Understanding Society
Working Paper Series

No. 2022 – 05

June 2022

Consent to Data Linkage: Question Wording and Format Experiments

Annette Jäckle[1], Jonathan Burton[1], Mick P. Couper[2], Thomas F. Crossley[3] and Sandra Walzenbach[4]

[1]University of Essex (UK); [2]University of Michigan (USA); [3]European University Institute (Italy); [4]University of Konstanz (Germany)

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

UKRI Economic and Social Research Council

Non-technical summary

Survey data are increasingly being linked to administrative records to maximize the value of the data while minimizing respondent burden. But relatively little is known about how best to word and format such consent requests. In this paper, we describe a collection of experiments focused on the effect of wording and placement of the data linkage consent request on consent rates. Given the importance of consent being informed, we also focus on understanding of the linkage process and confidence in the consent decision. Further, we examine time taken to reach a decision, and whether respondents consulted additional materials describing the linkage process.

We experimentally varied: 1) the wording of the consent request (comparing an "easy" version with simplified text to a "standard" version), 2) placement of the consent request in the survey (early versus late), 3) consent as default (i.e., "Press 'next' to continue" or explicitly opt out) versus the standard opt-in consent question (yes/no), 4) offering additional information (i.e., adding a response option "I need more information before making a decision"), and 5) a trust priming experiment focusing on trust in the data holder.

The overarching goal is to increase the rates of consent to administrative data linkages in surveys without compromising understanding of the consent process and confidence in the decision made, that is, to increase the rates of *informed* consent. All experiments were designed with this goal in mind.

Our findings show that simplifying the wording of consent statement improves understanding of the linkage process. Placing the consent request early in the survey increases consent rates. Priming participants to consider trust in the agents involved increases consent rates without compromising understanding. Our research points to ways in which the informed consent process can be improved.

Consent to Data Linkage: Question Wording and Format Experiments

Annette Jäckle[1], Jonathan Burton[1], Mick P. Couper[2], Thomas F. Crossley[3] and Sandra Walzenbach[4]

[1]University of Essex (UK); [2]University of Michigan (USA); [3]European University Institute (Italy); [4]University of Konstanz (Germany)

Abstract: Survey data are increasingly being linked to administrative records to maximize the value of the data while minimizing respondent burden. But relatively little is known about how best to word and format such consent requests. We conducted a series of independent experiments in a longitudinal survey and an online access panel to understand how various features of the design of the consent request can affect the respondents' decision, their understanding of the linkage process, and their confidence in the decision made. Easy wording of the consent question increases objective understanding but does not increase the consent rate. Asking for consent early in the survey increases consent. Priming respondents to consider trust increases consent.

JEL classification: C81, C83

Keywords: informed consent; data linkage; wording experiments; consent understanding; trust

Acknowledgements: This research was funded by the Nuffield Foundation (www.nuffieldfoundation.org) with co-funding from the UK Economic and Social Research Council (ESRC, https://esrc.ukri.org/) (OSP/43279). Data collection on the Innovation Panel was funded by the ESRC grant for *Understanding Society*: The UK Household Longitudinal Study, Waves 9-11 (ES/N00812X/1). We are also grateful for the feedback and suggestions received from the project advisory board. The views expressed here are those of the authors and not necessarily of the Nuffield Foundation or the ESRC.

Corresponding author: Annette Jäckle, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom, email: aejack@essex.ac.uk; phone: +44 1206 873896.

1. Introduction

Survey data are increasingly being linked to administrative records to maximize the value of the data while minimizing respondent burden. But relatively little is known about how best to word and format such consent requests. In this paper, we describe a collection of experiments focused on the effect of wording and placement of the data linkage consent request on consent rates. Given the importance of consent being informed, we also focus on understanding of the linkage process and confidence in the consent decision. Further, we examine time taken to reach a decision, and whether respondents consulted additional materials describing the linkage process.

We experimentally varied: 1) the wording of the consent request (comparing an "easy" version with simplified text to a "standard" version), 2) placement of the consent request in the survey (early versus late), 3) consent as default (i.e., "Press 'next' to continue" or explicitly opt out) versus the standard opt-in consent question (yes/no), 4) offering additional information (i.e., adding a response option "I need more information before making a decision"), and 5) a trust priming experiment focusing on trust in the data holder.

The overarching goal is to increase the rates of consent to administrative data linkages in surveys without compromising understanding of the consent process and confidence in the decision made, that is, to increase the rates of *informed* consent. All experiments were designed with this goal in mind. Our expectation is that the experimental treatments will increase rates of consent while not having any negative effects on understanding of the consent process, or on respondent confidence in their decision (as another measure of feeling informed without feeling pressured to comply). We next review the literature relevant to our experiments before identifying research questions and offering specific hypotheses.

2. Background and review of relevant literature

The survey methods literature has relatively few examples of wording experiments, other than manipulations of the length and framing of the consent request. We therefore review both the survey methods literature on linkage consent and literature from related fields, especially clinical trials and psychological research.

Length

Length of the consent form for record linkage has been the focus of limited experimental research in surveys. Singer (1978) and Singer and Frankel (1982) found no effects of consent question length on consent in interviewer-administered surveys, but Bradford et al. (2021) found a significant positive effect of shorter wording on consent to an online survey among Facebook members. Das and Couper (2014) conducted a two-factor experiment examining the length of consent text (short versus extended) and mode of communication (letter versus email), in a study of opt-out consent. Their extended text resulted in lower opt-out rates, greater objective understanding of the consent request (measured with true/false knowledge test questions) and lower perceptions of risk associated with record linkage, than the shorter text. There was an interaction between the two conditions with the combination of a short message and email having the highest opt-out rates (9.6%), but otherwise the way in which respondents were informed did not have much effect (2.6%-4.9% opt out rate). Edwards and Biddle (2021) conducted an experiment in the online Life in

Australia (LINA) panel, with participants randomly allocated to a long or short form of the data linkage consent question for two different types of administrative records (income support receipt, pensions and benefits; and health records). The consent rates for the short form (28.4%) were slightly but not significantly higher than for the long form (26.6%). They also found no interaction between length and type of data linkage request. Objective understanding and perceptions of risk did not differ significantly by length of the consent request.

We acknowledge that it is hard to manipulate the length of the consent request without also affecting content and format. But aside from the framing experiments reviewed later, we know of no other survey studies that have varied length or content. To find relevant literature, we need to look to the informed consent literature from psychology and biomedical research.

Focusing first on length, a number of studies have varied the length of informed consent forms or patient information leaflets for clinical trials, or consent forms for online surveys. Brierley, Richardson, and Torgerson (2012) found no differences in recruitment rates by information leaflet length. Three other clinical trials focused on objective comprehension (Enama et al. 2012; Matsui et al. 2012; Stunkel et al. 2010). They found no differences in understanding by form length, with Stunkel et al. (2010) concluding that "volunteers had the same level of comprehension after reading a 14-page or a four-page consent form." However, an online survey quasi experiment (with students at different institutions exposed to different length forms; Perrault and Nazione 2016) and designed experiment (Varnhagen et al. 2005) both reported higher rates of recall of key information for those exposed to the shorter forms.

Readability

A wide variety of different enhancements to consent forms have been tested in biomedical studies, including both content (e.g., simplified sentences and words, use of active voice) and design enhancements (layout, font, use of tables, images, and/or bullet points) to convey key information. Many of these studies evaluated the enhancements using the Flesch Kincaid Grade Level scale (FKGL; see Kincaid, Fishburne, Rogers and Chissom 1975), available in Microsoft Word. As with the length manipulations, several studies found no significant effect of these enhancements on objective measures of comprehension (Antonacopoulos and Serin 2016; Coyne et al. 2003; Grady et al. 2017; Paris et al. 2010, 2015; Perrault and Keating 2018; Walters and Hamrell 2008). However, Kim and Kim (2015) found that levels of both objective and subjective understanding were higher for the enhanced consent form. Tait, Voepel-Lewis, Nair, Narisetty and Fagerlin (2013) also reported higher rates of gist (main point) and verbatim (actual) understanding of the risks and benefits for enhanced versions of consent forms. Coyne et al. (2003) found lower levels of consent anxiety and higher levels of satisfaction among those exposed to the enhanced form. Two of the studies also looked at reading time: Paris et al. (2010) and Perrault and Keating (2018) found no differences by form version. Again, few of these studies examined actual consent rates. Cockayne et al. (2017) and Hall, Sanson-Fisher, Lynagh, Threlfall and D'Este (2013) report no significant differences in recruitment rates between standard and enhanced invitation letters. Similarly, Coyne et al. (2003) found no differences in participation rates by consent form type. However, Paris et al. (2015) reported lower rates of enrolment among those receiving the enhanced consent form.

Framing

In the survey methods literature, most of the wording experiments on record linkage consent have focused on framing of the request. Pascale (2011) reported on a study that varied whether the request mentioned accuracy of the data, reduction of data collection costs, or time saving reasons. The first two are benefits to the agency, while the last was framed as a benefit to the respondent. None of these had any effect on consent rates. More recent experiments have varied the framing of the request, with gain-framing mentioning the benefits from consent and loss-framing emphasising that not linking will reduce the value of the respondent's survey data. These studies have yielded mixed results. Kreuter, Sakshaug and Tourangeau (2016) found that loss-framing increased consent compared to gain-framing. However, Sakshaug, Wolter and Kreuter (2015) found that gain-framing was more effective than loss-framing. Similarly, gain-framing (expressed as a time saving for the respondent) yielded higher consent rates than a neutral framing in one study (Sakshaug and Kreuter 2014) but not in another (Sakshaug, Tutz and Kreuter 2013). Sakshaug, Schmucker, Kreuter, Couper and Singer (2019) found that the effect of framing (gain vs. loss) was evident only in one mode (a web survey, but not in a telephone survey), where loss-framing yielded a higher consent rate than gain-framing, but only when the consent request came at the end of the survey. Finally, Welch et al. (2017) found no significant differences between loss-framing and control conditions in consent to contact children's vaccination providers in a telephone survey. They also found no effect of normalized wording (e.g., "Most people we interview give us permission…") on consent.

Placement or location of the request

Other studies have varied when the request for record linkage consent is made. Both Eisnecker and Kroh (2016) and Sala, Knies and Burton (2014) found no effect of asking for linkage in an earlier versus later wave of a longitudinal study. Within a survey wave, asking for consent after a module of questions related to the content of the data to be linked increased consent compared to asking at the end of the questionnaire (Sala, Knies and Burton 2014), and asking it at the beginning of the survey rather than the end had a positive effect (Eckman and Haas 2017; Sakshaug et al. 2019).

In sum, the literature on psychology and biomedical research has mainly focused on the effects of wording and formatting of consent materials on objective and subjective understanding of the request. In contrast, few survey experiments examine outcomes beyond the consent decision itself, with the exception of Das and Couper (2014) and Edwards and Biddle (2021). Both studies point to relatively poor comprehension of the data linkage process. No studies (to our knowledge) of administrative record linkage consent in surveys have examined subjective confidence in the consent decision. Few studies have examined the other features of the consent request that we test in this paper: the provision of additional information, default wording of the consent request, and trust priming.

Additional information

Experimentally varying the provision of additional information on request has not been the subject of much research. This is predicated on the assumption that participants do not seem to invest significant effort in reading and understanding consent forms (see, e.g., Ghandour, Yasmine and El-Kak 2013; McNutt et al. 2008; Ripley, Hance, Kerr, Brewer and Conlon 2018). We expect the offer of additional information, coupled with design and wording enhancements to shorten and/or simplify

the consent information provided, to serve two purposes. One, it may reduce the length of the initial information provided, potentially increasing the likelihood of it being read. Two, it may give the respondent greater agency or choice in what information they are exposed to, potentially thereby also increasing engagement in the material. This is consistent with Annas' (2017) notion of "informed choice," giving participants the choice to become more informed if they wanted to.

Perrault and McCullock (2019) report on a non-experimental study where they developed a short initial consent for an online study on sexual health among U.S. college students that allowed participants a choice to either continue directly to the study or learn more about the study. All participants (100%, n=429) decided to continue directly to the study, choosing to forgo additional information about the study and the institutional review board (IRB) approval process.

In another non-experimental study, Graves and colleagues (2019) conducted a regression tree analysis of factors associated with consent to data linkage at the end of an open-access online survey of a cohort of 18-23 year old Australian women. Provision of residential address or not (as a measure of attitudes towards privacy), was the most important factor in classifying the data into similar groups of consenters, followed by incentive type. Two types of incentives were offered: AU$50 plus basic information on linkage, or designer leggings plus additional data linkage information. In the latter group, if the respondent did not consent to data linkage additional information popped up giving her a chance to change her mind. This additional information included further reassurances that health records provided via data linkage are confidential, examples of the type of information that the data linkage would provide and a link to an infographic illustrating how data is linked anonymously using keys. The linkage consent rates were 61% versus 79% for the AU$50 group and the designer legging group respectively ($p<.01$).

Default wording

Our experiment on default wording is informed by the literature on default options in behavioural economics (see, e.g., Thaler and Sunstein 2008). While opt-in or explicit consent is the norm for most research, there are some exceptions based on regulations mostly applying to government statistical agencies. For example, the U.S. Census Bureau uses opt-out consent for most of its data linkage work, as does Statistics Canada, and the U.K. Office for National Statistics (see Bates 2017; Hewison and Haines 2006). Yang, Fricker and Eltinge (2019) provide one recent example of opt-out consent. Little research has explored opt-in (active or explicit) versus opt-out (passive) consent to record linkage in surveys.

There is a growing body of literature across a number of different behavioural domains suggesting that small "nudges," including default options, can increase participation in various activities (see, e.g., Dinner, Johnson, Goldstein and Liu 2011; Dranseika and Piasecki 2020; Hummel and Maedche 2019; Loewenstein, Bryce, Hagmann and Rajpal 2015; Reisch and Sunstein 2016). We know of no research on using a default option for record linkage consent in the survey literature, but expect the findings in other fields to apply here too. In the consent-as-default approach we tested, respondents could press "next" to continue with the survey (and thereby consent to linkage) or explicitly opt-out. We contrast this with the standard approach with an explicit yes/no consent question.

Priming trust

Our final experiment is a trust prime. Again, there is a dearth of relevant survey literature. Our choice of this experiment was based on analyses of correlates of consent, that suggested trust in the organizations involved was a key predictor, and the association remained statistically significant after controlling for sociodemographic characteristics (Jäckle et al. 2021a). Trust was also frequently mentioned in qualitative interviews as a reason for consenting to record linkage (see Jäckle, Beninger, Burton and Couper 2021; Thornby, Calderwood, Kotecha, Beninger and Gaia 2018). Similarly, in a question asking for how respondents reached the consent decision, the option "I thought about how much I trust the organisations involved" was the most frequently chosen response option in the samples in which this was offered (Burton, Couper, Crossley, Jäckle and Walzenbach 2021).

Priming experiments are common in the psychology literature (see, e.g., Bargh 2006; Bargh and Chartrand 2000; Molden 2014). Priming is also called "digital nudging" in the information systems literature (see Dennis, Yuan, Feng, Webb and Hsieh 2020), and is a close ally of default options and other nudging techniques. In the context of medical decisions, Sepucha and colleagues (2010) found that trust in the doctor was associated with feeling informed, suggesting a "rational delegation of consent". That is, if respondents trust the organisations involved, they might feel they do not need to know the details of the request in order to consent (see also Kasperbauer et al. 2022; Kongsholm and Kappel 2017).

Couper, Singer, Conrad and Groves (2010) tested a privacy prime in a vignette-based web experiment on the effect of disclosure risk on hypothetical survey participation. Privacy was made salient at the outset by asking respondents a few questions about privacy concerns, while the control group was asked questions about a neutral topic, computer use. They found a significant negative effect of the privacy prime relative to the neutral (computer use) prime. That is, those for whom privacy was made salient have significantly lower levels of willingness to participate in the survey than those given the neutral prime. They also found that the privacy prime had significant effects on subsequent measures of perceived risk and perceived harm, and on general attitudes toward privacy, suggesting that the privacy prime was effective in raising general privacy concerns. Our expectation is that a trust prime will have a similar salutary effect on consent to record linkage, again without compromising comprehension or confidence in the decision.

Response time

Finally, very few studies of consent have examined the time taken to read the consent form. In part this is because many consents are administered verbally or on paper (this is especially true of consents in clinical trials). McNutt et al. (2007) used observers to record the time taken by women in a primary care setting in the U.S. to read informed consent forms for two studies related to partner violence. They reported that "the majority of patients spent less than 30 seconds to read their consent form." No study subject spent even half as much time as expected, based on the length of the forms and average reading speeds. Matsui et al. (2012) used a self-report measure of reading in a study comparing standard and short consent forms for a genetic cohort study in Japan. Only half of both groups reported reading at least half the documents before the informed consent procedure in the study (47.0% standard form, 53.6% short form). About a third in each group (34.6% standard and 30.8% short) admitted reading none of the consent form. Differences by treatment were not

significant (p=.50). Perrault and Keating (2018) compared several different variations for consent to participate in an online survey among U.S. college students. In addition to using timing paradata from the online survey, they included a self-report measure of reading. Only 9.7% of participants indicated they read the consent form entirely, while 69% said they skimmed it. The authors report a significant positive correlation (r=.46) between time spent on the form (measured passively) and comprehension.

Two other studies measured reading speed using survey paradata. Desch et al. (2011) measured the time taken to read online consent documents to a minimal-risk genetic study among U.S. college students. The consent document for adults was 2,833 words long, which would take a minimum of 566 seconds to read at a reading speed of 300 words per minute. The observed median time to consent was 53 seconds, with 23% of participants consenting within 10 seconds, and 93.6% consenting "without spending sufficient time to thoroughly read and comprehend the informed consent document." Ghandour, Yasmine and El-Kak (2013) measured reading speed in an online survey on sexuality and sexual practices among college students in Lebanon. A median time of 18.7 seconds was taken to read the 815-word informed consent form, with 65% of participants consenting within 30 seconds and 90% in less than the minimum predicted time (2.7 minutes). They did not find differences in response time between those who consented to the study and those who declined.

These studies suggest that relatively few participants fully read the consent documents, whether measured passively using paradata or based on self-report. Given the limited evidence, it is important to investigate the relationship between response time (reading speed) and the outcomes of interest (consent and understanding of the request).

3. Hypotheses and Research Questions

With the above discussion in mind, we turn to a specification of research questions and hypotheses. Our primary focus is on each of the experiments in turn.

RQ1: Do variations in question wording affect rates of consent to administrative data linkage, objective understanding, subjective understanding, and confidence in the consent decision?

Overall, the primary aim of these experiments is to positively affect (increase) consent rates without negatively affecting understanding or confidence (i.e., without compromising being informed or feeling informed). We measure understanding both objectively (using knowledge test questions) and subjectively (using self-reports of how well the respondent feels they understand the request). That is, we have a directional hypothesis for the primary outcome and a non-inferiority hypothesis for the secondary outcomes. Given this, we posit two specific hypotheses:

> H1: For each of the experiments, the treatment groups will result in a higher rate of consent to record linkage than the control group.
>
> H2: For each of the experiments, objective understanding, subjective understanding, and confidence will be no lower in the treatment groups than in the control group.

RQ2: What effects do these experimental variations have on the time taken to respond to the consent question and whether respondents click on hyperlinks to consult additional materials (a leaflet and diagram describing the linkage process)?

We have no firm expectations regarding the direction of the experimental effects on response speed (time to answer the consent question) and whether respondents availed themselves of the opportunity to consult the additional materials. For example, presenting the consent request in an easier (albeit longer) format may make it easier (faster) to read, but may also encourage more respondents to read the entire description of the process (as opposed to just skimming it). Similarly, will the offer of additional information increase access (review) of the additional material? Consistent with the biomedical literature reviewed above, we expect that relatively few respondents will take up the offer to consult additional information, but we are interested in the effect of the experimental manipulations (especially the explicit offer of additional information) on the use of this information.

In addition to looking at the bivariate relationships of the experimental conditions on outcomes, we also examine multivariable models to test whether these findings hold after controlling for other covariates of consent. We also test for selected interactions, in the case where experimental manipulations were crossed. For these analyses we focus on the two primary outcomes: consent and objective understanding.

4. Methods

Here we briefly describe the overall design and implementation of the surveys used for these experiments. We provide further details of the experimental manipulations in the Results section, when describing each experiment in turn. The experiments were embedded in four surveys using two different studies: the *Understanding Society* Innovation Panel and the PopulusLive online access panel.

The Innovation Panel is a probability sample of households in Great Britain that is used for methodological testing and experimentation. It is part of *Understanding Society:* The UK Household Longitudinal Study and its design mirrors that of the main panel. We implemented our study in wave 11, which was fielded in May to October 2018 by Kantar Public and NatCen Social Research (University of Essex, Institute for Social and Economic Research 2021). We refer to this survey as IP11. IP11 was conducted with random assignment to sequential mixed-mode designs, with some panel members assigned to web-first (followed by face-to-face) and others assigned to face-to-face first (followed by web). The response rates were similar between the two mode treatment groups: 80.5% of households allocated to FTF-first (and 80.8% of individuals in those households) responded, compared to 77.6% of households allocated to web-first (and 83.2% of individuals in those households). For documentation of the IP survey design and implementation and the IP11 questionnaire see the online survey documentation.[1]

[1] The IP User Guide is available at https://www.understandingsociety.ac.uk/documentation/innovation-panel/user-guide. The IP11 questionnaire is available at https://www.understandingsociety.ac.uk/documentation/innovation-panel/questionnaires.

Since the IP11 sample size constrained the number of experimental treatment groups we could implement, we fielded parallel surveys with additional experiments in an access panel (Jäckle, Burton, Couper, and Crossley 2022). The PopulusLive access panel (AP) is a non-probability online panel in the UK with around 130,000 active sample members at that time, who are recruited through web advertising, word of mouth, and database partners. To enable some comparison with the Innovation Panel sample, the sample was restricted to Great Britain and quotas based on age, gender and education were set to match the characteristics of the IP11 sample. Two samples were selected in this way. The first was surveyed in May 2018 and a sub-set was surveyed again in May 2019. We refer to the surveys from this two-wave panel as AP1.1 and AP1.2. A total of 46,206 panellists were invited to the first survey, of whom 6,532 started the survey and 5,633 completed it (401 broke off and 498 were screened out), for a participation rate of 12.2% (see AAPOR 2016). A sub-set of 2,053 respondents who completed AP1.1 were invited to the second wave (AP1.2); of these, 1,630 respondents completed the second survey for a conditional response rate of 79.4%. A second access panel (AP2) sample was independently selected in December 2019 and surveyed only once. A total of 30,682 panellists were invited to the survey, of whom 6,459 started the survey and 3,850 completed it (301 broke off and 2,308 were screened out), for a participation rate of 12.5%. Each of the AP samples included other experiments not reported on here, so the samples sizes reported in Table 1 below are a subset of cases. The implementation of these surveys was led by NatCen Social Research, in collaboration with the PopulusLive panel.[2]

Table 1 summarizes the various experiments and surveys. Further details on the surveys and experimental manipulations can be found in Jäckle (2021a). The final experiment on trust also varied the linkage domain, crossing the trust prime with a linkage request to income tax data (Her Majesty's Revenue and Customs, HMRC) or health data (held by the National Health Service, NHS). The AP surveys also included experiments with asking for multiple consents to data linkage (reported in Walzenbach, Burton, Couper, Crossley and Jäckle 2022). Sample members allocated to the multiple consent treatment groups are excluded from the analyses in this paper and from the sample sizes reported in Table 1. The IP11 analysis sample excludes respondents who completed the survey by telephone (n=1), who did not answer the consent question (n=1), and face-to-face respondents who did not complete the self-completion module of the questionnaire (n=233).

In terms of the outcomes of interest, we measured objective understanding using a series of eight true/false questions about the data linkage process[3]. Subjective understanding was measured with a single item asking *"How well do you think you understand what would happen with your data …,"* with a 4-point response scale ranging from *"I do not understand at all"* to *"I understand completely."* Confidence was similarly measured with a four-point response scale ranging from *"Very confident in my decision"* to *"Not confident in my decision"*. We reversed the scoring so a high score means greater confidence. These questions were all included in the self-completion module of the questionnaire. We also captured paradata, including time stamps (to measure how long a

[2] The AP User Guide by Jäckle, Burton, Couper, Crossley and Walzenbach (2022) and the AP questionnaires are available from the UK Data Service at https://reshare.ukdataservice.ac.uk/855464.

[3] A small number of cases (n=18 in IP11-FTF, n=142 in IP11-Web, n=5 in AP1.1, n=0 in AP1.2, and n=3 in AP2) did not answer any of the 8 items. We dropped these cases and coded the remaining missing responses as incorrect. Basing the percent correct on the answered questions or doing a complete case analysis (dropping cases missing on any of the 8 items) did not change the substantive conclusions presented here.

respondent spent on a given question) and, for online respondents, whether they clicked on links to additional information about the data linkage.

Table 1: Summary of experiments by survey

No.	Experiment	IP11-FTF	IP11-Web	AP1.1	AP1.2	AP2
1	Easy vs. standard wording of consent question	x	x	x	x	
2	Early vs. late placement of consent in questionnaire	x				
3	Additional information question wording			x		
4	Consent as default vs. standard yes/no question			x		
5	Trust priming					x
	Sample sizes	1,363	1,298	2,563	817	1,921

Note that in the AP samples, we did not actually link respondent data to administrative records. At the end of the surveys, respondent were informed that *"… we will not actually link the answers from this survey to any other data sources: the purpose of this survey is to collect information about the attitudes and concerns of the general public about data sharing."* In the case of the IP, consenting respondents will have their data linked with the relevant administrative records.

The analyses were conducted in Stata 15. We did not account for the complex sample design in the IP as our focus is on inference to the experimental conditions, not the larger population, and to parallel the AP analyses which are from a non-probability sample. Most of the analyses are bivariate associations, using chi-square tests of equivalence of proportions and F-tests of equivalence of means from ANOVA. Response time is positively skewed, so we present both means and medians. We test for the equivalence of medians between two groups using a two-sample Wilcoxon rank-sum test (yielding a z-score), and between three groups using a Kruskal-Wallis test of the Wilcoxon rank-sum scores (yielding a chi-square test).

5. Results

Easy versus standard wording

This experiment was designed to manipulate the difficulty of the request. Half of the sample was allocated to the 'standard' question wording, which had been used previously in the main *Understanding Society* survey. The other half were allocated to an 'easy' version, where the text was rewritten to reduce reading difficulty and to provide all essential information about the linkage in the question text rather than an additional information leaflet. The revisions were based on findings from our prior qualitative in-depth interviews about wording that hampered respondents' understanding of the consent request (Beninger, Digby, Dillon and MacGregor 2017) and on criteria used for reading level statistics. We assessed reading difficulty using the Flesch-Kincaid Grade level scores implemented in Microsoft Word. The revisions reduced the Flesch-Kincaid Grade level (which rates reading difficulty on a U.S. school grade level) from 14.3 to 8.8, meaning that an ninth grader (aged 14-15) would be able to understand the easy question wording. Both versions offered a leaflet

and diagram, but the diagram was enhanced for the easy version. The wording of the response options was also changed, from *"1) I have read the leaflet and am happy to give consent, 2) I do not want to give consent"* to simple *"1) Yes, 2) No"* responses. Note that the easy version was actually longer than the standard version (206 versus 128 words including response options). At a reading speed of 250 words per minute (wpm), it should take 49.4 seconds and 30.7 seconds respectively to read; at a reading speed of 300 wpm it should take 41.2 seconds for the easy version and 25.6 seconds for the standard version (see Brysbaert 2019 for a recent meta-analysis of reading speeds).

This experiment was implemented in IP11 and AP1.1. In IP11 the allocation to question wording was crossed with the mode in which respondents completed the survey. Tables 2 and 3 present the results for face-to-face and web respondents in the Innovation Panel respectively, while Table 4 presents the results for web respondents in the access panel (AP1).

Table 2. Standard (control) versus easy wording (IP11 face-to-face interviews)

Outcome	Standard wording	Easy wording	Test of difference
Consent rate (%)	71.1%	73.6%	$\chi2(1)= 1.05$, p=0.304
Objective understanding (0-8) (mean, SD)	4.57 (1.54)	5.08 (1.55)	F(1, 1344)= 36.32, p<0.001
Subjective understanding (1-4) (mean, SD)	2.78 (0.96)	2.94 (0.93)	F(1, 1362)= 9.86, p=0.002
Confidence (1-4) (mean, SD)	2.94 (0.82)	3.02 (0.80)	F(1, 1362)= 3.15, p= 0.076
Response time (median, mean, SD)	69 81.36 (56.47)	91.00 99.46 (61.15)	z = 6.205, p<0.001 F(1, 1361)= 32.27, p<.001
Consulted additional materials (%)	46.0%	47.0%	$\chi2(1)= 0.138$, p=0.711
(n)	(705)	(658)	

Table 3. Standard (control) versus easy wording (IP11 web respondents)

Outcome	Standard wording	Easy wording	Test of difference
Consent rate (%)	39.3%	44.6%	$\chi2(1)= 3.81$, p= 0.051
Objective understanding (0-8) (mean, SD)	4.4 (1.62)	4.8 (1.58)	F(1, 1155)= 19.25, p<0.001
Subjective understanding (1-4) (mean, SD)	2.37 (0.98)	2.42 (1.00)	F(1, 1297)= 0.72, p=0.397
Confidence (1-4) (mean, SD)	2.91 (0.83)	2.99 (0.82)	F(1, 1297)= 3.47, p= 0.063
Response time (median, mean, SD)	26.00 47.35 (59.31)	34.00 45.68 (38.48)	z = 4.691, p<0.001 F(1, 1297)= 0.36, p=0.547
Consulted additional materials (%)	17.5%	2.6%	$\chi2(1)= 78.34$, p<0.001
(n)	(657)	(641)	

It can be seen from Tables 2 and 3 is that while the easy wording has slightly higher consent rates, the differences do not reach significance at the 5%-level in either mode. Objective understanding is significantly higher with the easy version in both modes, while subjective understanding is significantly higher only among face-to-face cases. Response times are significantly longer for the easy version, reflecting the increased length of that version. The fact that fewer respondents consulted the additional material in the easy web version also reflects that some of this material was made more readily accessible in the question text, reducing the need to click on the links.

We tested whether the effects of the question wording experiment on consent, subjective and objective understanding, and confidence in the consent decision were different in web and face-to-face interviews. We found no significant interactions between easy or standard question wording and the mode of interview (reported in Jäckle et al. 2021b).

Table 4. Standard (control) versus easy wording (AP1.1)

Outcome	Standard wording	Easy wording	Test of difference
Consent rate (%)	46.4%	49.0%	$\chi^2(1)=0.71$, p=0.400
Objective understanding (0-8) (mean, SD)	4.29 (1.49)	4.68 (1.38)	F(1, 1023)=18.4, p<0.001
Subjective understanding (1-4) (mean, SD)	2.30 (0.90)	2.30 (0.94)	F(1, 1026)<0.01, p=.950
Confidence (1-4) (mean, SD)	2.94 (0.88)	2.94 (0.90)	F(1, 1026)=0.01, p=.940
Response time (seconds) (median, mean, SD)	41.27 65.39 (68.80)	38.03 61.81 (65.65)	z = 0.707, p=0.240 F(1, 1025)= 0.75, p=0.40
Consulted additional materials (%)	48.3%	27.6%	$\chi^2(1)=46.78$, p<0.001
(n)	(513)	(514)	

In AP1.1 we again find no main effect of the wording manipulation on consent rates. However, objective understanding of the consent process is significantly higher for those exposed to the easy wording, with 53.6% in the easy wording group getting more than half the items correct, compared with 43.4% in the standard wording group. Subjective understanding and confidence do not differ by treatment. In contrast to the IP, the median response time for the easy version is lower than that for the standard version.

We again see fewer respondents clicking on links to view the additional materials in the easy wording version. This again suggests that the easy wording version – by providing the information in a more accessible format – reduces the need to consult additional materials. It is interesting to note that both response times and the percent clicking links is higher in the access panel than in the IP. This may reflect the fact that requests for data linkage are fairly common in the IP, and panellists have developed trust in the survey organisations over time. In contrast, such requests for data linkage are rare in the access panel, potentially requiring greater scrutiny by participants.

One possible reason we do not find a significant effect of the alternative wording on consent rates is that the two versions were not sufficiently different. As a manipulation check, we asked a debriefing

question of respondents in AP1.1: "How easy or difficult was it for you to understand the question asking for permission to add data held by HMRC to the answers you gave in this survey?" We find no differences in mean scores (p=0.29). In general, respondents reported both versions to be very or somewhat easy to understand (83.3% in the easy version and 82.5% in the standard version, p=0.73). Nonetheless, the easier version did lead to better objective understanding of the linkage consent request.

We repeated the easy versus standard wording experiment in AP1.2, with respondents completing the survey a second time one year later, keeping the same allocation to treatment groups as in AP1.1. We obtained similar results to AP1.1: slightly, but not significantly (p=0.43) higher consent rates (54.1% for easy, 51.4% for standard); significantly higher levels of objective understanding (4.70 for easy, 4.32 for standard; p=0.0002); and no differences in subjective understanding or confidence. We find a similarly large difference in the percentage consulting additional materials (51.9% for the easy wording, 25.6% for the standard wording.

In summary, while the easy wording version is associated with significantly higher levels of objective understanding in both surveys and modes, this does not translate into significantly higher rates of consent.

Early versus late placement of consent in questionnaire

This experiment was included on the Innovation Panel only, and implemented only for face-to-face respondents. Half of the sample was asked for consent early in the interview (after a series of socio-demographic questions); the other half was asked at the end of the survey (before the self-completion CASI module), which is the usual placement of consent questions in *Understanding Society*.

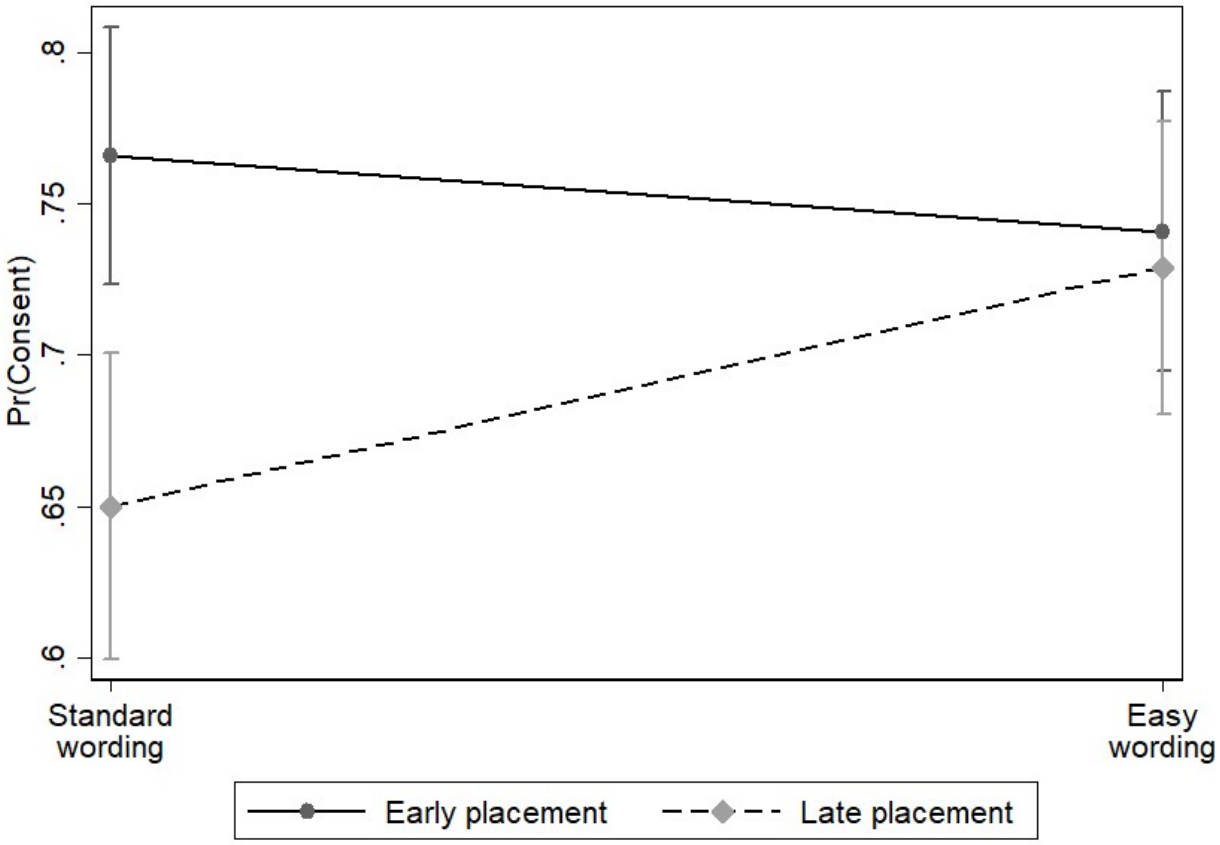
Table 5. Late versus early placement (IP11 face-to-face interviews)

Outcome	Late placement (control)	Early placement	Test of difference
Consent rate (%)	68.9%	75.4%	$\chi^2(1)=7.04$, p= 0.008
Objective understanding (0-8) (mean, SD)	4.87 (1.57)	4.77 (1.56)	F(1, 1344)= 1.48, p= 0.220
Subjective understanding (1-4) (mean, SD)	2.85 (0.95)	2.86 (0.94)	F(1, 1362)= 0.05, p=0.817
Confidence (1-4) (mean, SD)	2.98 (0.81)	2.98 (0.81)	F(1, 1362)= 0.00, p=0.967
Response time (seconds) (median, mean, SD)	69.00 81.12 (61.36)	90.00 98.34 (56.44)	z = 6.761, p<0.001 F(1, 1361)=29.15, p<0.001
Consulted additional materials (%)	45.33%	47.46%	$\chi^2(1)=0.624$, p= 0.430
(n)	(653)	(710)	

Here we find – consistent with prior literature – that asking consent for record linkage early in the interview is associated with a significantly 6.5 percentage point higher rate of consent. None of the other outcomes are affected, suggesting that the processing of the request does not change by position in the interview.

Given the wording and placement experiments were fully crossed for IP face-to-face respondents, we tested the interaction of these two manipulations. This interaction is marginally significant (Chi^2=4.6 d.f.=1, p=0.032 when controlling for covariates (see Appendix Table B.1 for details). Figure 1 illustrates the interaction, showing average marginal effects and 95% confidence intervals from the multivariable model. Figure 1 suggests that when the standard wording is used, early placement of the request has a significantly higher consent rate than late placement, but this effect is not seen when the easy wording is used. Put another way, the standard wording suppresses consent relative to the easy wording when the request is presented later in the survey. One interpretation is that the effect of more difficult wording is exacerbated when respondents are more fatigued.

Figure 1: Interaction effect of consent question wording and placement on probability of consent



Additional information question wording

All three versions of this experiment used the 'easy' wording from Experiment 1. In the control condition, the consent question simply had a "Yes/No" response. We tested two alternative versions, both of which had the following response options: "1) Yes, 2) I need more information before making a decision, 3) No." In one group, those who said "I need more information" were presented with a follow-up screen with additional information and were again asked for consent, as follows: "For more information on the data linkage, please read this **leaflet** and look at this **diagram**. Do you give permission for us to pass your name, address, sex and date of birth to HMRC to link your data? (Yes/No)." We refer to this as the "Need more information with follow-up" condition. The other group did not get the follow-up question, and we refer to this as "Need more information without follow-up". Results are presented in Table 6.

Table 6. Easy wording (control) versus offer of additional information with or without follow-up (AP1.1)

Outcome	(1) Easy wording (control)	(2) More information with follow-up	(3) More information without follow-up	Tests of differences
Consent rate (%)	49.0%	48.0%	39.9%	$\chi^2(2)=10.3$, p=0.006
Objective understanding (0-8)	4.68 (1.37)	4.61 (1.46)	4.60 (1.51)	F(2, 1533)=0.45, p=0.64
Subjective understanding (1-4)	2.30 (0.94)	2.14 (0.93)	2.18 (0.92)	F(2, 1538)=4.38, p=0.013
Confidence (1-4)	2.94 (0.90)	2.94 (0.88)	2.89 (0.89)	F(2, 1538)=0.65, p=0.520
Response time (seconds) (median, mean, SD)	38.03 61.81 (65.65)	36.67 50.21 (45.86)	41.98 61.27 (80.74)	F(2, 1538)=2.18, p=0.11 F(2, 1538)=8.17, p<0.01
Consulted additional materials (%)	27.6%	11.0%	22.5%	$\chi^2(2)=46.0$, p<0.001
(n)	(514)	(511)	(516)	

From Table 6 we see that offering additional information does not have a positive effect on consent rates. In the latter two groups, 19.8% and 17.8% respectively requested additional information. In the group that was subsequently followed up (group 2), 39.3% initially consented, while an additional 8.6% consented after providing additional information. In other words, almost a fifth of respondents expressed a desire for additional information, but offering that option does not increase consent rates. In both the conditions where additional information was offered, subjective understanding appears to be lower than in the control condition (p=0.004 for group 2 vs. 1, and p=0.310 for group 3 vs. 1).

The response time story is a complex one, and is linked to the consultation of additional materials. In group 2 (where the links to the leaflet and diagram are only presented to those who indicated a need for more information), response time is shorter than for the other two groups. But response times are substantially longer for those who click on the links (median=109 seconds) than those who do not (median=32 seconds), suggesting the lower proportion clicking on the link in group 2 accounts for the response time difference. Overall, we can conclude that giving respondents an opportunity to obtain additional information does not increase use of the links or response time substantially. Coupled with the findings that such offers do not increase consent rates or objective understanding, we conclude that these experimental treatments were not effective.

Default question wording

This experiment was implemented in AP1.1. This experiment used the 'easy' request for consent to link to tax data described above. In the control condition, respondents were explicitly asked "Do you give permission for us to pass your name, address, sex and date of birth to HMRC for this purpose? (Yes/No)" In the default version we dropped the yes/no question at the end of the question text, and instead presented linkage as the default, unless the respondent explicitly opted out. The respondent was instructed to "Press 'next' to continue" and given the option to click "I do not want HMRC records to be added to my answers to this survey". Table 7 presents the results.

Table 7. Easy wording (control) versus consent as default wording (AP1.1)

Outcome	Easy wording (control)	Consent as default wording	Test of difference
Consent rate (%)	49.0%	47.7%	$\chi^2(1)=0.289$, p=0.591
Objective understanding (0-8)	4.68 (1.38)	4.69 (1.50)	F(1, 1018)=0.04, p=0.839
Subjective understanding (1-4)	2.30 (0.94)	2.24 (0.97)	F(1, 1021)=0.94, p=0.33
Confidence (1-4)	2.94 (0.90)	2.91 (0.91)	F(1, 1021)=0.32, p=0.57
Response time (seconds) (median, mean, SD)	38.03 61.81 (65.365)	43.43 59.38 (53.47)	z=1.79, p=0.073 F(1, 1021)=0.42, p=0.52
Consulted additional materials (%)	27.6%	25.7%	$\chi^2(1)=0.47$, p=0.49
(n)	(514)	(509)	

Wording the consent response as the default surprisingly decreases the consent rate. However, the effect is modest and non-significant (p=0.591). Similarly, this experimental manipulation has no discernible effect on objective or subjective understanding, confidence in the decision, or whether the respondent clicked on the links for further information. The default wording approach took slightly (but not significantly) longer, possibly resulting from the less-straightforward question wording. Again, we conclude that our experimental manipulation was not particularly effective.

Trust prime

This experiment was included in AP2 and crossed with data linkage domain or data holder (HMRC versus NHS). We added an introductory screen saying, "The next question is about linking the information you provide in this survey, to data that [HMRC / the NHS] hold about you." Those in the trust prime group saw an additional statement: "[HMRC / The NHS] is a trusted data holder". This was followed by an icon symbolizing data security: a shield and lock symbol with the heading "Trust" (see the question wording in Appendix A). The control group did not see the additional statement or icon. We present the main effects (combining HMRC and NHS) for the trust manipulation in Table 8. We expected lower consent rates for HMRC than NHS but did not hypothesize an interaction of trust with data holder.

Table 8. No prime (control) versus trust prime (AP2)

Outcome	No prime	Trust prime	Test of difference
Consent rate (%)	51.0%	55.7%	$\chi^2(1)=4.34$, p=0.037
Objective understanding (0-8)	4.67 (1.45)	4.76 (1.42)	F(1, 1914)=1.64, p=0.20
Subjective understanding (1-4)	2.27 (0.98)	2.32 (0.95)	F(1, 1919)=1.30, p=0.25
Confidence (1-4)	2.92 (0.92)	2.95 (0.88)	F(1, 1919)=0.37, p=0.55
Response time (seconds) (median, mean, SD)	29.09 47.96 (56.67)	29.57 49.87 (71.32)	z=1.12, p=0.26 F(1, 1919)=0.42, p=0.52
Consulted additional materials (%)	25.1%	24.4%	$\chi^2(1)=0.13$, p=0.72
(n)	(961)	(960)	

As shown in Table 8, we find that the trust prime is associated with a significantly (p=0.037) higher level of consent than the no prime group, as hypothesized. As expected, we also find a main effect of data holder on consent (not shown in Table 8): 49.0% for HMRC, 57.8% for NHS ($\chi^2(1)=14.7$, p<0.001). Testing the interaction in a logistic regression model controlling for covariates (see Table B.2), we find main effects for both data holder (p<.0001) and trust (p=0.024), but no significant interaction (p=0.97) between the two on consent. In sum, this relatively modest manipulation has a significant positive effect on the proportion consenting, while not affecting any of the other outcomes of interest.

Multivariable models and interactions

Finally, we examine the effect of the experimental treatments on the two key outcomes (consent and objective understanding) in a multivariable context, controlling for socio-demographic variables. We briefly discuss these models here, and include the models in Appendix B, as the story described above does not change substantially in the presence of multivariate controls.

As previous studies have found, we find no socio-demographic characteristics that consistently predict consent in all samples. Age is associated with consent in the IP FTF and IP web samples, with older people having lower consent rates (see Table B.1). However, the effect of age is not as clear in

the AP samples (see Table B.2). Similarly, education is associated with consent in a non-linear pattern in IP FTF but not IP web or any of the AP samples.

In terms of objective knowledge, higher education is positively associated with knowledge in both IP samples and in all the AP surveys. Given this, we tested the interaction of education with the question wording manipulating, with the expectation that the easy wording may help more with those with less formal education. We find no significant interactions in any of the models tested.

6. Discussion and Conclusions

We conducted a series of experiments in which the wording, placement and format of the request for consent to the linkage of survey data with administrative records were manipulated. The aim of these experiments was to find ways to increase informed consent. Increased rates of consent, holding subjects' comprehension and confidence constant, are desirable. Equally, though, increased comprehension and confidence, for a given consent rate, is also a positive outcome. Of course, concomitant increases in consent, comprehension and decision confidence would be the best outcome of all.

We found small positive, but non-significant, effects on consent of the easier (but slightly longer) question wording, and this replicated across three samples (IP-FTF, IP-web, and AP). This finding is consistent with prior studies finding modest effects of wording manipulations on rates of consent. The easier wording did significantly increase objective knowledge across all three samples, increasing correct answers by about half a point on average. Subjective understanding (self-reported but in the presence of an interviewer) was significantly higher in the easy version in the face-to-face sample, but not in the two web samples. Confidence in the decision did not differ by question version.

We also found that significantly fewer respondents in the web samples clicked on the link for additional information in the easy wording condition than in the standard condition. In part we think this is because some of this material was incorporated into the question itself, reducing the need for additional explanation. In the face-to-face mode, the interviewer hands the respondent the leaflet in the same way regardless of wording, so we would expect no difference by wording there.

Consistent with prior research, we found that asking consent early in the interview was associated with a higher rate of consent. From the multivariable analyses we found a signification interaction between placement and wording. It appears that respondents are less likely to consent to data linkage if the question is asked late in the questionnaire and the wording of the request is difficult. Position has no effect on consent if the wording is easy; wording has no effect on consent if the position is early. But we did not test this in the web samples. Further research is needed to establish how robust this interaction is, and why the wording matters more when the question is asked late.

Turning to the experiments in the access panel, our attempts to provide the respondent with some agency or control over how much information they receive were not successful. Similarly, "nudging" the respondent to consent by framing it as the default option did not work. It is possible that access panel participants are not paying careful attention to the details of the request. However, 1) the median response times suggests that most respondents are reading much of the material presented, 2) the objective knowledge scores are similar to those for the Innovation Panel respondents, and 3)

the rates of consulting the additional materials are higher than that for IP-Web respondents. These observations suggest that lack of effort (or satisficing) may not be the explanation for the lack of effect of these manipulations. It may well be that our experimental manipulations were simply not strong enough to produce differences in consent rates. Alternatively, it may be that respondents are making decisions based on the key elements of the request, rather than on the peripheral factors that we were able to manipulate while keeping within the permissible boundaries for making such requests. This would be consistent with our findings in a related article (Burton et al. 2021), where we found that most respondents base their consent decisions on heuristics. Further work is needed to investigate how these heuristics can be influenced to increase informed consent.

Finally, we found that priming respondents to think about how much they trust the organisations involved in the data linkage increases consent, without negatively affecting comprehension or confidence in the decision. We tested this in only a single access panel sample, and so this manipulation this deserves further exploration. In a related article examining the effect of the survey mode on consent outcomes (Jäckle et al. 2021b), we found that answering the linkage consent question online, rather than with a face-to-face interviewer, led respondents to report higher levels of concern about data security and privacy. Whether trust priming would have a similar effect in interviewer administer surveys as it did in the present web survey remains to be tested.

Our research adds to the literature on consent to data linkage by expanding the outcomes studied beyond the rate of consent. Our work has shown that it is important to focus on other outcomes, including objective knowledge of the consent process (i.e., is the consent truly informed?), as well as subjective knowledge and confidence in the decision (do respondents feel informed, and are they comfortable with their decision?). For example, even though improvements in wording do not reliably and consistently increase consent rates, the fact that they significantly increase objective knowledge of the consent process is a positive finding.

In sum, our experimental manipulation of the consent request wording increased comprehension, with no decrease in consent. We replicated a prior finding that early placement of the request increases consent, and, in a novel finding, demonstrated that priming respondents to think about how much they trust the organisations involved. Importantly, both manipulations that increased consent rates did so without any negative effect on comprehension or confidence: they increased informed consent.

While the effects of these experimental manipulations are not large, they do produce incremental gains in consent rate or comprehension at negligible cost. Further work in this area is promising. As surveys increasingly seek consent to administrative data linkages to expand both the depth and breadth of data available for researchers, finding the best ways to inform and reassure participants about how such data linkages may be used remains a key research objective.

References

Annas, G.J. (2017). Informed consent: charade or choice? *The Journal of Law, Medicine & Ethics*, 45(1), 10-11, https://doi.org/10.1177/1073110517703096.

Antonacopoulos, N.M.D., & Serin, R.C. (2016). Comprehension of online informed consents: can it be improved? *Ethics & Behavior*, 26(3), 177-193, https://doi.org/10.1080/10508422.2014.1000458.

Bargh, J.A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology*, 36(2), 147-168, https://doi.org/10.1002/ejsp.336.

Bargh, J.A., & Chartrand, T.L. (2000). Studying the mind in the middle: a practical guide to priming and automaticity research. In H. Reis and C. Judd (eds.), *Handbook of Research Methods in Social Psychology*. New York: Cambridge University, pp. 253-285.

Bates, N. (2017). Consent and confidentiality for linked data. Paper presented at the CNSTAT Workshop on Developing a Methodological Research Program for Longitudinal Studies, Washington, DC, June 5-6.

Beninger, K., Digby, A., Dillon, G., & MacGregor, J. (2017). How people decide whether to give consent to link their administrative and survey data. Understanding Society Working Paper 2017-13. Colchester: University of Essex. Available at https://www.understandingsociety.ac.uk/research/publications/524692.

Bradford, A., Chiu, S., Kreuter, F., & Presser, S. (2021). Informed consent in social media surveys. Paper presented at the virtual annual conference of the American Association for Public Opinion Research.

Brierley, G., Richardson, R., & Torgerson, D.J. (2012). Using short information leaflets as recruitment tools did not improve recruitment: a randomized controlled trial. *Journal of Clinical Epidemiology*, 65(2), 147-54, https://doi.org/10.1016/j.jclinepi.2011.06.005.

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. https://doi.org/10.31234/osf.io/xynwg.

Burton, J., Couper, M.P., Crossley, T.F., Jäckle, A., & Walzenbach, S. (2021). How do survey respondents decide whether to consent to data linkage? Institute for Social and Economic Research, University of Essex: Understanding Society Working Paper Series No. 2021-05, https://www.iser.essex.ac.uk/research/publications/working-papers/understanding-society/2021-05.

Cockayne, S., Fairhurst, C., Adamson, J., et al. (2017). An optimised patient information sheet did not significantly increase recruitment or retention in a falls prevention study: an embedded randomised recruitment trial. *Trials*, 18, 144, https://doi.org/10.1186/s13063-017-1797-7.

Couper, M.P., Singer, E., Conrad, F.G., & Groves, R.M. (2010). Experimental studies of disclosure risk, disclosure harm, incentives, and survey participation. *Journal of Official Statistics*, 26(2), 287-300, https://www.scb.se/dokumentation/statistiska-metoder/JOS-archive/.

Coyne, C.A., Xu, R., Raich, P., Plomer, K., Dignan, M., Wenzel, L.B., Fairclough, D., Habermann, T., Schnell, L., Quella, S., & Cella, D. (2003). Randomized, controlled trial of an easy-to-read informed consent statement for clinical trial participation: a study of the eastern cooperative oncology group. *Journal of Clinical Oncology*, 21(5), 836-842, https://doi.org/10.1200/JCO.2003.07.022.

Das, M., & Couper, M.P. (2014). Optimizing opt-out consent for record linkage. *Journal of Official Statistics*, 30(3), 479-97, https://doi.org/10.2478/jos-2014-0030.

Dennis, A.R., Yuan, L., Feng, X., Webb, E., & Hsieh, C.J. (2020). Digital nudging: numeric and semantic priming in e-commerce." *Journal of Management Information Systems*, 37 (1), 39-65, https://doi.org/10.1080/07421222.2019.1705505.

Desch, K., Li, J., Kim, S., Laventhal, N., Metzger, K., Siemieniak, D., & Ginsburg, D. (2011). Analysis of informed consent document utilization in a minimal-risk genetic study. *Annuals of Internal Medicine*, 155, 316-322, https://doi.org/10.7326/0003-4819-155-5-201109060-00009.

Dinner, I., Johnson, E.J., Goldstein, D.G., & Liu, K. (2011). Partitioning default effects: why people choose not to choose. *Journal of Experimental Psychology: Applied*, 17(4), 332-41, https://doi.org/10.1037/a0024354.

Dranseika, V., & Piasecki, J. (2020). Transparent defaults and consent for participation in a learning health care system: an empirical study. *Journal of Empirical Research on Human Research Ethics*, 15(4), 261-270, https://doi.org/10.1177/1556264620904272.

Eckman, S., & Haas, G-C. (2017). Does granting linkage consent in the beginning of the questionnaire affect data quality? *Journal of Survey Statistics and Methodology*, 5, 535-551, https://doi.org/10.1093/jssam/smx016.

Edwards, B., & Biddle, N. (2021). Consent to data linkage: experimental evidence on the impact of data linkage requests and understanding and risk perceptions. In *Advances in Longitudinal Survey Methodology*, edited by P. Lynn: Wiley, https://doi.org/10.1002/9781119376965.ch8.

Eisnecker, P.S., & Kroh, M. (2016). The informed consent to record linkage in panel studies: optimal starting wave, consent refusals, and subsequent panel attrition. *Public Opinion Quarterly* 81(1), 131-143, https://doi.org/10.1093/poq/nfw052.

Enama, M.E., Hu, Z., Gordon, I., Costner, P., Ledgerwood, J.E., & Grady, C. (2012). Randomization to standard and concise informed consent forms: development of evidence-based consent practices. *Contemporary Clinical Trials*, 33(5), 895-902, https://doi.org/10.1016/j.cct.2012.04.005.

Ghandour, L., Yasmine, R., & El-Kak, F. (2013). Giving consent without getting informed: a cross-cultural issue in research ethics. *Journal of Empirical Research on Human Research Ethics*, 8(3), 12-21, https://doi.org/10.1525/jer.2013.8.3.12.

Grady, C., Touloumi, G., Walker, A.S., Smolskis, M., Sharma, S., Babiker, A.G., et al. (2017). A randomized trial comparing concise and standard consent forms in the START trial. *PLoS ONE*, 12(4), e0172607, https://doi.org/10.1371/journal.pone.0172607.

Graves, A., McLaughlin, D., Leung, J., et al. (2019). Consent to data linkage in a large online epidemiological survey of 18-23 year old Australian women in 2012-13. *BMC Medical Research Methodology*, 19, 235, https://doi.org/10.1186/s12874-019-0880-z.

Hall, A.E., Sanson-Fisher, R.W., Lynagh, M.C., Threlfall, T., & D'Este, C.A. (2013). Format and readability of an enhanced invitation letter did not affect participation rates in a cancer registry-based study: a randomized controlled trial. *Journal of Clinical Epidemiology*, 66(1), 85-94, https://doi.org/10.1016/j.jclinepi.2012.07.016.

Hewison, J., & Haines, A. (2006). Overcoming barriers to recruitment in health research. *British Medical Journal*, 333(7562), 300-302, https://doi.org/10.1136/bmj.333.7562.300.

Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, 80, 47-58, https://doi.org/10.1016/j.socec.2019.03.005.

Jäckle, A., Beninger, K., Burton, J., & Couper, M.P. (2021). Understanding data linkage consent in longitudinal surveys. In P. Lynn (ed.), *Advances in Longitudinal Survey Methodology*. New York: Wiley, pp. 122-150, https://doi.org/10.1002/9781119376965.ch6.

Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F., & Walzenbach, S. (2021a). Understanding and improving data linkage consent in surveys. *Understanding Society Working Paper 2021-01*. Colchester: University of Essex. Available at https://www.understandingsociety.ac.uk/research/publications/526574

Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F., & Walzenbach, S. (2021b). How and why does the mode of data collection affect consent to data linkage? *Understanding Society Working Paper 2021-04*. Colchester: University of Essex. Available at https://www.understandingsociety.ac.uk/research/publications/536848

Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F., & Walzenbach, S. (2022). Understanding and improving data linkage consent in surveys: User Guide, Version 2.0, January 2022. Colchester: University of Essex. Available at https://reshare.ukdataservice.ac.uk/855464

Jäckle, A., Burton, J., Couper, M.P., & Crossley, T.F. (2022). *Understanding and improving data linkage consent in surveys*, 2018-2019. [data collection]. Colchester, Essex: UK Data Service. https://reshare.ukdataservice.ac.uk/855464/

Kasperbauer, T.J., Halverson, C., Garcia, A., Schwartz, P.H. (2022). Biobank participants' attitudes toward data sharing and privacy: the role of trust in reducing perceived risks. *Journal of Empirical Research on Human Research Ethics*, 17(1-2), 167-176, https://doi.org/10.1177/15562646211055282.

Kincaid, J.P., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). Derivation of new readability formulas (automated readability index, Fog count, and Flesch reading ease formula) for navy enlisted personnel. Memphis, TN: Naval Air Station, Research branch report 8-75.

Kim, E.J., & Kim, S.H. (2015). Simplification improves understanding of informed consent information in clinical trials regardless of health literacy level. *Clinical Trials*, 12(3), 232-236, https://doi.org/10.1177/1740774515571139.

Kongsholm, N.C.H. and Kappel, K. (2017), Is consent based on trust morally inferior to consent based on information? *Bioethics*, 31, 432-442. https://doi.org/10.1111/bioe.12342.

Kreuter, F., Sakshaug, J.W., & Tourangeau, R. (2016). The framing of the record linkage consent question. *International Journal of Public Opinion Research* 28(1), 142-52, https://doi.org/10.1093/ijpor/edv006.

Loewenstein, G., Bryce, C., Hagmann, D., & Rajpal, S. (2015). Warning: You are about to be nudged. *Behavioral Science & Policy*, 1(1), 35-42, https://doi.org/10.1353/bsp.2015.0000.

Matsui, K., Lie, R.K., Turin, T.C., & Kita, Y. (2012). A randomized controlled trial of short and standard-length consent forms for a genetic cohort study: is longer better? *Journal of Epidemiology*, 22(4), 308-316, https://doi.org/10.2188/jea.JE20110104.

McNutt, L.-A., Waltermaurer, E., Bednarczyk, R.A., Carlson, B.E., Kotval, J., McCauley, J., Campbell, J.C., & Ford, D.E. (2007). Are we misjudging how well informed consent forms are read? *Journal of Empirical Research on Human Research Ethics*, 3(1), 89-97, https://doi.org/10.1525/jer.2008.3.1.89.

Molden, D.C. (2014). Understanding priming effects in social psychology: what is "social priming" and how does it occur?" *Social Cognition*, 32 (Supplement), 1-11, https://doi.org/10.1521/soco.2014.32.supp.1.

Paris, A., Brandt, C., Cornu, C., Maison, P., Thalamas, C., & Cracowski, J.-L. (2010). Informed consent document improvement does not increase patients' comprehension in biomedical research. *British Journal of Clinical Pharmacology*, 69, 231-237, https://doi.org/10.1111/j.1365-2125.2009.03565.x.

Paris, A., Deygas, B., Cornu, C., Thalamas, C., Maison, P., Duale, C., Kane, M., Hodaj, E., & Cracowski, J.-L. (2015). Improved informed consent documents for biomedical research do not increase patients' understanding but reduce enrolment: a study in real settings. *British Journal of Clinical Pharmacology*, 80, 1010-1020, https://doi.org/10.1111/bcp.12716.

Pascale, J. (2011). Requesting consent to link survey data to administrative records. Paper presented at the Fourth Conference of the European Survey Research Association (ESRA), Lausanne, Switzerland, July 18-22.

Perrault, E.K., & Keating, D.M. (2018). Seeking ways to inform the uninformed: improving the informed consent process in online social science research. *Journal of Empirical Research on Human Research Ethics*, 13(1), 50-60, https://doi.org/10.1177/1556264617738846.

Perrault, E.K., & McCullock, S.P. (2019). Concise consent forms appreciated—still not comprehended: applying revised common rule guidelines in online studies. *Journal of Empirical Research on Human Research Ethics*, 14(4), 299-306, https://doi.org/10.1177/1556264619853453.

Perrault, E.K., & Nazione, S.A. (2016). Informed consent—uninformed participants: shortcomings of online social science consent forms and recommendations for improvement. *Journal of Empirical Research on Human Research Ethics*, 11(3), 274-280, https://doi.org/10.1177/1556264616654610.

Reisch, L.A., & Sunstein, C.R. (2016). Do Europeans like nudges? *Judgment and Decision Making*, 11, 310-25, http://journal.sjdm.org/16/16202b/jdm16202b.html.

Ripley, K.R., Hance, M.A., Kerr, S.A., Brewer, L.E., & Conlon, K.E. (2018). Uninformed consent? The effect of participant characteristics and delivery format on informed consent. *Ethics & Behavior*, 28, 517-543, https://doi.org/10.1080/10508422.2018.1456926.

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), 166-76, https://doi.org/10.1093/poq/nfu001.

Sakshaug, J.W., Tutz, V., & Kreuter, F. (2013). Placement, wording, and interviewers: identifying correlates of consent to link survey and administrative data. *Survey Research Methods*, 7(2), 133-44, https://doi.org/10.18148/srm/2013.v7i2.5395.

Sakshaug, J.W., Wolter, D., & Kreuter, F. (2015). Obtaining record linkage consent: results from a wording experiment in Germany. *Survey Insights: Methods from the Field,* https://doi.org/10.13094/SMIF-2015-00012.

Sakshaug, J.W., Schmucker, A., Kreuter, F., Couper, M.P., & Singer, E. (2019). The effect of framing and placement on linkage consent. *Public Opinion Quarterly*, 83 (S1), 289-308, https://doi.org/10.1093/poq/nfz018.

Sala, E., Knies, G., & Burton, J. (2014). Propensity to consent to data linkage: experimental evidence on the role of three survey design features in a UK longitudinal panel. *International Journal of Social Research Methodology*, 17(5), 455-73, https://doi.org/10.1080/13645579.2014.899101.

Sepucha, K.R., Fagerlin, A., Couper, M.P., Levin, C.A., Singer, E., & Zikmund-Fisher, B.J. (2010). How does feeling informed relate to being informed? The DECISIONS survey. *Medical Decision Making*, 30 (Supplement), 77s-84s, https://doi.org/10.1177/0272989X10379647.

Singer, E. (1978). Informed consent: consequences for response rate and response quality in social surveys. *American Sociological Review*, 43 (2), 144-162, https://doi.org/10.2307/2094696.

Singer, E., & Frankel, M.R. (1982). Informed consent procedures in telephone interviews. *American Sociological Review*, 47 (3), 416–426, https://doi.org/10.2307/2094999.

Stunkel, L., Benson, M., McLellan, L., Sinaii, N., Bedarida, G., Emanuel, E., & Grady, C. (2010). Comprehension and informed consent: assessing the effect of a short consent form. *IRB: Ethics & Human Research*, 32(4), 1-9, https://www.jstor.org/stable/25703705.

Tait, A.R., Voepel-Lewis, T., Nair, V.N., Narisetty, N.N., & Fagerlin, A. (2013). Informing the uninformed: optimizing the consent message using a fractional factorial design. *JAMA Pediatrics*, 167(7), 640-646, https://doi.org/10.1001/jamapediatrics.2013.1385.

Thaler, R.H., & Sunstein, C.R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Penguin Press.

Thornby, M., Calderwood, L., Kotecha, M., Beninger, K., & Gaia, A. (2018). Collecting multiple data linkage consents in a mixed-mode survey: evidence from a large-scale longitudinal study in the UK." *Survey Methods: Insights from the Field.* Retrieved from https://surveyinsights.org/?p=9734, http://doi.org/10.13094/SMIF-2018-00004.

University of Essex, Institute for Social and Economic Research. (2021). Understanding Society: Innovation Panel, Waves 1-13, 2008-2020. [data collection]. 11th Edition. UK Data Service. SN: 6849, http://doi.org/10.5255/UKDA-SN-6849-14.

Varnhagen, C.K. Gushta, M., Daniels, J., Peters, T.C., Parmar, N., Law, D., Hirsch, R., Takach, B.S., & Johnson, T. (2005). How informed is online informed consent? *Ethics & Behavior*, 15(1), 37-48, http://doi.org/10.1207/s15327019eb1501_3.

Walters, K.A., & Hamrell, M.R. (2008). Consent forms, lower reading levels, and using Flesch-Kincaid readability software. *Drug Information Journal*, 42, 385-394, https://doi.org/10.1177/009286150804200411.

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,* https://doi.org/10.1093/jssam/smab053**Error! Hyperlink reference not valid.**.

Welch, V., Ravanam, M.S., Skalland, B., Ramsey, R., Reagan-Steiner, S., Knighton, C., Hill, H.A., Elam-Evans, L., & Smith, C. (2017). Questionnaire experiments to increase provider consent: the National Immunization Survey. Paper presented at the annual meeting of the American Association for Public Opinion Research, New Orleans, LA, May 17-20.

Yang, D., Fricker, S., & Eltinge, J. (2019). Methods for exploratory assessment of consent-to-link in a household survey. *Journal of Survey Statistics and Methodology*, 7(1), 118–155, https://doi.org/10.1093/jssam/smx031.

Appendix A: Question Wording

A1. Standard (control) versus easy wording

Note: The leaflet and diagrams referred to in the consent questions are archived with the fieldwork documents for the Innovation Panel wave 11. See the "Interviewer materials" available at https://www.understandingsociety.ac.uk/documentation/innovation-panel/fieldwork-documents. The diagrams are under the headings "Data Linkage Flowchart Version A (standard)" and "Data Linkage Flowchart Version B (new)". The leaflet is under the heading "IP11 Information on adding economic records v3".

Standard consent question wording

We would like to add records held by HM Revenue and Customs, or HMRC, containing information on your employment and self-employment history, your income, National Insurance contributions and tax credits. All information will be used for research purposes only by academic or policy researchers under restricted access arrangements which make sure that the information is used responsibly and safely.

Please read this **leaflet** and look at this **diagram** [Version B] explaining how we would like to attach your HMRC records to the answers you have given in this study.

Do you give permission for us to pass your name, address, sex and date of birth to HMRC for this purpose?

1. I have read the leaflet and am happy to give consent
2. I do not want to give consent

Easy consent question wording

We would like to add records held by HM Revenue and Customs, or HMRC, to the answers you have given in this study. If you agree:

- We will send HMRC your name, address, sex and date of birth so that they can identify the records they have about you. The HMRC records contain information about your current and previous employment, your income, National Insurance contributions and tax credits.
- We will not send HMRC the answers you have given in this study.
- HMRC will send us your records. These will contain an anonymous identification number but not your name, address, sex or date of birth.
- We will add the HMRC records to the answers you have given in this study.
- We will make the combined anonymous data available for academic and policy research purposes only.
- Access to the data will be restricted and controlled, to make sure that researchers use the information responsibly and safely.
- This will not affect the way that you deal with the HMRC in any way.

Please read this **leaflet** and look at this **diagram** [Version A] for further information.

Do you give permission for us to pass your name, address, sex and date of birth to HMRC for this purpose?

1. Yes
2. No

A2. Offer of additional information with or without follow-up

Additional information with follow-up

[PAGE 1]

We would like to add records held by HM Revenue and Customs, or HMRC, to the answers you have given in this study. If you agree:

- We will send HMRC your name, address, sex and date of birth so that they can identify the records they have about you. The HMRC records contain information about your current and previous employment, your income, National Insurance contributions and tax credits.
- We will not send HMRC the answers you have given in this study.
- HMRC will send us your records. These will contain an anonymous identification number but not your name, address, sex or date of birth.
- We will add the HMRC records to the answers you have given in this study.
- We will make the combined anonymous data available for academic and policy research purposes only.
- Access to the data will be restricted and controlled, to make sure that researchers use the information responsibly and safely.
- This will not affect the way that you deal with the HMRC in any way.

Do you give permission for us to pass your name, address, sex and date of birth to HMRC for this purpose?

1 Yes

2 I need more information before making a decision

[PAGE 2]

For more information on the data linkage, please read this **leaflet** and look at this **diagram**.

Do you give permission for us to pass your name, address, sex and date of birth to HMRC to link your data?

1 Yes

2 No

Additional information without follow-up

We would like to add records held by HM Revenue and Customs, or HMRC, to the answers you have given in this study. If you agree:

- We will send HMRC your name, address, sex and date of birth so that they can identify the records they have about you. The HMRC records contain information about your current and previous employment, your income, National Insurance contributions and tax credits.
- We will not send HMRC the answers you have given in this study.
- HMRC will send us your records. These will contain an anonymous identification number but not your name, address, sex or date of birth.
- We will add the HMRC records to the answers you have given in this study.
- We will make the combined anonymous data available for academic and policy research purposes only.
- Access to the data will be restricted and controlled, to make sure that researchers use the information responsibly and safely.
- This will not affect the way that you deal with the HMRC in any way.

Please read this leaflet and look at this diagram for further information.

Do you give permission for us to pass your name, address, sex and date of birth to HMRC for this purpose?

1 Yes

2 I need more information before making a decision

3 No

A3. Consent as default wording

We plan to add records held by HM Revenue and Customs, or HMRC, to the answers you have given in this study. This involves the following steps:

- We will send HMRC your name, address, sex and date of birth so that they can identify the records they have about you. The HMRC records contain information about your current and previous employment, your income, National Insurance contributions and tax credits.
- We will not send HMRC the answers you have given in this study.
- HMRC will send us your records. These will contain an anonymous identification number but not your name, address, sex or date of birth.
- We will add the HMRC records to the answers you have given in this study.
- We will make the combined anonymous data available for academic and policy research purposes only.
- Access to the data will be restricted and controlled, to make sure that researchers use the information responsibly and safely.
- This will not affect the way that you deal with the HMRC in any way.

Please read this **leaflet** and look at this **diagram** for further information.

Press “next” to continue.

1. I do not want HMRC records to be added to my answers to this survey.

A4. No prime (control) versus trust prime

The next question is about linking the information you provide in this survey, to data that HM Revenue and Customs, or HMRC, hold about you.

[IF trust treatment group: HMRC is a trusted data holder [display trust symbol]]

1. Continue

(Easy version of consent question used in both control and treatment group.)

A5. Follow-up questions about consent decision process

Objective understanding of data linkage

To help us understand whether the explanation we gave you about linking HMRC data and your answers to this study was clear or unclear, here are a few statements about how the linkage is done. Please specify whether you think each of the statements is true or false.

Answer categories: True/false for each row

- Every researcher can access the combined data via the Internet [false]
- HM Revenue and Customs will combine the information they have with your answers to this study [false]
- Researchers using the data will only have access to anonymous data [true]
- The combined data can be used by HM Revenue and Customs to check that you have been paying your taxes [false]
- HM Revenue and Customs will send us the information they have about you [true]
- Your name, address, sex, and date of birth will be saved with the linked data [false]
- We will send your name, address, sex, and date of birth to HM Revenue and Customs [true]
- HM Revenue and Customs will send us future data about you, unless you object in writing [true]

(Correct responses [indicated in brackets] summed to create index of objective knowledge, ranging from 0-8.)

Subjective understanding of consent request

How well do you think you understand what would happen with your data, if you allowed us to link it to records held by HM Revenue and Customs?

1. I do not understand at all
2. I understand somewhat
3. I mostly understand
4. I completely understand

Confidence in linkage consent decision

We are interested in how people decide whether or not to give us permission to add data held by HM Revenue and Customs to the answers they have given in this study.

How confident are you about the decision decisions you made?

1. Very confident in my decision
2. Confident in my decision
3. Somewhat confident in my decision
4. Not confident in my decision

Appendix B: Multivariable Models of Consent and Objective Knowledge

B.1 Probability of consent: average marginal effects (AME), standard errors (in italics) and p values (in parentheses) estimated from logit models: Innovation Panel (IP)

	IP11 FTF (1)	IP11 FTF (2)	IP11 web
Easy wording	0.025	0.025	0.055
	0.024	*0.024*	*0.027*
	(0.305)	(0.302)	(0.043)
Late placement	-0.066	-0.065	
	0.024	*0.024*	
	(0.006)	(0.006)	
Interaction			
Easy if placement=early		-0.025	
		0.032	
		(0.434)	
Easy if placement=late		0.079	
		0.036	
		(0.027)	
Female	0.026	0.027	-0.007
	0.024	*0.024*	*0.027*
	(0.289)	(0.253)	(0.798)
Age			
[Omitted: age 16-40]			
41-59	-0.125	-0.124	-0.092
	0.031	*0.031*	*0.034*
	(<0.001)	(<0.000)	(0.006)
60+	-0.054	-0.051	0.024
	0.038	*0.038*	*0.044*

	(0.153)	(0.176)	(0.592)
Education			
[Omitted: degree]			
A/AS level	0.098	0.098	-0.003
	0.037	*0.036*	*0.043*
	(0.007)	(0.007)	(0.947)
GCSE or lower	-0.008	-0.010	-0.032
	0.027	*0.027*	*0.031*
	(0.757)	(0.715)	(0.295)
In work	0.093	0.092	-0.015
	0.029	*0.029*	*0.034*
	(0.001)	(0.002)	(0.657)
Household size	-0.005	-0.004	0.010
	0.010	*0.010*	*0.011*
	(0.607)	(0.701)	(0.352)
Home tenure			
[Omitted: home owned outright]			
Home owned with mortgage	-0.037	-0.033	-0.013
	0.035	*0.035*	*0.035*
	(0.300)	(0.351)	(0.714)
Home rented or other	0.027	0.026	0.042
	0.034	*0.034*	*0.041*
	(0.424)	(0.443)	(0.312)
Pseudo R^2	0.030	0.033	0.013
N	1,363	1,363	1,298
Joint tests			
Age	Chi^2=15.87	Chi^2=15.66	Chi^2=11.19
	d.f.=2	d.f.=2	d.f.=2
	P=.0004	P=.0004	P=.0037
Education	Chi^2=9.81	Chi^2=9.91	Chi^2=1.24
	d.f.=2	d.f=2	d.f.=2

	P=.0074	P=.0070	P=.5385
Tenure	Chi^2=3.69	Chi^2=3.16	Chi^2=2.02
	d.f.=2	d.f.=2	d.f.=2
	P=.1581	P=.2059	P=.3647
Wording*placement interaction		Chi^2=4.6	
		d.f.=1	
		P=.032	

B.2 Probability of consent: average marginal effects (AME), standard errors (in italics) and p values (in parentheses) estimated from logit models: Access Panel (AP)

	AP1.1 (1)	AP1.1 (2)	AP1.1 (3)	AP1.2	AP2
Easy wording	0.026			-0.030	
	0.031			*0.034*	
	(0.407)			(0.382)	
Additional information					
[Omitted: easy wording]					
More info with follow up		-0.011			
		0.031			
		(0.734)			
More info without follow up		-0.091			
		0.031			
		(0.003)			
Consent as default			-0.015		
			0.031		
			(0.615)		

Data holder: HMRC					-0.087
					0.022
					(<0.001)
Trust priming					0.051
					0.022
					(0.024)
Female	-0.066	-0.092	-0.091	-0.112	-0.106
	0.031	*0.025*	*0.031*	*0.034*	*0.022*
	(0.037)	(<0.001)	(0.003)	(0.001)	(<0.001)
Age					
[Omitted: age 16-40]					
41-59	-0.047	-0.008	0.035	-0.009	-0.025
	0.040	*0.032*	*0.040*	*0.046*	*0.029*
	(0.238)	(0.814)	(0.373)	(0.846)	(0.387)
60+	0.078	0.098	0.180	0.097	0.056
	0.051	*0.041*	*0.050*	*0.058*	*0.037*
	(0.128)	(0.016)	(<0.001)	(0.097)	(0.132)
Education					
[Omitted: degree]					
A/AS level	0.051	0.071	0.036	-0.051	0.041
	0.043	*0.035*	*0.043*	*0.048*	*0.030*
	(0.230)	(0.041)	(0.406)	(0.296)	(0.172)
GCSE or lower	0.017	0.009	-0.016	-0.005	0.046
	0.036	*0.029*	*0.036*	*0.041*	*0.026*
	(0.648)	(0.755)	(0.655)	(0.905)	(0.081)
In work	-0.005	-0.021	-0.047	-0.051	-0.060
	0.037	*0.030*	*0.036*	*0.043*	*0.027*
	(0.888)	(0.472)	(0.196)	(0.230)	(0.026)
Household size	0.011	0.001	0.004	0.034	-0.004
	0.014	*0.011*	*0.014*	*0.016*	*0.010*
	(0.421)	(0.912)	(0.762)	(0.031)	(0.691)

Home tenure					
[Omitted: home owned outright]					
Home owned with mortgage	0.010	0.009	0.042	0.016	0.123
	0.045	*0.036*	*0.043*	*0.049*	*0.031*
	(0.815)	(0.807)	(0.332)	(0.747)	(<0.001)
Home rented or other	0.033	0.007	0.061	0.078	0.060
	0.045	*0.034*	*0.043*	*0.048*	*0.030*
	(0.462)	0.841	(0.152)	(0.106)	(0.048)
Pseudo R^2	0.010	0.020	0.023	0.024	0.025
N	1,027	1,541	1,023	816	1,921
Joint tests					
Age	Chi^2=7.28	Chi^2=8.84	Chi^2=14.32	Chi^2=4.47	Chi^2=6.15
	d.f.=2	d.f.=2	d.f.=2	d.f.=2	d.f.=2
	P=0.026	P=0.012	P<.001	P=0.107	P= 0.046
Education	Chi^2=1.45	Chi^2=4.54	Chi^2=1.45	Chi^2=1.20	Chi^2=3.50
	d.f.=2	d.f.=2	d.f.=2	d.f.=2	d.f.=2
	P=0.485	P=0.103	P=0.484	P=0.548	P= 0.174
Tenure	Chi^2=0.61	Chi^2=0.07	Chi^2=2.09	Chi^2=3.06	Chi^2=15.51
	d.f.=2	d.f.=2	d.f.=2	d.f.=2	d.f.=2
	P=0.738	P=0.968	P=0.351	P=0.216	P<0.001

Non-significant interaction tests not included in above models:

AP1.1 model (1): Wording*Education: Chi^2=1.03, d.f.=2, p=0.5980

AP1.2: Wording*Education: Chi^2=0.93, d.f.=2, p=0.629

AP2: Trust*Data holder: Chi^2=0, d.f.=1, p=0.9861; Trust*Education: Chi^2=0.41, d.f.=2, p= 0.8159

B.3 Objective Knowledge: coefficients, standard errors (in italics) and p values (in parentheses) estimated from OLS regression models: Innovation Panel (IP)

	IP11 FTF	IP11 web
Easy wording	0.498	0.440
	0.081	*0.093*
	(<0.001)	(<0.001)
Late placement	0.100	
	0.081	
	(0.217)	
Female	-0.014	-0.102
	0.082	*0.094*
	(0.861)	(0.277)
Age		
[Omitted: age 16-40]		
41-59	0.203	0.051
	0.106	*0.114*
	(0.055)	(0.658)
60+	-0.041	0.291
	0.136	*0.148*
	(0.760)	(0.050)
Education		
[Omitted: degree]		
A/AS level	-0.058	-0.489
	0.132	*0.145*
	(0.661)	(0.001)
GCSE or lower	-0.655	-0.611
	0.091	*0.105*
	(<0.001)	(<0.001)
In work	0.423	0.037

	0.098	*0.116*
	(<0.001)	(0.752)
Household size	-0.070	0.032
	0.034	*0.038*
	(0.040)	(0.392)
Home tenure		
[Omitted: home owned outright]		
Home owned with mortgage	-0.056	0.211
	0.117	*0.118*
	(0.634)	(0.073)
Home rented or other	-0.222	0.019
	0.117	*0.140*
	(0.058)	(0.894)
Constant	4.864	4.479
	0.192	*0.207*
	(<.000)	(<.001)
Pseudo R^2	0.107	0.048
N	1,345	1,156
Joint tests		
Age	F=2.81	F=2.01
	d.f.=2, 1333	d.f.=2, 1145
	P=.0608	P=.1347
Education	F=29.31	F=17.82
	d.f.=2, 1333	d.f.=2, 1145
	P=<.0001	P=<.0001
Tenure	F=2.05	F=2.00
	d.f.=2, 1333	d.f.=2, 1145
	P=.1287	P=.1364

B.4 Objective Knowledge: coefficients, standard errors (in italics) and p values (in parentheses) estimated from OLS regression models: Access Panel (AP)

	AP1.1 (1)	AP1.1 (2)	AP1.1 (3)	AP1.2	AP2
Easy wording	0.381			0.370	
	0.089			0.103	
	(<0.001)			(<0.001)	
Additional information					
[Omitted: easy wording]					
More info with follow up		-0.066			
		0.090			
		(0.462)			
More info without follow up		-0.083			
		0.090			
		(0.354)			
Consent as default			0.020		
			0.090		
			(0.828)		
Data Holder: HMRC					-0.156
					0.065
					(0.016)
Trust priming					0.092
					0.065
					(0.156)
Female	0.038	-0.111	-0.021	0.000	-0.038
	0.091	*0.074*	*0.091*	*0.104*	*0.065*
	(0.678)	(0.135)	(0.819)	(0.999)	(0.562)
Age					
[Omitted: age 16-40]					
41-59	0.257	0.182	0.107	0.451	0.191
	0.114	0.095	0.116	0.137	0.082

	(0.024)	(0.055)	(0.358)	(0.001)	(0.021)
60+	0.196	0.152	0.226	0.260	0.209
	0.147	*0.118*	*0.146*	*0.174*	*0.107*
	(0.183)	(0.198)	(0.122)	(0.136)	(0.051)
Education					
[Omitted: degree]					
A/AS level	-0.146	-0.026	0.075	-0.141	-0.079
	0.123	*0.101*	*0.125*	*0.144*	*0.088*
	(0.234)	(0.796)	(0.549)	(0.330)	(0.367)
GCSE or lower	-0.534	-0.504	-0.359	-0.501	-0.387
	0.105	*0.086*	*0.106*	*0.121*	*0.076*
	(<0.001)	(<0.001)	(0.001)	(<0.001)	(<0.001)
In work	-0.153	-0.091	-0.106	-0.023	-0.150
	0.106	*0.087*	*0.106*	*0.127*	*0.078*
	(0.150)	(0.292)	(0.317)	(0.857)	(0.055)
Household size	-0.023	-0.005	-0.008	-0.027	0.006
	0.040	*0.033*	*0.040*	*0.047*	*0.029*
	(0.570)	(0.890)	(0.846)	(0.573)	(0.844)
Home tenure					
[Omitted: home owned outright]					
Home owned with mortgage	-0.065	-0.095	-0.099	-0.289	0.066
	0.128	*0.104*	*0.127*	*0.148*	*0.093*
	(0.613)	(0.363)	(0.439)	(0.050)	(0.476)
Home rented or other	0.068	-0.170	-0.170	-0.263	-0.194
	0.128	*0.126*	*0.126*	*0.144*	*0.088*
	(0.597)	(0.177)	(0.177)	(0.069)	(0.027)
Constant	4.509	4.928	4.865	4.527	4.918
	0.206	*0.172*	*0.204*	*0.241*	*0.152*
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
R^2	0.048	0.030	0.027	0.061	0.032
N	1,024	1,536	1,020	816	1,916

Joint tests					
Age	F=2.58 d.f.=2, 1013 P=0.076	F=1.90 d.f.=2, 1524 P=0.150	F=1.21 d.f.=2, 1009 P=0.298	F=5.45 d.f.=2, 805 P=0.004	F=3.05 d.f.=2, 1904 P= 0.048
Education	F=13.58 d.f.=2, 1013 P<0.001	F=20.21 d.f.=2, 1524 P<0.001	F=8.20 d.f.=2, 1009 P<0.001	F=8.90 d.f.=2, 805 P<0.001	F=14.11 d.f.=2, 1904 P<0.001
Tenure	F=0.66 d.f.=2, 1013 P=0.515	F=0.46 d.f.=2, 1524 P=0.630	F=0.91 d.f.=2, 1009 P=0.402	F=2.32 d.f.=2, 805 P=0.099	F=5.14 d.f.=2, 1904 P= 0.006

Non-significant interaction tests not included in above models:

AP1.1 (1): Wording*Education: F=1.76, d.f.=2, 1011, p=0.172

AP1.2: Wording*Education: F=2.18, d.f.=2, 803, p=0.114

AP2: Trust*Data holder: F=0, d.f.=1, 1903, p=0.991; Trust*Education: F=0.06, d.f.=2, 1902, p=0.9420