

Understanding Society Working Paper Series

No. 2021 – 02

February 2021

Experiments on multiple requests for consent to data linkage in surveys

Sandra Walzenbach¹, Jonathan Burton², Mick P. Couper³, Thomas F. Crossley⁴ and Annette Jäckle²

¹University of Konstanz

² University of Essex

³ University of Michigan

⁴ European University Institute





Economic and Social Research Council



Non-technical summary

It is increasingly common for researchers to link survey data to administrative data. If several administrative data sources are of interest, respondents are required to give consent to each of them, meaning that multiple consent questions have to be included in one survey. Existing literature suggests that individual consent varies widely between administrative data sources and over time, but little is known about how respondents process multiple requests for consent in a single survey.

Using an online access panel in the UK, we conducted a set of experiments in two surveys to explore multiple consent requests (covering five domains or data sources).

In the first study we experimentally varied the format of the request, testing three versions: 1) a sequence of pages (with one response per domain), 2) all five requests on the same page (with one response per domain), and 3) a single request (with one joint request covering all five domains). We also varied the order of the domains. We find that average consent rates do not differ by format, but asking a less sensitive or easier-to-comply request first yields slightly higher average consent rates than asking the more request first.

We repeated the order experiment in a second study, using an independent sample from the same panel, and adding two more order conditions. We find average consent rates are not affected much by order, but the consent to individual domains is affected by order. However, we fail to replicate the pattern of consents found in the first study. We conclude that the order in which multiple consent requests is asked does matter, but in complicated ways that depend on the particular outcomes in which one is interested. Objective knowledge and subjective comprehension of the consent process, and confidence in the decision are largely unaffected by format or order.

Experiments on multiple requests for consent to data linkage in surveys

Sandra Walzenbach (University of Konstanz) Jonathan Burton (University of Essex) Mick P. Couper (University of Michigan) Thomas F. Crossley (European University Institute) Annette Jäckle (University of Essex)

Abstract

Surveys are increasingly seeking informed consent from respondent to link to several administrative data sources. This paper explores how best to administer consent requests for multiple data linkages in a survey. We experimentally varied the format of the consent request and the order of five data sources in two surveys. The format of the request did not have much effect on consent rates. The order does matter, but in complicated ways that depend on the particular outcomes of interest. Objective knowledge and subjective comprehension of the consent process, and confidence in the consent decision are unaffected by format or order.

Keywords: Informed consent; record linkage; administrative data; multiple requests

JEL classification: C81, C83

Acknowledgements: This research was funded by the Nuffield Foundation (<u>www.nuffieldfoundation.org</u>) with co-funding from the UK Economic and Social Research Council (ESRC, <u>https://esrc.ukri.org/</u>) (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). The views expressed here are those of the authors and not necessarily of the Nuffield Foundation or the ESRC.

Corresponding author: Sandra Walzenbach, University of Konstanz, Germany, Sandra.Walzenbach@uni-konstanz.de.

Introduction

With advancements in technology and analysis methods to deal with so-called "big data," in recent years researchers have been faced with many potential new data sources that can be linked to survey data. While big data comes in many forms, including social media data, transaction data and sensor data, our focus is on administrative records (namely tax and income, pension, energy consumption, education and health records), typically held by government departments. Administrative records are increasingly becoming accessible to researchers and integrating longitudinal survey and administrative data provides wide-ranging opportunities for new research (Benzeval et al. 2020). For the two data sources to be combined they need to be linked and – except in the case of some government surveys – respondents must consent to such linkage. Consent, and more specifically informed consent, is a necessary requirement for exploiting the power of integrated data to address important research questions and policy issues.

As the availability of administrative data sources grows and as the demand for linkage increases, there is a growing need to seek consent from respondents to link to multiple data sources. While there is a growing body of research on differences between consenters and non-consenters, and emerging studies on how best to ask respondents for consent, little is known about multiple requests for linkage consent in a single survey. This paper presents evidence from several survey experiments that were designed to find out if respondents' consent can be maximised by presenting the consent requests in a certain order or format (e.g. on the same versus on separate pages). These order and format effects are examined with respect to three different outcomes: average consent rates, share of respondents who refused or agreed to all requests, and the consent rates for each domain over an individual sequence of consent requests. Additional outcomes include objective and subjective measures of knowledge, i.e. how informed the consent is. The paper also touches on how respondents process such requests and which mechanisms might explain the patterns we find.

Previous Research and Background Literature

So far, methodological research on data linkage has largely focused on single consent questions. Previous research has shown that the probability of consent varies by topic matter (Sakshaug, Couper, Ofstedal and Weir 2012), between interviewers (Korbmacher and Schroeder 2013; Sakshaug, Tutz and Kreuter 2013), and even within respondents over time (Weir, Faul and Ofstedal 2014). This evidence suggests that people do not have strong fixed views on consent and that the decision to consent can be influenced by features of the request. However, studies examining what influences the decision to consent have had limited success. Findings on the effects of respondent

characteristics on consent are inconsistent (Sakshaug et al. 2012; Sala, Burton and Knies 2012), as is evidence of the effects of interviewer characteristics (Sala, Knies and Burton 2014). One consistent finding is on mode, with web surveys yielding substantially lower levels of consent than intervieweradministered surveys (Jäckle, Beninger, Burton and Couper in press; Sakshaug, Hülle, Schmucker and Liebig 2017).

Experimental studies examining how the consent question is worded (Pascale 2011; Sakshaug et al. 2013), or whether it is asked in an earlier or later wave of a panel survey (Eisnecker and Kroh 2016; Sala et al. 2014) found no effects. Asking for consent after a module of questions related to the content of the data to be linked increases consent compared to asking at the end of the questionnaire (Sala et al. 2014), and asking it at the beginning of the survey rather than the end has a positive effect (Sakshaug, Schmucker, Kreuter, Couper and Singer 2019). Several experiments have varied the framing of the request (gain vs. loss), with mixed results (see Kreuter, Sakshaug and Tourangeau 2016; Sakshaug et al. 2013; Sakshaug, Wolter and Kreuter, 2015; Sakshaug, Stegmaier, Trappmann and Kreuter, 2019; Sakshaug et al. 2019).

Research on multiple consent requests within one questionnaire is still very rare. One recent exception is a survey experiment on order effects by Weiß et al. (2019), who hypothesized a possible fatigue or ceiling effect: "As respondents do not know how many data linkage requests they will see, they might be willing to consent to the first ones, and then reach a critical point where they are not willing to share more information. Besides such a ceiling effect, every question is an additional intrusion into privacy, so that also the contextual sensitivity of the request might increase with each question" (p. 2). They found higher consent rates to the first consent request of a sequence, irrespective of the consent domain.

Thornby, Calderwood, Kotecha, Beninger, and Gaia (2018) conducted a qualitative study on how to present multiple consent requests for administrative data linkage. They provide interesting anecdotal evidence on variables that increased reported consent in their qualitative interviews: trust, sensitivity, perceived benefits, being given assurances, the feeling of having "nothing to hide", and a linkage request referring to already collected data rather than data to be collected in the future. In addition, the authors give a hint that fatigue is a response mechanism worth thinking about in the context of multiple consent questions. They report decreasing levels of invested cognitive effort over the consent sequence. However, their study does not provide any quantitative data on trust and sensitivity levels, size of effects on consent or the relative importance of variables.

Given the paucity of research on multiple consent requests in surveys, are there other literatures that may inform our expectations of the effects? There is a relatively large literature on sequential

requests in social psychology (see Dillard 1991). Two well-studied mechanisms within this literature are the foot-in-the-door (FITD) and door-in-the-face (DITF). FITD can be traced back to the research on "compliance without pressure" by Freedman and Fraser (1966, p. 195) who posited that "once a person has been induced to comply with a small request he is more likely to comply with a larger demand." The effect is generally explained in term of self-perception theory (see Bem 1972): complying with an easy or small request engenders the perception of being a cooperative person, which leads to compliance with larger subsequent requests (Guadagno, Asher, Demaine and Cialdini 2001). In contrast, DITF posits that "a request made is more likely to be agreed to if preceded by the offer and refusal of a more expensive one" (Cialdini et al. 1975). See Dolinski (2011) for a fuller discussion of these two mechanisms.

These two contrasting views have generated a lot of research, including several meta-analyses, on the conditions under which one or the other is likely to prevail. Fern, Monroei and Avila (1986) conclude from their meta-analysis that FITD is likely to be a more effective strategy than DITF, whereas Pascual and Guéguen (2005) found that the two strategies had similar mean compliance rates. Dillard, Hunter and Burgoon (1984) and Burger (1999) both conclude from their meta-analyses that the effectiveness of the strategies depends on a number of other factors, such as the timing and relative size of the requests and whether incentives were used.

We know of only two applications of these ideas in the survey literature. Brehm (1994) used FITD to explain the effect of small incentives in gaining compliance to a survey. Acquisti, John and Loewenstein (2012) varied the order of intrusiveness of sensitive question to examine the effect on disclosure. They argued (p. 161) that "altering the order of intrusiveness of a set of questions is akin to asking respondents to comply with requests of different magnitude and therefore is comparable to the literature on "foot-in-the-door" (FITD) (Freedman and Fraser 1966) or "door-in-the-face" (DITF) techniques (Cialdini et al. 1975)." Contrary to the FITD expectation and the prevailing survey literature advocating for starting with less-sensitive questions, their results suggest that "beginning with milder questions only to move toward more intrusive questions may actually elicit lower overall willingness to divulge" (Acquisti et al. 2012, pp. 172-3).

To what extent are these mechanisms relevant for the multiple consent request situation? First, almost all of these studies examine a pair of requests, where only the second one is the outcome of interest. However, in the linkage consent setting, the goal is to maximize consent across all requests. Second, the first request in most DITF experiments is designed to be sufficiently large that most people will decline, generating reciprocal concessions when a smaller second request is made. In our

case, the requests are all of relatively similar magnitude. Nonetheless, consistent with Acquisti and colleagues, we view the psychological literature on sequential requests as relevant to our research.

In summary, both the social psychological and survey literature suggest that the order in which the consent requests are made is likely to matter. The research also suggests that the format of presentation should matter, but is less clear on what the expected effects are. But from a practical perspective, how the sequence of requests is presented is likely to be of importance. We focus on these two elements of multiple consent requests. We formulate our specific research questions and hypotheses below.

Research Questions / Hypotheses

Following the review of relevant literature above, we focus on three research questions.

RQ1: Does the order matter?

Specifically we examine whether starting with a less sensitive or easier-to-comply request may lead to greater overall compliance than starting with a more sensitive or harder-to-comply request. The perceived sensitivity of the data linkage request could be affected by a number of factors, including the sensitivity of the information held by the data holder, knowledge of the data holder and the data it has, trust in the data holder, the risk of disclosure, etc., and thus varies across both individuals and data domains. Here we use the term to describe relative overall rates of consent observed in the population from prior studies, thus a less-sensitive request is one that is easier to comply with on average (i.e. has higher overall consent rates).

RQ2: Does the format matter?

From a practical perspective, there are many different ways to present a sequence of requests. One approach may be to reveal each request in sequence, such that when responding to the first consent request, respondents are unaware that later requests will follow. An alternative is to reveal all consent requests together, so that the respondent is aware that they will be asked for consent to linkage to multiple administrative datasets. Yet another alternative is to show the multiple consent requests, but with a single yes/no question indicating consent to all or none. Of course, there are many other ways in which the multiple consent requests could be presented, including, for example, dispersed throughout the questionnaire (embedded in the relevant context). Similarly the number of requests could be varied. We focus on the three alternative formats outlined above, and described in more detail below.

RQ3: How can we explain what we find?

Our research design was not set up to explicitly test the competing mechanisms (i.e., both FITF and DITF would predict higher consent to later items), but we can use these to help understand the results we obtain. We expect that starting with an easier-to-comply request may lead to greater compliance with later requests (the FITD effect). We also expect to see a decline in consent across individual consent questions (regardless of order), suggesting a possible fatigue effect.

Outcomes of Interest

When examining multiple consents, there are several outcomes of interest. Which one is crucial depends on the goals of the specific research project. Should the average consent rate across all domains be maximised? Or do we want to minimise the share of respondents who do not agree to any type of linkage? Alternatively, we might want to maximise the consent rate in a specific consent domain. It is quite obvious from this example that there is not one most informative outcome, but several that might be affected by design choices in different ways. As a result, we analyse the effects of question order and format with regard to three different consent outcomes:

- the average consent rates
- the share of respondents who said YES or NO to all five requests
- the patterns of individual consents

It is only in the last of these analyses that we can explore the underlying mechanisms of sequential requests.

In addition to obtaining record-linkage consent, it is important from a research ethics perspective that such consent is informed. This could be both objective (do respondents understand the consent process) and subjective (are they confident they made the right decision). Thus, as secondary outcomes, we also examine the effect of the experimental manipulations on indicators of objective and subjective comprehension and confidence, time taken to answer the consent questions, and whether respondents consulted additional materials describing the linkage process.

We conducted two sets of experiments. Given that the second experiment was designed to further explore some of the findings from the first experiment, we describe each of the experiments in turn.

Experiment 1

1) Data Source and Experimental Design

Experiment 1 was implemented in the online access panel PopulusLive (see https://www.populuslive.com/). This is a non-probability online panel which had around 130,000 active sample members at the time, who are recruited through web advertising, word of mouth, and database partners. Quotas based on age, gender and education were set to match the characteristics of the *Understanding Society* Innovation Panel sample¹. Once the target number of respondents for a quota was reached, the survey was closed to further respondents with that characteristic.

The implementation of these surveys was led by NatCen Social Research, in collaboration with the PopulusLive panel. Respondents were asked permission to link their survey data with five different administrative data sources: tax data (held by HM Revenue and Customs, HMRC), benefit and pensions data (held by the Department for Work and Pensions, DWP), household energy data (held by the Department for Business, Energy and Industrial Strategy, BEIS), education data (held by the Department for Education in England, DfE, and equivalent departments in Scotland and Wales, EDUC), and health data (held by the National Health Service, NHS).

We employed a 2*3 experimental design, varying both the order and format of the consent requests, as described earlier. With regard to order, we tested two sequences:

- HMRC DWP BEIS EDUC NHS ("HMRC first")
- NHS EDUC BEIS DWP HMRC ("NHS first")

These two orders were chosen to vary the sensitivity of the request being made. This was determined on an empirical basis, using observed consent rates to these domains from a variety of sources, including the Millennium Cohort Study (see Mostafa, 2016), Next Steps (Thornby et al. 2018) and *Understanding Society* (Jäckle et al. in press). The ranking was consistent across these data sources, with NHS achieving the highest consent rates and HMRC the lowest. The order was also supported by evidence from qualitative interviews (Jäckle et al. in press).

¹ This research is part of a larger project, involving experiments on the Innovation Panel as well as the access panel. For further details, see Jäckle, Burton, Couper, Crossley and Walzenbach (2021) available at https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2021-01.pdf.

With regard to format, three different versions were tested:

- Sequence of pages (with one response per domain)
- Same page (with one response per domain)
- Single request (with one joint request covering all 5 domains)

In the first scenario, respondents were presented with five consecutive pages with requests for data, each page asking about one of the linkage domains. The second scenario contained the same requests but presented all of the requests on one survey page. This means that respondents had to scroll through the questions, but could see all of the requests before answering any of them. In the third scenario, all the consent domains were listed on one page, but respondents were asked to make a single joint decision for linkage to all five domains. Appendix A1 contains wording of the consent questions for the sequences of pages and single request versions. The wording of the same page version is identical to that for the sequence of pages but without the page breaks.

In addition to the consent questions, we asked a number of background questions. The survey also included a series of follow-up questions related to the consent request, e.g., how well respondents understood the request and how they made a decision. Additional questions on how much people trusted the data holders and organisations involved and how sensitive they found the content of the administrative records were asked. The full questionnaire can be found at https://www.iser.essex.ac.uk/research/projects/understanding-and-improving-data-linkage-consent-in-surveys. We note that survey responses were not actually linked to administrative data. Respondents were informed of this at the end of the survey.

The study was conducted in June 2018. A total of 46,206 panelists were invited to the survey, of whom 6,532 started the survey and 5,633 completed it (401 broke off and 498 were screened out). The survey took about 11 minutes to complete (median value for the experimental groups that answered multiple consent questions; 80% of respondents took between 6 and 21 minutes). The fully-crossed experimental design resulted in six cells, with 511-521 respondents in each.

2) Results

As outlined above, order and format effects can have effects on different outcomes of interests. Namely, these are (a) the overall average consent rates in each experimental condition, (b) the share of respondents who said YES or NO to all five requests, and (c) the pattern of individual consents over the whole sequence of questions (for the respondents who could give separate answers for the five consent domains). We are also interested in the effect of order and format of asking multiple consents on d) comprehension of the consent request, subjective understanding (how informed respondents feel) and confidence in the consent decision.

a) Average consent rates

Figure 1 shows the average consent rates for the six experimental groups (see Appendix Table B1 for additional details)². Note that the horizontal axis is truncated to show the effect. Markers in the same row stem from the same formats (sequences of pages / same page / single request), while equal marker shapes indicate the same question order (HMRC first vs NHS first). It is clear from the figure that format has little effect on average consent rates, while order does appear to have an effect, with lower average consent rates in the three HMRC-first groups (48.0% on average; indicated by circles) than in the NHS- first groups (51.6% on average; indicated by diamonds). To test these effects, we ran a logistic regression model, with cluster-robust standard errors to account for the multilevel-structure of the data (domains nested within respondents). Regressing all of the individual responses on order and format, we find that question order has a significant effect on consent (p=0.036), while format fails to reach statistical significance (p=0.79) (see Appendix B2 for the full regression model).

² For the single request version, this is the average across respondents. For the multiple request versions, this is the average across both respondents and domains.

Figure 1 Average consent rates by experimental condition



b) Yes-to-all / No-to-all requests

The second outcome of interest is the share of respondents who said yes or no to all five requests for data linkage. Figure 2 distinguishes between respondents who always gave consent, always denied consent, and respondents who gave mixed answers, that is, they agreed to data linkage for some domains and refused for others. Groups that differ in question order but share the same format are plotted next to each other. In the groups asking for consent to each of the five domains separately, 76% to 87% of respondents say either yes-to-all or no-to-all, that is, they answer consistently to all five requests. The last pair of bars shows the groups that were asked a single consent question for all five domains together. By design, they had to say yes or no to the set of domains; mixed answers were not possible.



Figure 2 Yes-to-all / No-to-all requests by experimental condition

Comparing the share of respondents who said yes (or no) to all requests across *formats* reveals no differences between *sequence of pages* and *same page*. The *single request* group differs significantly from the other two formats (p<.01 in tests of equal proportion and p<.001 in regression models controlling for order). The single request has a higher rate of "yes to all" than the other two groups. However, it also has a higher rate of "no to all". Respondents who are unable to give a mixed answer in this group appear to distribute their answers roughly equally between these two responses. On the one hand, this suggests that asking a single yes/no consent question does not produce higher rates of non-consent. This is potentially good news, considering that respondents in the qualitative study (Beninger et al. 2017) had expressed a preference for separate questions for each domain over such a "catch all" or joint question. On the other hand, a sizeable minority of respondents (ranging from 12.9% to 23.7%) differentiated between domains in their consent decisions, pointing to the potential need to provide respondents with this opportunity. Further, individual data holders may not accept such a joint decision request for ethical or legal reasons. Our conclusion is thus that the format of presentation of multiple consent domains has little impact on the consent rates.

With regard to *order* effects, there are no differences between the shares of yes-to-all answers within any format group (see each pair of adjacent bars for comparison). However, there is some evidence of significant differences in the shares of no-to-all answers in both the sequence of pages and same page formats: when HMRC consent is asked first, a higher share of respondents say no to all requests than when NHS consent is asked first (p=.06 for sequence of pages and p=.01 for same page in tests of equal proportion and p=.002 in a regression model controlling for format).

b) Individual consents

Next we examine order and format effects across the sequence of individual consents. We can observe patterns of answers in the four experimental groups where respondents answered the five linkage requests separately. Figure 3 illustrates the two orders and two formats in which the five questions were asked. Letters indicate the domain (N for NHS, H for HMRC, etc.). The individual rates are also shown in Table B3 in the Appendix.



Figure 3 Consent to individual domains by order and format

Note again that the vertical axis is truncated to illustrate the effect. An inspection of Figure 3 suggests that order affects consent rates more than format. The two formats show very similar curves, while the question order is the characteristic that divides the curves into the upper and the lower part of a scissor shape: testing the differences in consent rates within the same consent domain and question order (e.g. the difference between same page and sequence of pages for BEIS linkage if HMRC was asked first), none of the comparisons yields significant format effects.

It is also clear from Figure 3 that asking for NHS consent first results in higher consent rates to NHS than when asking NHS last; a difference of 8.9 percentage points. Asking HMRC consent first does not appear to change HMRC consent rates relative to asking HMRC last; a difference of 1.9 percentage points. Focusing just on the first consent request, the fact that NHS consent rates are significantly higher than HMRC consent rates (on average 53.8% for NHS and 45.6% for HMRC; p<0.01) is consistent with previous findings about differences between domains.

To test for order or position effects, we need to take into account that the experimental design was limited to only two orders. This means that the middle domain (BEIS) is the only one with a constant position. In all other cases, the consent domain and position are confounded. For this reason, Table 1 focuses on the consent rates for the BEIS request by order and format. Since its location is constant throughout all experimental groups, we can get valid measures of the order effect of the preceding questions (HMRC-first or NHS-first) and format on BEIS consent.

BEIS consent rate	HMRC first	NHS first	Difference	р
Sequence of pages	47.4%	53.8%	6.4	0.04
Same page	50.5%	54.4%	3.9	0.21
Combined	48.9%	54.1%	5.2	0.02

Table 1 Order and format effects on BEIS consent rates

As already seen in Figure 3, the differences by format are small and non-significant (p=0.32 for HMRC first and p=0.85 for NHS first), so we focus on differences by order. In both formats, BEIS gets a higher consent rate when NHS is asked first than when HMRC is asked first. This reaches statistical significance in the sequential format (p=0.04), but not in the same page format (p=0.21). This

suggests a carryover effect: when the more sensitive request (HMRC) is asked first, consent to BEIS linkage is lower than when the less sensitive request (NHS) is asked first.

In summary, across these three measures of consent rates, we find some evidence that asking a smaller or less sensitive request (NHS) first yields higher average consent rates than asking the more sensitive request (HMRC) first (see Figure 1). We also see that BEIS consent is increased slightly when following NHS consent than when following HMRC consent (see Table 1). However, we see little effect of order on the proportion who consent to all five requests (see Figure 2). Figure 3 also suggests little evidence of a foot-in-the-door carryover effect across all five domains. In addition, there is little evidence of a fatigue effect, which should lead to decreasing rates of consent across the five domains. We thus conclude that the order in which multiple consent requests are asked matters, but in complicated ways that depend on the particular outcomes in which one is interested.

d) Comprehension, feeling informed, and confidence

As noted above, it is not sufficient to simply maximize consent *rates*. It is also important to ensure that *informed* consent is not negatively affected by experimental manipulations designed to increase rates of consent. To this end, we examined the effect of the order and format manipulations on objective knowledge of the consent process (measured by 8 true-false questions; see Jäckle, Burton, Couper, Crossley and Walzenbach 2021), subjective understanding of the consent request (using the question "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-not at all, 4=completely), and confidence in the linkage consent decision ("How confident are you about the consent decisions you made?" 1=very confident, 4=not confident), all asked about a single domain (HMRC).

In results not shown here (see Appendix Table B4), we find no differences in these three measures by order of the consent requests. We find small differences by format, but these are not readily interpretable. For instance, we find slightly higher objective knowledge scores for the sequence of pages than for the single-page and single-request versions, but slightly higher subjective knowledge ratings for the single-request version, and no differences in confidence in the decision. We thus conclude that the format and order of presentation of multiple consent requests has no reliable effect on how well informed respondents are about the linkage process, and on how informed and confident they feel about their decision.

One final factor to consider is how much time it takes to make the consent decision. Across both orders, we find that the sequence of pages takes the longest (median of 71.6 seconds), followed by the single page version (67.1 seconds) and then the joint request version (44.6 seconds) (see Appendix Table B4). We also examined whether respondents looked at additional materials (the leaflet and diagram referred to in Appendix A) at different rates by experimental condition, and found no evidence of this (p=0.23; results not shown). Across all conditions, between 75.3% and 81.5% did not access the additional materials, while between 13.7% and 16.2% looked at both. These results suggest that the joint request condition (with a single yes/no question) may be more efficient (i.e. takes less time), without compromising objective or subjective knowledge or confidence, or negatively affecting consent rates. However, such as approach may not meet the legal requirements of data holders.

Experiment 2

Given our limited variation of different order sequences and the confounding of order and domain (by design) in Experiment 1, we could only examine the effect of order on a single consent domain, BEIS. We had an opportunity to replicate the experiment a few months later on the same panel. This second experiment is described in further detail below, focusing on differences from the first experiment.

1) Data Source and Experimental Design

Experiment 2 was implemented in the same online access panel PopulusLive, with invitations restricted to those who had not completed the first experiment. Similar quotas and procedures were employed with the goal of keeping the two samples as comparable as possible.

- We added two additional order sequences to the two used in the first experiment. In these new sequences, we switched the order of the second and fourth requests (DWP and EDUC) to give us additional orders. We tested the following four sequences:
 - HMRC DWP BEIS EDUC NHS ("HMRC 1st, DWP 2nd")
 - HMRC EDUC– BEIS– DWP NHS ("HMRC 1st, EDUC 2nd")
 - NHS EDUC BEIS DWP HMRC ("NHS 1st, EDUC 2nd")
 - NHS DWP–BEIS– EDUC– HMRC ("NHS 1st, DWP 2nd")

The first and third sequences are the same as those used in Experiment 1. The additional of the second and fourth sequences allows us to explore the effects of asking NHS or HMRC first on consent rates to DWP and EDUC, in addition to BEIS.

Given the lack of effect of the format manipulation, all four orders were asked using a sequence of separate pages. This results in a single-factor experiment with four conditions.

The study was conducted in December 2019. A total of 30,682 panelists 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). The survey included four additional experimental conditions (single-consent questions) not used here. Median completion time for the four experimental groups that answered multiple consent questions was 10 minutes. Each sequence was answered by between 479 and 487 respondents.

2) Results

We examined the same three consent rate outcomes as in Experiment 1. Table 2 shows the average consent rates by order.

Order	Average consent rate	(s.e.)	(n)
HMRC 1 st , DWP 2 nd	50.2%	2.1%	487
HMRC 1 st , EDUC 2 nd	49.3%	2.1%	479
NHS 1 st , DWP 2 nd	49.1%	2.1%	482
NHS 1 st , EDUC 2 nd	45.8%	2.0%	481

Table 2 Average consent rates by order, Experiment 2

The first thing we note from Table 2 is that while the overall consent rates are similar to those from Experiment 1 (on average 49.0% for the sequence of pages conditions in Experiment 1 versus 48.6% for all groups in Experiment 2), we did not replicate the slight positive effect of asking NHS consent first. Directly comparing the two groups that were exact replications of the first study, we find an average consent rate of 50.2% if HMRC was asked first and 45.8% if NHS was asked first. In fact, the group where NHS consent was asked first, followed by EDUC, has a noticeably lower consent rate than the other three groups. We have conducted a number of additional analyses to explore

possible errors in the implementation of Experiment 2 and disproportionate allocation of quotas across the groups, but find no evidence of errors, and no explanation for this finding.

Figure 4 shows the results for yes-to-all, no-to-all, and mixed responses across the five domains, similar to Figure 2 above for Experiment 1. The fourth group (NHS 1st, EDUC 2nd) also yielded a lower rate of "yes-to-all" (34.3%) than the other groups in Experiment 2 (40.2% for the third group, also with NHS first), but also lower than the corresponding group in Study 1 (39.4%).



Figure 4 Yes-to-all / No-to-all requests by experimental condition, Experiment 2

Figure 5 shows the pattern of consents to individual domains, by order. In contrast with Experiment 1 (see Figure 3), we no longer see the scissor pattern we observed (see Appendix C1 for details). Recall that in Figure 3, consent to NHS differed by whether it was asked first or last, while HMRC consent did not differ much by position. In Experiment 2 (Figure 5), order seems to matter more for HMRC than for NHS.





Finally, Table 3 shows the individual consent rates for the three "target" domains (DWP, EDUC and BEIS) for evaluating carryover effects.

Domain	Position	(n)	HMRC first	NHS first	Difference	Significance
BEIS	third	(1,929)	51.6%	50.4%	-1.2	p=0.60
DWP	second	(969)	46.6%	47.1%	0.5	p=0.88
	fourth	(960)	45.3%	39.7%	-6.6	p=0.08
EDUC	second	(960)	52.6%	50.3%	-2.3	p=0.48
	fourth	(969)	52.8%	48.8%	-4.0	p=0.21

Table 3 Order effects on target domains, Experiment 2

We find that the results for BEIS consent also do not replicate. Recall that in Experiment 1, BEIS had a 6.4 percentage point *higher* consent rate when following NHS consent than when following HMRC consent. In Experiment 2 this difference trends in the opposite direction (1.2 percentage points *lower* when following NHS) but does not reach traditional levels of significance. That is, we no longer find a positive carryover effect on BEIS consent of asking NHS consent first. The results for the other two consent domains (DWP and EDUC) also do not show a consistent pattern by order.

The differences between the consent rates for NHS and HMRC when each is asked first (i.e. unaffected by subsequent domains) are in the expected direction (on average 48.7% for HMRC and 53.0% for NHS, a 4.3 percentage point difference) but not as large as observed in the first experiment, and fail to reach traditional levels of significance (p=0.06). These surprising failures to replicate our earlier results suggest caution in interpreting the results from the first experiment.

We also looked at objective knowledge, subjective comprehension, confidence in the decision and time to respond by order of the consent domains (see Appendix Table C2). We find no significant differences in these outcomes by experimental treatment. In addition, we examined differences in consulting the leaflet and brochure, and again found no significant differences (p=0.78).

Overall Conclusions and Discussion

Across two experiments, we find that asking multiple consents across several separate pages versus asking them together on one page does not seem to affect consent rates or other key outcomes. While quicker, asking a joint consent request (yes/no to all 5 domains) does not have significant advantages in terms of consent rates or understanding and confidence. The order in which a series of linkage consent requests is made affects both overall rates of consent and consent rates to individual domains. While the effect is not yet clearly understood, and does not replicate across the two experiments, it is clear that the order of the requests matters.

We do not find clear evidence of a foot-in-the-door-effect (a positive effect on later requests of an easier initial request), a door-in-the-face request (a positive effect on later requests of a harder initial request), or a fatigue effect (declining rates of consent across the five requests). The fact that we do observe carryover effects, but that the effects do not align with expectations from the sequential requests literature, is an area for further research.

From a practical perspective, it is clear that practitioners need to take care when asking for consent to multiple domains within a single survey. The order in which these are asked has potential consequences for the consent rates obtained to each of the domains. While we did not discern a clear pattern across the two experiments that would lead to a specific recommendation, we urge

caution in implementing multiple consent requests within a single survey. More research is needed to better understand the process of asking consent for multiple linkage domains. We did not test separating the consent requests across the survey (i.e. embedded within questions related to the domain in question). This is worth considering in future research, although it likely can only be done in the context of a longer survey in which questions on each relevant domain are asked.

References

Acquisti, A., John, L.K., and Loewenstein, G. (2012), "The Impact of Relative Standards on the Propensity to Disclose." *Journal of Marketing Research*, 49 (2): 160-174.

Bem, D.J. (1972), "Self-perception Theory." In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology (Vol. 6)*. New York: Academic Press, pp. 1-62.

Benzeval, M., Bollinger, C.R., Burton, J., Couper, M.P., Crossley, T.F., and Jäckle, A. (2020), "Integrated Data: Research Potential and Data Quality." *Understanding Society Working Paper 2020-*02. Colchester: University of Essex. Available at

https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2020-02.pdf.

Brehm, J. (1994), "Stubbing Our Toes for a Foot in the Door? Prior Contact, Incentives and Survey Response." *International Journal of Public Opinion Research*, 6 (1): 45-63.

Burger, J.M. (1999), "The Foot-in-the-door Compliance Procedure: A Multiple-process Analysis and Review." *Personality and Social Psychology Review*, 3: 303-325.

Cialdini, R.B., Vincent, J.E., Lewis, S.K., Catalan, J., Wheeler, D., and Darby, B.L. (1975), "Reciprocal Concessions Procedure for Inducing Compliance: The Door-in-the-face Technique." *Journal of Personality and Social Psychology*, 31: 206-215.

Dillard, J.P. (1991), "The current status of research on sequential-request compliance techniques." *Personality and Social Psychology Bulletin*, 17 (3): 283-288.

Dillard, J.P., Hunter, J.E., and Burgoon, M. (1984), "Sequential-Request Persuasive Strategies: Metaanalysis of Foot-in-the-door and Door-in-the-face." *Human Communication Research*, 10 (4): 461-488.

Dolinski, D. (2011), "A Rock or a Hard Place: The Foot-in-the-Face Technique for Inducing Compliance Without Pressure." *Journal of Applied Social Psychology*, 41: 1514-1537.

Eisnecker, P.S., and 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.

Fern, E., Monroei, S., and Avila, R. (1986), "Effectiveness of multiple request strategies: A synthesis of research results." *Journal of Marketing Research*, 23: 144-152.

Freedman, J.L., and Fraser, S.C. (1966). "Compliance without pressure: The foot-in-the-door technique." *Journal of Personality and Social Psychology*, 4: 195–202.

Guadagno, R.E., Asher, T., Demaine, L.J., and Cialdini, R.B. (2001), "When Saying Yes Leads to Saying No: Preference for Consistency and the Reverse Foot-in-the-Door Effect." *Personality and Social Psychology Bulletin*, 27 (7): 859–867.

Jäckle, A., Beninger, K., Burton, J., and Couper, M.P. (in press), "Understanding Data Linkage Consent in Longitudinal Surveys." To appear in P. Lynn (ed.), *Advances in Longitudinal Survey Methodology*. New York: Wiley.

Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F., and Walzenbach, S. (2021) "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/sites/default/files/downloads/working-papers/2021-01.pdf

Korbmacher, J.M., and Schroeder, M. (2013), "Consent When Linking Survey Data with Administrative Records: The Role of the Interviewer." *Survey Research Methods*, 7 (2): 115-31.

Kreuter, F., Sakshaug, J.W., and Tourangeau, R. (2016), "The Framing of the Record Linkage Consent Question." *International Journal of Public Opinion Research* 28(1):142-52.

Mostafa, T. (2016), "Variation within Households in Consent to Link Survey Data to Administrative Records: Evidence from the UK Millennium Cohort Study." *International Journal of Social Research Methodology*, 19 (3): 355-375.

Pascale, J. (2011), "Requesting consent to link survey data to administrative records: Results from a split-ballot experiment in the survey of health insurance and program participation." Technical report, survey methodology series 2011-03, United States Census Bureau.

Pascual, A., and Guéguen, N. (2005), "Foot-in-the-Door and Door-in-the-Face: A Comparative Meta-Analytic Study." *Psychological Reports*, 96 (1): 122–128.

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

Sakshaug, J.W., Hülle, S., Schmucker, A., and Liebig, S. (2017), "Exploring the Effects of Interviewerand Self-Administered Survey Modes on Record Linkage Consent Rates and Bias." *Survey Research Methods* 11(2):171-188. Sakshaug, J.W., Schmucker, A., Kreuter, F., Couper, M.P., and Singer, E. (2019), "The Effect of Framing and Placement on Linkage Consent." *Public Opinion Quarterly*, 83 (S1): 289-308.

Sakshaug, J.W., Stegmaier, J., Trappmann, M., and Kreuter, F. (2019), "Does Benefit Framing Improve Record Linkage Consent Rates? A Survey Experiment." *Survey Research Methods*, 13(3), 289-304

Sakshaug, J.W., Tutz, V., and Kreuter, F. (2013), "Placement, Wording, and Interviewers: Identifying Correlates of Consent to Link Survey and Administrative Data." *Survey Research Methods* 7(2):133-44.

Sakshaug, J.W., Wolter, D., and Kreuter, F. (2015), "Obtaining Record Linkage Consent: Results from a Wording Experiment in Germany." *Survey Insights: Methods from the Field*. Retrieved from http://surveyinsights.org/?p=7288.

Sala, E., Burton, J., and Knies, G. (2012), "Correlates of Obtaining Informed Consent to Data Linkage: Respondent, Interview and Interviewer Characteristics." *Sociological Methods and Research* 41(3):414-39.

Sala, E., Knies, G., and Burton, J. (2014), "Propensity to Consent to Data Linkage: Experimental Evidence on the Role of Three Survey Design Features in a UK Longitudinal Panel." *International Journal of Social Research Methodology* 17(5):455-73.

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

Weir, D.R., Faul, J.D., and Ofstedal, M.B. (2014), "The Power of Persistence: Repeated Consent Requests for Administrative Record Linkage and DNA in the Health and Retirement Study." Presented at the Panel Survey Methods Workshop. Ann Arbor, MI.

Weiß, B., Beuthner, C., Silber, H., Keusch, F., Menold, N., and Schröder, J. (2019), "Consenting to Data Linkage: The Role of Data Domain, Framing, Device, Incentives, and Respondent Characteristics." Presented at the Mobile Apps and Sensors in Surveys (MASS) workshop (University of Mannheim and Utrecht University).

Appendix A: Consent Question Wording

a) Sequence of pages

Most sensitive first sequence: HMRC

[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.

Please read this <u>leaflet</u> and look at this <u>diagram</u> 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

Most sensitive first sequence: DWP

[Page 2]

We would also like to add records held by the Department for Work and Pensions, or DWP, to the answers you have given in this study. The DWP records contain information on your benefit claims and time on employment programs. The process of linking is the same as described for HMRC.

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

1 Yes

2 No

Most sensitive first sequence: BEIS

[Page 3]

We would also like to add records held by the Department for Business, Energy and Industrial Strategy (BEIS), to the answers you have given in this study. The BEIS records contain information on energy use and energy efficiency for each address in the country. The process of linking is the same as described earlier.

Do you give permission for us to pass your name and address to BEIS for this purpose?

1 Yes

Most sensitive first sequence: EDUC

[Page 4]

We would also like to add records held by the Department for Education {if COUNTRY=England} / the Education Analytical Services {if COUNTRY=Scotland} / the Department for Education and Skills {if COUNTRY=Wales}, to the answers you have given in this study. These records contain information on schools or colleges you've attended, the courses you've studied and the qualifications you've gained. The process of linking is the same as described earlier.

Do you give permission for us to pass your name, address, sex and date of birth to the Department for Education {if COUNTRY=England} / the Education Analytical Services {if COUNTRY=Scotland} / the Department for Education and Skills {if COUNTRY=Wales} for this purpose?

1 Yes

2 No

Most sensitive first sequence: NHS

[Page 5]

We would also like to add records held by the National Health Service, or NHS, to the answers you have given in this study. The NHS records contain information on your health treatment and use of health services. The process of linking is the same as described earlier.

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

1 Yes

b) Single Response

Most sensitive first

We would like to add records held by various government departments to the answers you have given in this study. If you agree:

- We will send the departments your name, address, sex and date of birth so that they can identify the records they have about you.
- We will not send the answers you have given in this study to the departments.
- The departments 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 records from the departments 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 departments in any way.

Please read this <u>leaflet</u> and look at this <u>diagram</u> for further information.

These are the data we would like to link:

- Records held by HM Revenue and Customs, or HMRC, contain information about your current and previous employment, your income, National Insurance contributions and tax credits.
- Records help by the Department for Work and Pensions, or DWP, contain information on your benefit claims and time on employment programs.
- Records help by the Department for Business, Energy and Industrial Strategy (BEIS), contain information on energy use and energy efficiency for each address in the country.
- Records held by the Department for Education {if COUNTRY=England} / the Education Analytical Services {if COUNTRY=Scotland} / the Department for Education and Skills {if COUNTRY=Wales} contain information on schools or colleges you've attended, the courses you've studied and the qualifications you've gained.
- Records held by the National Health Service, or NHS, contain information on your health treatment and use of health services.

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

1 Yes

Appendix B: Experiment 1 Results

Order	Format	Average consent rate	(s.e.)	(n)
HMRC first	Sequence of pages	47.0	2.0%	511
HMRC first	Same page	48.3	2.1%	521
HMRC first	Single request	48.6	2.2%	518
NHS first	Sequence of pages	51.0	2.0%	515
NHS first	Same page	52.2	2.0%	515
NHS first	Single request	51.6	2.2%	519

Table B1 Average consent rates by order (Experiment 1)

B2 Logistic regression of consent on order and format (Experiment 1)

	β (t-statistics)
HMRC first	-0.156*
	(-2.10)
Same page	0.0507
	(0.63)
Single request	0.0466
	(0.55)
constant	0.0373
	(0.55)
Ν	11,347
McFadden Pseudo R ²	0.001

* p<0.05 ** p<0.01 *** p<0.001. Based on logistic regression with cluster-robust standard errors.

Logit coefficients with t-statistics in parentheses.

B3 Consent to individual domains by o	order and format (Experiment 1)
---------------------------------------	---------------------------------

	Question position				
	1	2	3	4	5
Sequence of pages, HMRC first	46.2	44.4	47.4	48.7	48.1
Same page, HMRC first	45.1	45.5	50.5	50.7	49.5
Sequence of pages, NHS first	53.4	55.7	53.8	47.6	44.5
Same page, NHS first	54.2	56.3	54.4	48.5	47.8

B4 Objective knowledge, subjective understanding, confidence and response time (Experiment 1)

2	Ohiactiva	knowladge k	w order an	d format	(maanc)	۱.
a	Objective	KIIUWIEUge L	y uluel all	u ioiniat	(IIICalis)	1

	Sequence of pages (1)	Single Page (2)	Single Request (3)	p-value (1 vs 2)	p-value (1 vs 3)	p-value (2 vs 3)
Objective score (both orders)	4.5	4.4	4.4	0.02	0.06	0.66
Objective score (HMRC 1 st)	4.6	4.3	4.4	0.02	0.13	0.42
Objective score (NHS 1 st)	4.5	4.4	4.4	0.37	0.27	0.86
p-value (HMRC 1 st vs NHS 1 st)	0.55	0.45	0.83			
Significance tests using Student's t-test						

b) Subjective understanding by order and format (means)

	Sequence	Single				
	of pages	Page	Single Request	p-value	p-value	p-value
	(1)	(2)	(3)	(1 vs 2)	(1 vs 3)	(2 vs 3)
Subj understanding (both orders)	2.2	2.2	2.3	0.73	0.02	0.01
Subjective understanding (HMRC 1 st)	2.2	2.2	2.3	0.68	0.09	0.20
Subjective understanding (NHS 1 st)	2.2	2.2	2.3	0.35	0.11	0.01
p-value (HMRC 1 st vs NHS 1 st)	0.42	0.60	0.47			
Significance tests using Student's t-test						

c) Confidence by order and format (means)

		Single				
	Sequence of pages	Page	Single Request	p-value	p-value	p-value
	(1)	(2)	(3)	(1 vs 2)	(1 vs 3)	(2 vs 3)
Confidence (both orders)	2.8	2.8	2.9	0.77	0.34	0.50
Confidence (HMRC 1 st)	2.8	2.9	2.9	0.74	0.45	0.65
Confidence (NHS 1 st)	2.8	2.8	2.8	0.94	0.57	0.62
p-value (HMRC 1 st vs NHS 1 st)	0.64	0.45	0.47			
Significance tests using Student's t-test						

d) Response times by order and format (medians)

	Sequence of pages (1)	Single Page (2)	Single Request (3)	p-value (1 vs 2)	p-value (1 vs 3t)	p-value (2 vs 3)
Time to answer all consents (both orders)	71.6	67.1	44.6	0.01	0.00	0.00
Time to answer all consents (HMRC 1 st)	69.5	67.8	45.1	0.62	0.00	0.00
Time to answer all consents (NHS 1 st)	73.0	65.8	43.9	0.00	0.00	0.00
p-value (HMRC 1 st vs NHS 1 st)	0.06	0.29	0.74			

Significance tests conducted using Wilcoxon Rank Sum tests

Appendix C: Experiment 2 Results

	Position					
	1	2	3	4	5	
HMRC 1 st , DWP 2 nd	49.8	46.6	51.3	52.8	50.7	
HMRC 1 st , EDUC 2 nd	47.6	52.6	52.0	45.3	49.1	
NHS 1 st , DWP 2 nd	55.2	47.1	51.0	48.8	43.4	
NHS 1 st , EDUC 2 nd	50.7	50.3	50.0	39.8	38.0	

C1 Consent to individual domains by order (Experiment 2)

C2 Objective knowledge, subjective understanding, confidence and response time (Experiment 2)

	Objective knowledge	Subjective knowledge	Confidence (mean)	Time (median)
	(mean)	(mean)		
HMRC 1 st , DWP 2 nd	4.4	2.2	2.8	60.3
HMRC 1 st , EDUC 2 nd	4.4	2.2	2.8	61.9
NHS 1 st , DWP 2 nd	4.3	2.2	2.8	63.3
NHS 1 st , EDUC 2 nd	4.3	2.2	2.8	61.4
p-value (ANOVA/Kruskal-Wallis test)	p=0.47	p=0.62	p=0.73	p=0.70