U.S.A. Email: mcouper@umich.edu

Acknowledgments: We are grateful to the Associate Editor and three anonymous referees for helpful comments and suggestions for improvement. We also thank Eleanor Singer and Annette Scherpenzeel for detailed comments on the setup of the experiment and follow-up survey, and on the first draft of the article. The LISS panel data were collected by CentERdata (Tilburg University, The Netherlands) through its MESS project funded by the Netherlands Organization for Scientific Research (grant number 176.010.2005.017).

The panel is based on a probability sample of households drawn from the population register by SN. All household members aged 16 and over are asked to join the panel and participate in the monthly questionnaires. Households that could not otherwise participate are provided with a computer and Internet connection. Survey data collected in the LISS panel can easily be linked to the administrative data available at SN, since the original sample was drawn by SN.

When linking survey data to administrative records one should obey the law with respect to confidentiality and privacy issues. The Dutch Data Protection Authority (Dutch DPA; www.dutchdpa.nl) supervises the fair and lawful use and security of personal data in the Netherlands. Rules are strict, but several exceptions hold for scientific research. There is no explicit legal requirement or rule to ask respondents (again) for consent to link their survey data to administrative records, given the fact – when joining the panel – they had consented that their data can be used for (purely) scientific purposes, and linking and analysis takes place in the secure environment at SN.

However, there is also an ethical issue. Institutes such as SN and the Netherlands Institute for Social Research (SCP) commonly use an opt-out version informing the respondent about the linkage of survey data to administrative records, giving the respondent the opportunity to object. We also decided to use an opt-out version when plans were first made to link survey data collected in the LISS panel to administrative data available at SN. However, before presenting the opt-out statement to the entire panel, we conducted an experiment using only a small sample of the LISS panel to optimize the opt-out consent for record linkage.

This article adds to the limited knowledge on how best to present opt-out consent statements. The objective is to identify the optimal wording for persuading respondents to consent to survey data being linked to administrative records, to remove *unfounded* fears or distrust, but also to ensure that respondents understand what they are consenting to. The article is structured as follows. Section 2 gives some background on consent to record linkage. Section 3 describes the data and methods. The section starts with a brief introduction about the LISS panel and continues with a description of the cognitive interviews (to evaluate a draft version of the wording), the experiment fielded in the LISS panel (to test the effects of content manipulations and mode), and follow-up survey (to measure the knowledge about the linking process and test the effects on attitudes). Section 4 presents the results. Based on the results of the experiment and follow-up survey, an opt-out statement was then presented to the entire panel. Section 5 discusses the implementation and effects on attrition and responding behavior in the long(er) run. Finally, Section 6 concludes with a general discussion.

2. Background

There are a number of concerns driving the research on consent for record linkage. A primary concern is the issue of nonconsent bias, that is, the extent to which those who consent to record linkage may be different from those who do not, thereby biasing the estimates derived from the subset of the sample with linked data (see [Sakshaug and Kreuter 2012](#)). A second concern is that low consent rates may limit the sample size for analysis, increasing the variance of the estimates. A third concern is that of whether the

consent was informed, that is, the extent to which the respondent's decision to consent to linkage was based on a clear understanding of what was requested. Our article focuses on the latter two concerns.

Much of the research on consent to record linkage focuses on consent requests administered by an interviewer that require an explicit response from the respondent, whether in writing (e.g., a signature) or orally (see, e.g., [Sakshaug et al. 2012](#); [Sakshaug and Kreuter 2012](#); [Sala et al. 2012](#)). However, because of the renewed interest in self-administered survey methods such as mail and web, whether as stand-alone modes or as part of mixed-mode data collection (see [Couper 2011](#)), increased attention is being paid to ways of obtaining consent for record linkage in such modes.

There is evidence that the process of obtaining consent – and particularly the administrative requirement to document such consent – may affect respondents' willingness to consent. For example, [Singer \(1978, 2003\)](#) reported that some respondents who may have been willing to participate in a survey were not willing to sign a consent form. [Sala et al. \(2013\)](#) provide an estimate of the negative effect of having to sign a consent form: 3.9% of the sample who consented to record linkage verbally refused to sign a consent form. [Hunt et al. \(2013\)](#) found that requiring explicit opt-in consent (in the form of a reply card) prior to a mail survey significantly reduced participation. Requiring the collection of identifying data (such as social security number (SSN) or identification number) to facilitate data linkage may further lower consent rates (see, e.g., [Dahlhamer and Cox 2007](#); [Bates 2005](#)). Given the administrative burden of explicit consent – especially that of documenting the consent – and the concerns about low consent rates and nonconsent bias, researchers are exploring opt-out alternatives to the more traditional forms of explicit consent.

Writing in the context of medical research, [Junghans et al. \(2005, p. 1\)](#) stated: "Opting in is deemed ethically more defensible, as it relies on active participation of individuals, and some evidence shows that this is what patients expect. The opt-out method has come under scrutiny as it relies on both inertia and the moral assumption that most people are willing to help researchers in principle." In an experiment by [Junghans et al. \(2005\)](#), 510 patients were randomized to an opt-in (in which patients had to return the reply card or call if they wished to participate) or opt-out (in which patients had to return the reply card or call if they did not wish to participate) procedure. In both cases, patients who had consented or had not objected were subsequently contacted by phone to make an appointment. Of those in the opt-in condition, 48% returned the card or called and made appointments and 38% were seen in the clinic, while in the opt-out group, 59% made an appointment and 50% were seen in the clinic, while 20% actively opted out.

Writing similarly about medical research, [Vellinga and colleagues \(2011\)](#) note that active or opt-in consent limits participation and introduces consent bias. They argue that "if risks for the participants are very low, an opt-out arrangement or passive consent is generally the most efficient procedure without violating the option of providing choice." ([Vellinga et al. 2011, p. 1](#)).

In the field of education, research on school-based surveys has compared active consent procedures (where written parental consent is required for participation of minors) to passive consent (opt-out; where parental permission is assumed unless they explicitly object). In general, passive consent is found to result in higher participation rates and may

reduce selection bias in such research (see, e.g., [Anderman et al. 1995](#); [Ellikson and Hawes 1989](#); [Eaton et al. 2004](#); [Fendrich and Johnson 2001](#)).

The studies reviewed above are about consent to participate in research, not consent for record linkage. There is relatively little research in the survey world on opt-out consent alternatives, specifically with regard to record linkage. Two exceptions are from the US Census Bureau. In a telephone-based experiment of alternative consent wording, [Bates \(2005\)](#) found that asking for the last four digits of the social security number (SSN) increased consent rates over asking for the full 9-digit SSN (50.6% versus 36.8%) in an explicit consent request, but that framing the request as opt-out (“Do you have any objections?”) with no request for SSN further increased consent to 63.4%.

[Pascale \(2011\)](#) conducted a similar split-ballot experiment in the context of a US Census Bureau telephone survey. Where addresses were available an advance letter was sent, which included an explanation of linking plans, and instructed respondents to inform the interviewer during the interview if they did not want their data to be linked. During the telephone survey introduction, these respondents were asked if they had received the letter. If they said yes, and did not inform the interviewer that they objected to linking, this was considered implicit consent. If they did not say yes, an explicit request for consent was made. These cases were then randomized to three conditions, one mentioning accuracy as a reason for linkage, a second mentioning cost, and a third mentioning time. [Pascale \(2011\)](#) reports that 38% of household respondents who participated in the survey gave implicit consent to link and were not asked the explicit linking question. Among those who were asked the explicit linking question, she found no significant differences in the consent rates between reasons for linkage (83.0% for accuracy, 85.3% for cost, and 83.6% for time). Combining the implicit and explicit consent, 90% of the sample consented to record linkage.

The sparse research that exists suggests that opt-out (implicit) consent procedures increase consent rates over opt-in (explicit) procedures. No research as yet has documented whether they also reduce consent bias. And virtually no research exists on whether respondents who consent implicitly understand that they have consented, and what they have consented to. Two exceptions to this are the studies by [Ellikson and Hawes \(1989\)](#) and [Pascale and Mayer \(2004\)](#). The first-mentioned study shows that parents who did not return their consent form when asked for explicit consent often did not intend to withhold consent, whereas those who failed to return the form in the implicit condition did not object to their children’s participation. Pascale and Mayer find that among respondents who declined a request for consent to share the respondent’s data with other household members during later waves, more than 80% misunderstood the request, many believing it was a request for a subsequent interview with other household members.

In summary, very little experimental research exists on the effects of different ways of asking for consent to record linkage, and there is even less research on what respondents understand by the request and whether the request changes attitudes regarding confidentiality and data sharing. This article represents an attempt to begin to fill these gaps, by focusing on 1) methods of informing respondents about opt-out consent options in an Internet panel and 2) measuring knowledge and attitudes following exposure to an opt-out consent request.

3. Data and Methods

In this section we describe the methods we used to learn about how and whether respondents understand opt-out consent requests, whether opt-out rates vary by mode and question wording, and whether the opt-out consent request had any impact on attitudes to privacy and knowledge about record linkage. We first start with a detailed description of the LISS panel which plays a central role in the experiment and follow-up survey we carried out.

3.1. LISS Panel

The LISS panel is a representative sample of Dutch households who participate in monthly Internet surveys. The panel is based on a probability sample of households drawn from the population register with the help of Statistics Netherlands. Recruitment was carried out in a traditional way: First, an announcement letter was sent together with a brochure explaining the nature of the panel study. Second, an interviewer contacted the selected respondents by telephone or face to face, asking them to participate in a ten-minute interview. At the end of the interview the request to participate in the panel was made. Households that could not otherwise participate are provided with a computer and Internet connection.

Respondents *with* Internet access who consented to participate in the LISS panel received a confirmation by e-mail, as well as a letter with login code, an information booklet, and a reply card. Respondents confirmed their willingness to participate in the panel either by returning the reply card or via the Internet by using the login code provided in the letter. Respondents *without* Internet access confirmed their willingness to participate by returning the signed reply card, after which equipment and a broadband connection were provided. The confirmation procedure ensured double consent. For 63% of the total gross sample the contact person expressed willingness to participate in the panel at the end of the recruitment interview (first consent), while 48% of the total gross sample finally registered and started to participate in the monthly interviews (second consent). Detailed information about the LISS panel can be found at www.lissdata.nl or in [Scherpenzeel and Das \(2011\)](#).

The LISS panel follows changes in life course and living conditions and monitors trends in household composition, covering a broader range of topics and approaches than surveys typically cover. Panel members are provided with an incentive for each completed questionnaire. One member in the household provides the household data and updates this information at regular intervals. Researchers from the Netherlands and abroad are invited to submit research proposals, which after review and acceptance are fielded in the LISS panel free of charge. Data from the longitudinal core study as well as data from the individual research proposals are freely available for academic researchers. This yields an enormous amount of multidisciplinary data, and linking with administrative data increases the research opportunities even further.

3.2. Cognitive Interviews

To evaluate a draft version of the text including the opt-out question, we conducted cognitive interviews with eleven respondents in the fall of 2010. These cognitive

interviews were intended to provide insight into the comprehensibility and persuasiveness of the draft text, as well as into possible points for improvement. The interviews took around 40 minutes on average and were held at the premises of TNS NIPO, a market research agency located in Amsterdam, the Netherlands.

For the purpose of the cognitive interviews we did not select respondents from the LISS panel. Instead, TNS NIPO selected the respondents from their database of persons willing to participate in surveys. The selection of respondents was stratified by age, gender, and education. The youngest respondent in the sample was a 19-year-old female, and the oldest respondent was a 68-year-old female. The education level ranged from primary school to university level. The interviews were monitored by researchers from CentERdata from a separate room. Respondents were aware of that, and all gave consent to this.

At the start of each cognitive interview the respondent was informed about the LISS panel, and was asked to imagine he or she was one of the LISS panel members. The interviewer explained that members of the panel have agreed to participate in monthly interviews, complete the questionnaires voluntarily, and receive an incentive for their participation. The topics of the questionnaires vary substantially, and the collected data are only used for scientific research. All cognitive interview respondents confirmed they understood the setting before reading the draft consent text.

After the introduction the interviewer asked the respondent to read the following text:

We guarantee confidentiality in all our studies. Your answers to the questionnaires are used for scientific purposes only. We always comply with the Personal Data Protection Act. We never provide information to public agencies such as the Tax Administration or the UWV [Employee Insurances Implementing Agency].

For some surveys we complement the answers with information obtained from Statistics Netherlands (SN). This allows us to (better) answer research questions. The data are processed using secured computer systems. If you object to having your answers combined with SN data, please contact the CentERdata helpdesk: 0800 – 023 14 15 (or by e-mail: lisspanel@uvt.nl).

When read, the interviewer asked for a first impression, whether they would opt out, and a series of follow-up questions to get a clear picture of whether the respondent understood the text.

3.3. Experiment in the LISS Panel

In addition to the effect of level of detail in (or length of) a text about consent for linking, there is also the issue of which mode is best to use for communicating the information. As mentioned earlier, the LISS panel is an online panel for which (most) respondents have an e-mail account. In the cognitive interview we presented the text on paper. Instead of sending an e-mail to the panel members, sending the opt-out consent for record linkage by regular mail could be a better alternative. A letter may be more likely to be seen and/or read and it may appear more legitimate to respondents. On the other hand, sending the opt-out consent by mail may help to draw attention to the request, may increase respondents' suspicions and thus increase the opt-out rate. Moreover, it is more expensive and requires more effort in terms of logistics than sending an e-mail message.

To test the effects of content manipulations and mode we conducted a two-factor experiment in a random subsample of the LISS panel. For this experiment we randomly selected 500 households from the LISS panel. The two factors are: length of consent text (with short and extended text as levels) and mode of communication (with letter and e-mail as levels). We based the short text on that used in the cognitive interviews.

The body of the (Dutch) text contained 184 words. In the extended version we gave more examples, as well as (more) details on how the process of linking is carried out, how the linked data will be used by researchers, who exactly has access to the linked data, and where the linked data are physically stored. The body of the extended text contained 369 words. Translated versions of both the short and extended text are provided in the [Appendix](#).

We randomly assigned the selected households to one of the four conditions. Within each household the experimental condition was the same for all members. We sent every household member (aged 16+) who participates in the panel an e-mail or letter personally. We sent the letters on February 14, 2011 and e-mails one day later, to ensure as far as possible that all respondents received the text on the same day. Respondents could object to linking their records by sending an e-mail message or calling the (toll-free) number of the CentERdata helpdesk. The helpdesk staff was trained to answer questions and alleviate concerns about record linkage. If the respondent preferred to opt out, the helpdesk always confirmed by e-mail that respondents' individual records will never be used in any linking with administrative data.

3.4. Follow-Up Survey

The experiment described in the previous subsection focused on the effects of content and mode manipulations on opt-out rates. However, the objective was also to ensure that respondents understood what they were consenting to. Therefore, we included a series of questions in the monthly rounds of fieldwork two weeks after the panel members were informed about the linkage of their records to administrative data. The questions were about general attitudes regarding privacy, confidentiality, and trust, as well as attitudes toward survey organizations (part I) and some knowledge questions about the process of linking (to see whether they understood the opt-out statement; part II).

We included all 745 panel members who received an e-mail or letter in the sample for the follow-up survey. In addition, we randomly selected a separate group of 500 households in the LISS panel (containing a total of 776 panel members aged 16 and older). This latter group acted as a control group, to see whether exposure to the opt-out statement changed attitudes.

Part I replicated the questions that were used in [Singer and Couper \(2011\)](#). Both the experimental group and the control group received this part of the follow-up survey. With this part we tested the effects of presenting the consent statement on attitudes. The first question referred to personal privacy:

Overall, how concerned are you about your (personal) privacy?

We asked similar questions on concerns about violations of privacy rights by banks, credit card companies, tax authorities, research on public opinion (either by government or in

general), computers (storing large amounts of information), and continued confidentiality of information possessed by private and public organizations. Finally, we asked questions on trust that information gathered about the respondent is treated confidentially by three types of organizations.

Part II was only presented to the experimental group. The order in which part I and part II were presented to the respondents was randomized. Part II started with an introduction referring to the e-mail (or letter) that had been sent to this group previously. We then asked these respondents whether they recalled reading the opt-out statement. We presented the opt-out text in the survey (over several screens) to all respondents in the follow-up survey who indicated they had not read the e-mail (or letter), according to their original experimental treatment with respect to length of consent text (either short or extended). This was done to make sure that all respondents were informed about the record linkage before answering the knowledge question in part II. This question consisted of seven statements and was introduced as follows:

To better understand whether the explanation is perceived as clear or unclear, we present a few statements to you about the content of the [e-mail / letter]. Please specify whether you think each of the statements is true or false.

It is not an exam. You do not have to review the [e-mail / letter] again.

Exact formulations of the seven statements are presented in the Results section (Table 3). Respondents could answer with: true, false or don't know. On the basis of these seven statements we calculated a 'total score value' for each respondent, as an indication of how well the aspects in the process of record linkage are known. We constructed this score as follows. For each correct answer we assigned a score of +1, while for each incorrect answer we assigned a score of -1. For "don't know" answers, we assigned a score of 0. In this way the possible values for the total score value could range from -7 to +7, with a value of -7 for 'all answers incorrect' and a value of +7 for 'all answers correct'.

An alternative to the total score value could be a measure of perceived risk, with a score of +1 assigned if the answer the respondent gives indicates a higher perceived risk, regardless of the truth. For example, when a respondent believes name, gender, and date of birth will be sent to Statistics Netherlands, the perceived risk of disclosure is high. We assigned a score of -1 if the opposite holds, and as before we assigned a score of 0 to a "don't know" response. We defined the resulting sum score as a measure of perceived risk of disclosure. Again, this measure potentially ranges from -7 to +7.

4. Results

4.1. Cognitive Interviews

Based on the cognitive interviews, the following conclusions were drawn:

- The text does prompt respondents to consent;
- The text creates both trust and confusion;
- No one understands what they are consenting to.

Respondents generally took a trusting approach. They assumed that ‘it should be fine’, and the first reaction to the text was positive. Respondents perceived the message as providing information, rather than inquiring. The majority assumed that one could contact the helpdesk to ask questions, but did not grasp that panel members could also do so to register their objection to record linkage. While most respondents said they would immediately consent, a few expressed uncertainty.

The respondents perceived the first paragraph as clear and persuasive. Some fears or distrust only emerged after the text (and particularly the second paragraph) was read for a second time, and following further probing. Respondents took the most important message to be that anonymity is assured, but the intended message (linking the survey data to administrative data) did not come across clearly. Some thought that their data records would be linked to similar data records in the registers from other persons, “to get a more complete picture”. A similar finding is described in [Gray \(2010\)](#), based on cognitive interviews with respondents in preparation for the seventh wave of the Family and Children study.

Virtually no one in the cognitive interviews understood exactly what they were consenting to. This was mainly due to the formulation used in the text and the incomplete information about the different steps in the process of linking data. However, as was argued by [Singer et al. \(1992\)](#), providing more details and lengthy explanations might arouse respondents’ suspicions rather than alleviating their concerns. Furthermore, it does not necessarily increase the respondents’ willingness to respond. In the context of the collection of paradata – data about the process of data collection – [Singer and Couper \(2011\)](#) found that providing more information about the paradata reduces the willingness to participate in the research and, more substantially, the willingness to permit use of the paradata collected.

On the basis of the results of the cognitive interviews we fine-tuned the text for the experiment, and added some more information about the linking process (including an example) in the second paragraph (see Appendix).

4.2. Experiment in the LISS Panel

In total we sent 745 individuals (in 489 households) the opt-out consent text. At the time of sending out the letter and e-mails, we excluded eleven individuals (from eleven single-person households) from the initial selection because of unknown e-mail addresses or unknown mail addresses; one person stopped participating in the panel after we made the random selection. [Table 1](#) in the next subsection presents the distribution of individuals in the four conditions.

Out of the 745 respondents who were sent the opt-out consent text, only 38 respondents indicated they objected to linking (5.1%). Most sent a very short reply by e-mail; some asked for additional information and objected later. One respondent explicitly indicated that he has no problems with linking (which in fact was not asked for). The opt-out numbers per condition are presented in [Table 1](#). The short e-mail version resulted in the highest opt-out rate (by quite a large margin), and the extended text resulted in lower opt-out rates in the e-mail condition but not in the letter condition.

[Table 2](#) presents the results of a probit analysis in which the decision to opt out is explained by the two factors and the interaction between these factors. The results show significant main effects, as well as a significant interaction effect. The probability of

Table 1. Number and percentage of respondents who opted out across the four conditions

Mode of communication	Length of consent text		Total
	Short text	Extended text	
E-mail	18 out of 187 (9.6%)	6 out of 186 (3.2%)	24 out of 373 (6.4%)
Letter	5 out of 190 (2.6%)	9 out of 182 (4.9%)	14 out of 372 (3.8%)
Total	23 out of 377 (6.1%)	15 out of 368 (4.1%)	38 out of 745 (5.1%)

opting out is lower in the letter condition than in the e-mail condition. The same holds for the extended text, which resulted in lower opt-out rates than the shorter text. The significant interaction effect implies that the effect of an extended text is different for a letter than for an e-mail. A test for a difference in effect on the probability of opting out between ‘short text by e-mail’ and ‘extended text by letter’ does not indicate a significant difference. The same holds for ‘short text by letter’ versus ‘extended text by e-mail’. The combination of a short message and e-mail seems to be worst (see Table 1), but in general the results indicate that the way in which respondents were informed did not have much effect on opting out, and overall relatively few did so.

Note that for the experiment we randomly selected 500 households, and we sent the opt-out consent message to all panel members in these households. Conversations among household members might have induced a cluster effect, and the decision to opt out (or not) is not really an individual decision. The 38 respondents who objected were from 25 households, ten of them being a single-person household. There was one household with four (participating) members who all opted out, ten households with two (participating) members who all opted out, and four multi-person households with only one member opting out. If we base the probit analysis on household-level data (with a household being classified as opt out if any member of the household objected) the signs of the probit coefficients are the same as those presented in Table 2. However, due to the lower number of observations only the effect of mode is still significant.

4.3. Follow-Up Survey

In the discussion of the results of the follow-up survey, we refer to the experimental group as Group A and to the control group as Group B. In February and March 2011 a few panel

Table 2. Probit estimates in a model with opt out as dependent variable and factors mode of communication (Mode = 1 for letter, 0 for e-mail) and length of consent text (Length = 1 for extended, 0 for short) as independent variables, including interaction effect

Variable	Estimated coefficient	Z	P > z
Constant term	− 1.30	− 10.3	0.000
Mode (1 = letter)	− 0.635	− 2.78	0.005
Length (1 = extended)	− 0.545	− 2.49	0.013
Mode * Length	0.833	2.52	0.012

members from both groups A and B indicated they wanted to stop participating in the panel. There was no significant difference between the number of respondents that decided to stop from Group A (7) and Group B (6), indicating that there was no short-term effect on panel attrition because of the opt-out statement. The panel management system excluded some others (seven for Group A and twelve from Group B) before fielding the questionnaire for other reasons (e.g., because the respondent indicated they would be unavailable for a longer period due to holiday, illness, etc.). This yielded a final selection of 1,489 panel members who were asked to complete the follow-up survey (Group A: 731; Group B: 758).

The overall response rate to the follow-up survey was 73.2% (1,090 out of 1,489). The response rate was almost the same for both groups: 73.5% for Group A (537 out of 731) and 73.0% for Group B (553 out of 758). One respondent in Group A and one respondent in Group B had incomplete data.

4.3.1. Knowledge Questions

We first discuss the results of Part II (the knowledge questions), which was only presented to Group A. Almost 70% of the respondents said they had read the e-mail (or letter). There was no significant difference between the group who received the e-mail and the group who received the letter. That is, whether the opt-out statement was (said to be) read did not depend on mode.

We presented seven statements on the process of linking to the respondents. [Table 3](#) summarizes the responses.

For four of the seven statements a majority of the respondents gave the correct answer. However, for none of the statements was the percentage of respondents giving the correct

Table 3. Summary of the responses given to the seven statements concerning the record linkage. The correct answer for each statement is displayed in brackets

When linking your responses to our questionnaires to information that Statistics Netherlands has available about you . . .	Answer (in %)		
	Correct	Incorrect	Don't know
a) your name, gender, and date of birth will be sent to Statistics Netherlands. [TRUE]	34.7	44.2	21.1
b) researchers (from outside Statistics Netherlands) get access to your name, gender, and date of birth. [FALSE]	68.3	11.4	20.3
c) your name, gender, and date of birth will be saved with the linked data. [FALSE]	34.7	39.2	26.1
d) for each project the linked data will always stay at Statistics Netherlands, and will be destroyed after completion of the specific project. [TRUE]	39.9	16.4	43.7
e) results of the study can be traced to you as an individual. [FALSE]	65.5	9.7	24.8
f) every researcher can consult the linked data via the Internet. [FALSE]	66.8	7.5	25.7
g) the Dutch Data Protection Authority supervises the linking and analyses of the data. [TRUE]	65.5	5.6	28.9

answer very high. A majority seemed to understand that when linking survey data to administrative data researchers from outside Statistics Netherlands would not gain access to name, gender, and date of birth (68.3%), that results of the study cannot be traced to a specific individual (65.5%), that researchers cannot consult the linked data via the Internet (66.8%), and that the Dutch Data Protection Authority supervises the linking and analysis of the data (65.5%). However, only a minority knew that when linking information is sent to Statistics Netherlands, these variables are not saved with the linked data (34.7%), and that for each project the linked data will always stay at Statistics Netherlands, and will be destroyed after completion of the specific project (39.9%). We examined whether presenting the consent text in the survey only to the respondents who indicated they did not read the e-mail (or letter) affected the answer distributions. We did not find statistically significant associations, that is, the answer distributions did not differ between those who indicated they had read the letter or e-mail (and were not presented with the text again) and those who were presented with the text.

No respondent had all answers incorrect, and 3.5% of the respondents scored the maximum value (that is, all answers correct). When comparing the group of respondents who opted out with those who did not opt out, we found a striking result. Those who objected (opted out) had a significantly lower total score value (1.06 versus 2.51; $p < 0.001$). That is, they seem to have significantly *less* knowledge about the aspects of linking survey records to administrative data. Considering the responses to individual items for the group who opted out, the results not only show that the average frequency of correct answers to the individual items is much lower for the opt-out group (30.3% versus 55.2% for those who did not opt out), but also the average frequency of 'don't know' answers is much higher for the opt-out group (54.6% versus 25.4% for those who did not opt out).

When comparing the total score value of the panel members who received the extended text with those who received the short text, we found a significant difference. Panel members who received the extended text have a better knowledge about the linking process than those who received the short text (score values of 2.95 versus 1.88 respectively; $p < 0.001$). This seems to be a reassuring result: the more details one provides, the better respondents seem to understand what they are consenting to.

As mentioned in Subsection 3.4, we also calculated an alternative to the total score value: the measure of perceived risk of disclosure. We subtracted one point for a 'yes' answer to the statements d and g (see [Table 3](#)); for all the other statements a 'yes' answer increased the measure by one point.

Of the 536 respondents, 39 respondents (7.3%) scored -7 , or the lowest possible perceived risk; the highest value for perceived risk of disclosure was 6 (for only one respondent). There is (again) an obvious and significant difference between the group of respondents who opted out and the group who did not. The group of respondents who opted out have an average value of perceived risk of -0.94 , while this average for the group of respondents who did not opt out is -2.72 ($p < 0.001$). Those who objected to record linkage thus have a significantly higher perceived risk of disclosure.

Comparing the measure for perceived risk for those who received an extended text with those who received a short text, the perceived risk is significantly lower for the extended text (-3.20 versus -2.02 ; $p < 0.001$). So providing more detail is not only associated

with higher levels of knowledge, it is also associated with lower perceived risk of disclosure, regardless of the truth.

4.3.2. General Attitudes

The relative frequency distributions of the answers to the questions in Part I of the follow-up survey for both groups A and B are shown in Table 4. A chi-square test shows no significant difference between the answers given by Group A and B to the first question about personal privacy ($p = 0.44$). That is, there is no evidence that presenting the opt-out consent statement makes respondents more concerned about personal privacy in general. Table 4 also shows that the same holds true when asking questions about violations of privacy rights by banks, credit card companies, tax authorities, research on public opinion (either by government or in general), and computers (storing large amounts of information). The consent experiment did not appear to result in any changes in attitudes towards these issues. The experimental and control group also did not differ in concerns about continued confidentiality of information possessed by private and public organizations. We only found that the experimental group felt more strongly than the control group that different government agencies can get information about the respondent

Table 4. Relative frequency distributions (in %) of the answers to questions on concerns for both the experimental (A) and control (B) group

Questions on concerns	Experimental group (A)	Control group (B)
<i>Overall, how concerned are you about your (personal) privacy?</i>		
not at all concerned	13.2	14.5
not very concerned	45.6	49.2
a bit concerned	37.2	32.7
very concerned	3.9	3.6
<i>Please indicate whether you feel that your privacy is violated by the following entities (yes/no):</i> [frequencies of 'yes' answers are displayed]		
– banks and credit card companies, when they inquire after your financial situation	46.9	46.3
– the government, when you fill out your tax forms	14.5	14.1
– the government, when they conduct research projects among the population	24.8	25.3
– computers, which store a lot of information about you	80.4	80.1
– persons that ask questions as part of public opinion surveys	39.5	41.8
<i>Do you think that government entities can gather information about you if they try (yes/no)?</i> [frequency of 'yes' answers is displayed]	95.9	90.2
<i>All sorts of private and public organizations possess personal information about us. How concerned are you that this information will remain confidential?</i>		
not at all concerned	6.7	6.0
not very concerned	30.0	31.3
a bit concerned	55.5	52.4
very concerned	7.8	10.3

if they try to (95.9% versus 90.2%; $p < 0.001$). Overall the conclusion is that *concerns* about privacy hardly change after being exposed to the consent statement.

In terms of *trust* that information gathered about the respondent is treated confidentially, we found a few significant effects. Table 5 shows the frequency distributions of the answers to questions on trust for both the experimental and control group. Trust in research agencies that investigate public opinion and government agencies such as Statistics Netherlands to keep the information they collect from the respondent confidential turned out to be significantly different between those receiving the opt-out statement (Group A) and those not (Group B) ($p = 0.046$ and $p = 0.001$, respectively). Those who were exposed to the consent statement have *higher* levels of trust in these organizations.

5. Implementation and Longitudinal Effects

In September 2011 we informed all LISS panel members – except those we selected for the experiment – about the record linkage, and gave them the opportunity to opt out. Based on the results of the experiment and follow-up survey, we decided to send an extended text by e-mail to the balance of the LISS panel. The objective was not only to minimize opt-out rates but also to remove unfounded fears and to ensure that respondents understand what they are consenting to. The short letter had slightly lower opt-out rates than the extended e-mail in the experiment, but this difference was insignificant and the extended text increased knowledge and decreased perceived risk of disclosure, based on the results of the follow-up survey.

In total we sent 6,055 panel members the e-mail message; 551 (9.1%) opted out. This rate is significantly higher than that for the experiment (5.1% overall or 3.2% for the extended text by e-mail). A possible explanation is the fact that at the time the e-mail message was sent out, several cases of fraud related to data integrity were published in the

Table 5. Relative frequency distributions (in %) of the answers to questions on trust for both the experimental (A) and control (B) group

Questions on trust	Experimental group (A)	Control group (B)
<i>To what extent do you trust that information gathered about you is treated confidentially by:</i>		
– research agencies that investigate public opinion		
not at all	5.6	9.6
not very	27.4	29.7
a bit	50.7	46.6
a lot	16.2	14.1
– market research agencies		
not at all	14.6	19.6
not very	36.9	32.2
a bit	39.6	39.1
a lot	9.0	9.1
– government entities such as Statistics Netherlands		
not at all	3.7	4.9
not very	13.8	17.8
a bit	45.9	51.6
a lot	36.6	25.7

Netherlands and discussed extensively in the Dutch media. This may have triggered respondents to be more cautious in consenting to data linkage.

There is also the question of what the effects are of asking for consent in the long(er) run. Sending out a consent statement might affect the loyalty of panel members towards the fieldwork organization, and – if concern or distrust is stimulated by such a consent question – response rates may drop and panel attrition may increase. To examine the effects of sending out the opt-out consent statement, we compared the experimental group with all the other panel members who did not receive the consent statement (until September 2011). In the period February 2011–September 2011 the attrition rate on the individual level was equal to 8.3% for the panel members from the experimental group. For the remaining panel members the attrition rate in the same period was equal to 7.5%. The difference is not significant ($p = 0.43$).

In terms of response to the monthly questionnaires, we considered the following indicator. For the six-month period March–September 2011 we divided the number of months in which the panel member completed at least one of the questionnaires for which he was selected by the total number of months in which the panel member was selected for at least one questionnaire. This yields an indicator of participation. From the experimental group 92.8% were selected for questionnaires in all six months; for the non-experimental group this percentage was equal to 91.7%. The average percentage of participation is 72.2% for the panel members from the experimental group, compared to 70.4% for the panel members who were not included in the experimental group. Once again, the difference is not significant ($p = 0.22$).

The fact that the consent request could increase anxiety or distrust may be stronger for the group of panel members who opt out of linkage. To test for this we examined the group of respondents who did not participate in the experiment, but who were sent the opt-out statement in September 2011. We compared attrition and participation rates of the group of panel members who opted out with the group who did not opt out. We took data from the period September 2011–August 2012 (twelve months).

The attrition rate (on the individual level) was equal to 12.0% for the panel members who opted out (in response to the consent statement sent in September 2011), compared to 14.7% for those who did not opt out. This difference is not significant ($p = 0.06$). With respect to the participation rate (as defined above), among those who were active in the panel in the six months prior to September 2011, the opt-out respondents had an average percentage of participation of 84.7% in the twelve-month period after exposure to the statement, which is significantly higher than the 79.7% for those who did not opt out ($p < 0.001$). In other words, those who opted out in response to the presentation of the statement in September 2011 did *not* attrite at a higher rate and had *higher* participation rates in the following twelve months than those who did not opt out.

In summary, we found no evidence that sending an opt-out consent statement leads to a higher attrition or lower participation rates in future surveys in the panel.

6. Discussion

Results from our study show that sending a short e-mail message about record linkage between survey and administrative data yields the highest opt-out rate. A short e-mail

message may be considered as an attempt of the survey organization to get things done in a quick and easy way. It also provides the respondents with an easy way to opt out – simply by replying to the e-mail. A short letter or an extended e-mail yields lower opt-out rates. A letter appears to be more legitimate to respondents, and an extended text indicates the request is taken seriously by the survey organization.

The cognitive interviews as well as results from the experiment and follow-up survey indicate that respondents find it difficult to understand what linkage is all about. Whether we choose opt-in or opt-out consent statements, respondents seem to have little idea what is happening with the survey data they provide. The central question is whether it is possible at all to explain this to respondents and, if so, how best to do so. While researchers may be confident they can explain exactly what is happening in the case of record linkage, respondents in general population surveys may have little interest in these explanations. Results from the follow-up study, however, do indicate that the more details one provides the better respondents seem to understand what they are consenting to. These results run counter to the findings of [Singer et al. \(1992\)](#), which were based on a small convenience sample of university students and focused on survey participation. [Singer \(1978\)](#) found no effect of a short versus long statement on participation in an interviewer-administered survey. More research is needed on this topic. Our results also show that providing respondents with opt-out consent does not appear to increase concerns over privacy or trust in the survey organization.

While the proportion of LISS panel members who opted out of record linkage is relatively small (9.1%) compared to what might have occurred if explicit opt-in consent was required, it may still bias the results of projects using linked data. The rich amount of data collected earlier in the LISS panel offers the unique opportunity to get a clear picture of the respondents who opted out of record linkage, and the possible effects of their exclusion from key analyses. The LISS panel provides a further opportunity to explore the long-term effects of exposure to opt-out consent on later participation and attrition.

In summary, using an opt-out consent process in the LISS panel appears to have maximized the value of the linked data for researchers without apparent effects on panel members in terms of increased concerns over privacy or increased rates of attrition.

Appendix

Short and Extended Text Versions Used in the Experiment

SHORT TEXT VERSION (translated from Dutch)

Dear < XXX > ,

As member of the LISS panel, you are helping scientists gain valuable information. As you know, your privacy is guaranteed in all our research projects. Your answers to the questionnaires are only used for scientific research. We strictly comply with the Personal Data Protection Act, and never provide any information to other organizations like the Tax Administration or the UWV.

Some information is difficult or even impossible to acquire through your answers to our questions, for example because it would make the questionnaires extremely long or

complicated. Fortunately, Statistics Netherlands (SN) has additional information about you available; for example about your General Old Age and pension benefits, or data about your health and well-being.

For that reason we will soon start working with SN to combine information they have about you with your answers to our questionnaires. This will be done using highly protected computer systems. The results are always anonymous and cannot be traced to you in any way.

If you do not want us to combine your answers with data about you at SN, then please contact the CentERdata helpdesk on 0800–023 14 15 (free) or via lisspanel@uvt.nl.

With kind regards,

EXTENDED TEXT VERSION (translated from Dutch)

Dear < XXX > ,

As member of the LISS panel, you are helping scientists gain valuable information. As you know, your privacy is guaranteed in all our research projects. Your answers to the questionnaires are only used for scientific research. We strictly comply with the Personal Data Protection Act, and never provide any information to other organizations like the Tax Administration or the UWV.

Some information is difficult or even impossible to acquire through your answers to our questions, for example because it would make the questionnaires extremely long or complicated. Fortunately, Statistics Netherlands (SN) has additional information about you available; for example about your General Old Age and pension benefits, or data about your health and well-being.

How Will We Use the Additional Information?

Researchers using our data never have access to your name or address details. Our panel management does of course know which panel member number belongs to what person. SN also knows to whom the information that they have belongs. By comparing name, sex and date of birth, it is possible to combine the data of CentERdata and SN.

The information exchange with SN will be done using highly protected computer systems. After the data have been combined, your name, sex and date of birth will be removed from the database. Researchers that study the combined data will therefore never see your name, sex or date of birth. The results are strictly anonymous and cannot be traced to you in any way.

Who Will Have Access to the Combined Data?

Researchers can submit a request to SN to use the combined data for scientific research. If the request is accepted, the researcher is required to sign a contract with SN. The data will never leave SN. The researcher can access and use the SN data by means of a fingerprint reader and a secure connection. Research results will first be checked by SN before the researcher is permitted to publish them.

What Will CentERdata Do?

We will soon start working with SN to combine information they have about you with your answers to our questionnaires. If you do not want us to combine your answers with data about you at SN, then please contact the CentERdata helpdesk on 0800 – 023 14 15 (free) or via lisspanel@uvtnl.

With kind regards,

7. References

- Anderman, C., A. Cheadle, S. Curry, P. Diehr, L. Shultz, and E. Wagner. 1995. "Selection Bias Related to Parental Consent in School-Based Survey Research." *Evaluation Review* 19: 663–674. DOI: <http://dx.doi.org/10.1177/0193841X9501900604>.
- Bates, N.A. 2005. "Development and Testing of Informed Consent Questions to Link Survey Data with Administrative Records." In Proceedings of the American Statistical Association, May 12–15, 2005. 3786–3793, Miami Beach, FL. Available at: <http://www.amstat.org/committees/ethics/linksdire/Jsm2005Bates.pdf> (accessed May 21, 2014).
- Calderwood, L. and C. Lessof. 2009. "Enhancing Longitudinal Surveys by Linking to Administrative Data." In *Methodology of Longitudinal Surveys*, edited by P. Lynn, 55–72. Chichester: John Wiley & Sons.
- Couper, M.P. 2011. "The Future of Modes of Data Collection." *Public Opinion Quarterly* 75: 889–908. DOI: <http://dx.doi.org/10.1093/poq/nfr046>.
- Dahlhamer, J.M. and C.S. Cox. 2007. "Respondent Consent to Link Survey Data with Administrative Records: Results from a Split-Ballot Field Test with the 2007 National Health Interview Survey." In Proceedings of the Federal Committee on Statistical Methodology Research Conference, November 5–7, 2007, Arlington, VA. Available at: <http://www.fcsm.gov/07papers/Dahlhamer.IV-B.pdf> (accessed May 21, 2014).
- Eaton, D.K., R. Lowry, N.D. Brener, J.A. Grunbaum, and L. Kann. 2004. "Passive Versus Active Parental Permission in School-Based Survey Research." *Evaluation Review* 28: 564–577. DOI: <http://dx.doi.org/10.1177/0193841X04265651>.
- Ellikson, P.L. and J.A. Hawes. 1989. "Active vs. Passive Methods for Obtaining Parental Consent." *Evaluation Review* 13: 45–55. DOI: <http://dx.doi.org/10.1177/0193841X8901300104>.
- Fendrich, M. and T.P. Johnson. 2001. "Examining Prevalence Differences in Three National Surveys of Youth: Impact of Consent Procedures, Mode, and Editing Rules." *Journal of Drug Issues* 31: 615–642.
- Gray, M. 2010. "A Review of Data Linkage Procedures at NatCen", Working Paper. London: National Centre for Social Research. Available at: <http://www.natcen.ac.uk/media/205504/data-linkage-review.pdf> (accessed May 21, 2014).
- Hunt, K.J., N. Schlomo, and J. Addington-Hall. 2013. "Participant Recruitment in Sensitive Surveys: A Comparative Trial of 'Opt In' Versus 'Opt Out' Approaches." *BMC Medical Research Methodology* 13, DOI: <http://dx.doi.org/10.1186/1471-2288-13-3>.
- Junghans, C., G. Feder, H. Hemingway, A. Timmis, and M. Jones. 2005. "Recruiting Patients to Medical Research: Double Blind Randomised Trial of 'Opt-in' versus

- 'Opt-out' Strategies." *British Medical Journal* 331: 940. DOI: <http://dx.doi.org/10.1136/bmj.38583.625613.AE>.
- Pascale, J. 2011. "Requesting Consent to Link Survey Data to Administrative Records: Results from a Split-Ballot Experiment in the Survey of Health Insurance and Program Participation (SHIP)." *Study Series in Survey Methodology* #2011-03, (U.S. Census Bureau. Washington DC). Available at: <http://www.census.gov/srd/papers/pdf/ssm2011-03.pdf> (accessed April 2013).
- Pascale, J. and T.S. Mayer. 2004. "Exploring Confidentiality Issues Related to Dependent Interviewing: Preliminary Findings." *Journal of Official Statistics* 20: 357–377.
- Sakshaug, J., M.P. Couper, M.B. Ofstedal, and D. Weir. 2012. "Linking Survey and Administrative Records: Mechanisms of Consent." *Sociological Methods and Research* 41: 535–569. DOI: <http://dx.doi.org/10.1177/0049124112460381>.
- Sakshaug, J. and F. Kreuter. 2012. "Assessing the Magnitude of Non-Consent Biases in Linked Survey and Administrative Data." *Survey Research Methods* 6: 113–122.
- Sala, E., J. Burton, and G. Knies. 2012. "Correlates of Obtaining Informed Consent to Data Linkage: Respondents, Interview and Interviewer Characteristics." *Sociological Methods and Research* 41: 414–439. DOI: <http://dx.doi.org/10.1177/0049124112457330>.
- Sala, E., G. Knies, and J. Burton. 2013. "Propensity to Consent to Data Linkage: Experimental Evidence from the Innovation Panel on the Role of Three Survey Design Features." *Understanding Society Working Paper Series* 2013-05. Colchester, England: University of Essex, Institute for Social and Economic Research. Available at: <http://www.iser.essex.ac.uk/publications/working-papers/understanding-society/2013-05> (accessed June 29, 2014).
- Scherpenzeel, A.C. and M. Das. 2011. "'True' Longitudinal and Probability-Based Internet Panels: Evidence from the Netherlands." In *Social and Behavioral Research and the Internet*, edited by M. Das, P. Ester, and L. Kaczmirek, 77–104. New York: Taylor and Francis.
- Singer, E. 1978. "Informed Consent: Consequences for Response Rate and Response Quality in Social Surveys." *American Sociological Review* 43: 144–162.
- Singer, E. 2003. "Exploring the Meaning of Consent: Participation in Research and Beliefs about Risks and Benefits." *Journal of Official Statistics* 19: 273–285.
- Singer, E. and M.P. Couper. 2011. "Ethical Considerations in Internet Surveys." In *Social and Behavioral Research and the Internet*, edited by M. Das, P. Ester, and L. Kaczmirek, 133–162. New York: Taylor and Francis.
- Singer, E., H.-J. Hippler, and N. Schwarz. 1992. "Confidentiality Assurances in Surveys: Reassurance or Threat?" *International Journal of Public Opinion Research* 4: 256–268. DOI: <http://dx.doi.org/10.1093/ijpor/4.3.256>.
- Vellinga, A., M. Cormican, B. Hanahoe, K. Bennett, and A. Murphy. 2011. "Opt-Out as an Acceptable Method of Obtaining Consent in Medical Research: A Short Report." *BMC Medical Research Methodology* 11: 40. DOI: <http://dx.doi.org/10.1186/1471-2288-11-40>.

Received April 2013

Revised April 2014

Accepted April 2014