



**Understanding Society
Working Paper Series**

No. 2023 – 07

March 2023

The effects of position, context, and order on consent to data linkage in a survey

Jonathan Burton¹, Mick P. Couper² and Annette Jäcke¹

¹ University of Essex

² University of Michigan



**Economic
and Social
Research Council**



Non-technical summary

We report on an experiment conducted in the *Understanding Society* COVID-19 study in March 2021. Survey respondents were asked for consent to link their survey responses to two sources of government administrative data (NHS health records and Cancer and Death Register data), as well as to be mailed a COVID-19 antibody test kit.

We experimentally varied where in the survey we placed the consent requests (early, early after some questions about COVID-19, or late in the survey) and the order in which we asked the consent questions (linkage first or serology first). We also examine reasons respondents gave for consent or non-consent. We find that order of the requests does not make much difference to whether or not respondents consent, but making the requests early in the survey significantly increases consent rates compared to asking them after a series of content-related questions or later in the survey.

The most frequent reason chosen for not consenting was “too personal, shared enough information”. Respondents were more likely to give this reason if they were asked for consent at the end of the survey. This finding points to some form of fatigue, as does the effect of the placement of the consent request we find.

The most common reason chosen for consenting was “to support research”. Our finding that trust – whether in the organization (the National Health Service) or the study (*Understanding Society*) – is selected more often as a reason for consenting by those who got the consent questions early suggests that the prevailing belief that trust is built up over the course of a survey (used as justification for asking questions like these at the end of the survey) is not empirically supported in our study.

The results of this research have implications for survey design and how best to collect informed consent for data linkages, in a way that maximizes rates of consent without compromising understanding of the requests.

The effects of position, context, and order on consent to data linkage in a survey

Jonathan Burton (University of Essex), Mick P. Couper (University of Michigan), and Annette Jäckle (University of Essex)

Abstract

We report on an experiment where we ask survey respondents to consent to two forms of data linkage to health records and to consent to be mailed a serology kit. We varied the placement (early, early in context or late in the survey) and order (linkage first or serology first) of the consent requests. We also examine reasons for consent or non-consent. We find that order of the requests does not make much difference, but making the requests early in the survey significantly increases consent rates over asking them after a series of content-related questions or later in the survey.

Keywords

Administrative records, linked data, experiment

JEL classification

C81, C83

Acknowledgements

The *Understanding Society* COVID-19 study was funded by the [Economic and Social Research Council](#) (ES/K005146/1) and the [Health Foundation](#) (2076161). Fieldwork for the web survey was carried out by [Ipsos MORI](#); fieldwork for the telephone survey by [Kantar](#). *Understanding Society* is an initiative funded by the [Economic and Social Research Council](#) (ES/T002611/1) and various Government Departments, with scientific leadership by the [Institute for Social and Economic Research](#), University of Essex. Fieldwork for the annual survey is carried out by [Kantar](#) and [NatCen](#). The research data are distributed by the [UK Data Service](#).

Ethics approval was granted by the University of Essex Ethics Committee for the COVID-19 web and telephone surveys (ETH1920-1271). The March 2021 web survey was reviewed and ethics approval granted by the NHS Health Research Authority, London – City & East Research Ethics Committee (reference 21/HRA/0644).

Corresponding author

Jonathan Burton, Institute for Social and Economic Research, University of Essex, Colchester, Wivenhoe Park, Essex CO4 3SQ, United Kingdom, jburton@essex.ac.uk, +44 1206 872266.

1. Introduction

Surveys are increasingly asking respondents to do more than answer questions. They are being asked for consent to link data from administrative records, social media accounts, or transaction data to their survey responses (e.g., Mneimneh et al. 2021; Sakshaug et al. 2012; Sloan et al. 2020). They are being asked to perform additional tasks, such as provide blood or saliva samples (e.g., Dykema et al. 2017; O'Doherty et al. 2014; Sakshaug, Couper and Ofstedal 2010), wear devices or download apps to track their movements or activity (e.g., Bergmann, Franzese and Schrank 2022; Keusch et al. 2019; Kreuter et al. 2020), and so on. These enhancements to surveys may increase the breadth and depth of data available to researchers. With some exceptions, a key requirement is that survey respondents provide informed consent to the additional requests.

In contrast to studies relying on self-selected volunteers, an explicit goal of probability-based sample surveys is to make valid inferences to the broader population. Non-consent may lead to selection biases, threatening the inferential value of these linked datasets. Finding ways to maximize consent without compromising respondent autonomy is important for maximizing the benefit of such survey enhancements for research.

As the demand for more data rises and new sources of digital data become accessible for linkage, survey respondents are increasingly seeing multiple requests for consent. 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 (exceptions are Walzenbach et al. 2022; Beuthner et al. 2023). Understanding how survey respondents react to these requests and how the position and order of these requests affects consent is an important step in minimizing consent bias when multiple requests are made.

Much of the methodological research on data linkage requests has focused on single consent questions. 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 (Eckman and Haas 2017; Jäckle et al. 2023; Sakshaug et al. 2019). This runs counter to much anecdotal advice, which is to make such requests near the end of the survey, giving time to build rapport and trust. In our view, this might be due to respondent fatigue which might diminish the chances of a positive response by the end of a survey.

In the context of multiple consent requests, Beuthner et al. (2023), hypothesized a possible fatigue effect, with respondents consenting to the first one, but then reaching a critical point where they are not willing to share more information. They tested consent to seven different data types (e.g. administrative data, sensor data, social media data, biomeasures) in an opt-in panel. For each type, asking it as the first consent request resulted in higher consent rates than asking it after other consents. Similarly, Walzenbach et al. (2022) explored order effects for five different administrative linkage types. However, they did not find evidence of fatigue (declining rates of consent across the five requests). While they did not find clear patterns in consent rates across the five domains, they conclude that “the order in which these are asked has potential consequences for the consent rates obtained to each of the domains.” These limited studies suggest that the order of multiple requests matters, but more research is needed to see how consistent these effects are across domains and surveys, and to understand the reasons behind these mixed findings.

Given the limited research on how best to ask multiple consent requests, we designed an experiment to test the effect of position, context and order on the rate of consent to multiple requests. We address the following research questions:

RQ1. Does the placement of the consent questions affect consent rates?

Consistent with prior research, we expect to find higher rates of consent when the request is made early in the survey rather than later. We also expect that making the request in the context of relevant questions will result in higher consent rates than asking at the end of the survey.

RQ2. Does the order of consent questions affect consent rates? Is there an interaction between order and placement of consent questions?

Here we have no strong expectations, given the mixed evidence in the literature. But no prior study has explored placement and order together for multiple consent requests, so this is a novel contribution to the literature.

RQ3. Why does the placement of consent questions affect consent rates?

This analysis is largely exploratory. Given the previous research suggesting that early placement of the requests is beneficial, we seek to understand possible reasons for this somewhat surprising finding.

We describe the study and experimental design in the next section.

2. Methods

The experiment was embedded in the March 2021 COVID-19 study, conducted as part of *Understanding Society*: the UK Household Longitudinal Study. *Understanding Society* is a representative panel study of adults age 16+ in the United Kingdom. The COVID surveys, begun in April 2020, were designed as short (20 minute) frequent web surveys (with a parallel telephone survey in some waves) to measure the impact of the pandemic on panel members. All eligible members of the main *Understanding Society* panel were invited to the COVID surveys, unless they had expressly opted out. Invitations and reminders were sent by email and SMS. Respondents for whom no mobile number or email address were known were sent an invitation letter by post. Details of the COVID-19 studies are found in the User Guide (ISER, 2021). The data are available through the UK Data Service (see University of Essex, 2021).

The March 2021 survey included two requests to participants for their permission to link data held about them by the National Health Service (NHS) to their survey responses. We also included an invitation to have a serology kit sent to participants so they could take a blood sample and return it to test for COVID-19 antibodies. We thus consider responses to three consent questions: (1) to link to NHS data, (2) to link to information about cancer and death registrations held by the General Registrar, and (3) to allow us to send a serology test kit through the post.

We took the opportunity to implement an experiment around the placement and the order of the three consent questions in the survey. To investigate the effect of position on consent rates, we randomly allocated participants to be asked these questions (1) early in the survey (after the household relationships module, but before any substantive questions), (2) early-in-context (after the coronavirus vaccination module), or (iii) late in the survey (after the General Health Questionnaire (GHQ), containing a series of questions about mental health and stress). To examine the effect of order of the questions, participants were randomly allocated to receive the (1) health

linkage questions (NHS and Register data) first, or (2) the serology kit question first. Crossing these two factors results in a 3x2 design.¹

A total of 19,280 adults were invited to complete the March COVID-19 study. Of these 11,802 fully completed the survey (61.2% of issued), 233 partially completed or broke off (1.2%), and 7,245 were non-respondents (37.6%). Invitees were randomly allocated to the experimental conditions in equal proportions prior to fieldwork.

For each of the three consent requests, we coded respondents who answered “don’t know” or “prefer not to say” as non-consenters; observations where the answer to the consent question was “missing” or “ineligible” (because they dropped out prior to the consent question or were ineligible for the serology testing due to a health condition), were excluded from the analysis.

Following the consent questions respondents were asked a module of follow up questions. For each of the three consent questions, they were asked a closed question about why they had consented/not consented (depending on their answer). They were then asked four true/false questions to test their understanding of the linkage process for health data. The full wording of all questions is reproduced in Appendix A.

More information about the COVID-19 study is available in the User Guide. The full March 2021 (Wave 8) questionnaire is available at: <https://www.understandingsociety.ac.uk/documentation/covid-19/questionnaires>. The consent questions and other items used in our analyses are reproduced in Appendix A.

All analyses are unweighted and do not account for the complex sample design. Our focus is on differences between the experimental conditions, rather than inference to the broader population.

3. Results

Overall, 79.6% of respondents consented to receive the serology kits, while 71.2% consented to link to NHS data and 69.6% to register data. Around three-fifths (60.6%) consented to all three requests, while 11.7% declined all three, and the balance (27.6%) consented to some combination of the three. There is a strong overlap of NHS and Register linkage consent ($\phi^2=0.749$), but a weaker overlap between either of those and serology ($\phi^2=0.130$ and 0.117 respectively), suggesting that people think of data linkage and serology requests as different types.²

RQ1: Does the placement of the consent question affect consent rates?

A higher proportion of participants consented to each of the requests when they were asked earlier. Table 1, shows the consent rate for each request when asked early, compared to when asked late. When the consent questions were asked early, 66.3% of participants agreed to all three requests. This agreement rate was 9 percentage points lower (57.2%) when the requests were made late in the survey. The individual requests show the same pattern, consistent with prior research.

¹ The study included a 7th cell, with both the serology consent (after the coronavirus vaccination module) and the health linkage questions asked after the General Health Questionnaire) asked in context. We do not include this cell in our analyses.

² Consenting to receive a serology kit does not necessarily mean returning the kit. Overall, around 67% of those who consented returned the kit and provided a sample. This did not differ significantly by position of the consent question.

Table 1: Consent to link or send kit, by placement of question

Consent to...	Early	Early, in context	Late	Difference (and p-value) early vs. in-context	Difference (and p-value) early vs. late
NHS data	75.6%	72.2%	68.4%	3.4 (0.001)	7.2 (<0.001)
Register data	74.4%	70.2%	66.5%	4.2 (<0.001)	7.9 (<0.001)
Send serology kit	81.6%	77.9%	78.4%	3.7 (<0.001)	3.2 (0.001)
All three	66.3%	60.0%	57.2%	6.3 (<0.001)	9.0 (<0.001)

Notes: P-values from Chi² tests. Cell sizes are documented in Appendix Table B1.

Also shown in in Table 1 is the effect of asking consent in context (after the coronavirus vaccination module). Prior research would suggest that doing so would provide a rationale for the request, thereby increasing consent rates. We acknowledge that putting the consent questions in context (by definition) changes the placement of the request. Nonetheless, when the consents are asked in context (at around 4 minutes into the survey), the consent rate is significantly *lower* than when it is asked early in the survey before any of the health questions (around 1.7 minutes in). This suggest that placement is more important than context.

There is a caveat to the effect of placement. Respondents who are asked for consent early in the survey are more likely to break off (4.1%) than those asked early in context (2.7%) or late (2.1%) (See Appendix Table C1). The increase in the break off rate is smaller than the increase in the consent rate, though, so there still a net gain in consent, although there is a slight loss in completed (as opposed to partial) surveys.

The higher rate of consent from the early placement does not come at a cost of reduced understanding of the consent request. There was no difference between the three placement groups in their objective understanding of the linkage request: in each group respondents answered on average 2.2 of the (true/false) knowledge test questions correctly (see Appendix A for the wording of the objective understanding questions).

RQ2: Does the order of consent questions affect consent rates? Is there an interaction between order and placement of consent questions?

Table 2 shows the main effect of order on consent rates. There is no difference in the consent to link to administrative data or to be sent a serology kit if the health consents are asked first, or serology is asked first.

Table 2: Consent to link or send kit, health or serology asked first

Consent to...	Health asked first	Serology asked first	Difference	P-value
NHS data	71.8%	72.3%	-0.5%	0.588
Register data	71.0%	69.7%	1.3%	0.140
Send serology kit	79.0%	79.6%	-0.6%	0.480
All three	60.7%	61.6%	-0.9%	0.333

Notes: P-values from Chi² tests. Cell sizes are documented in Appendix Table B2.

To test whether the effect of the position of the consent request depends on the order of the requests, we test the interaction between the experimental treatments. For each of the consent requests, we estimate a logit model of the probability of consent, regressed on binary indicators for the position of the request (early/late), the order of the request (health/serology first), and the interaction between the two, to test whether the effect of the early versus late treatment is different when health is asked first compared to when serology is asked first.

Table 3 shows the predicted probabilities of consent estimated from the logit models, for each request, by position and order. If serology is asked first, asking NHS consent early in the survey yields a consent rate that is 5.1 percentage points higher than asking late. If NHS linkage is asked first, asking early yields a consent rate that is 9.2 percentage points higher than asking late. The difference-in-difference estimate of 4.1 percentage is significant at the 10% level. This indicates that asking health consent early in the survey increases consent rates regardless of the order of the consent requests; however, the effect of asking early is larger when the health consent is asked first.

For register data linkage and serology the interactions are not significant. This indicates that the effect of the placement of these requests (early or late) does not depend on which is asked first (health or serology). In sum, asking consent early in the survey is always better than asking late; for health consent it is in addition best to ask health as the first of the consent requests.

Table 3: Interaction of position and order of consent requests

	Early		Late				Diff-in-		P-
	Pr(consent)	S.E.	Pr(consent)	S.E.	Delta	S.E.	Diff	S.E.	value
NHS data									
Serology first	74.5	1.04	69.4	1.11	5.1	1.52			
Health first	76.6	1.00	67.4	1.10	9.2	1.48	4.1	2.12	0.052
Register data									
Serology first	72.5	1.07	65.8	1.14	6.8	1.56			
Health first	76.1	1.00	67.2	1.10	8.9	1.49	2.2	2.16	0.312
Send serology kit									
Serology first	80.8	0.95	78.8	1.00	2.0	1.38			
Health first	82.4	0.91	77.9	0.98	4.4	1.34	2.4	1.92	0.207

Notes: predicted probabilities of consent from logit model of consent, regressed on position, order and their interaction. Diff-in-Diff: difference in Delta between serology first and health first groups.

RQ3. Why does the placement of consent questions affect consent rates?

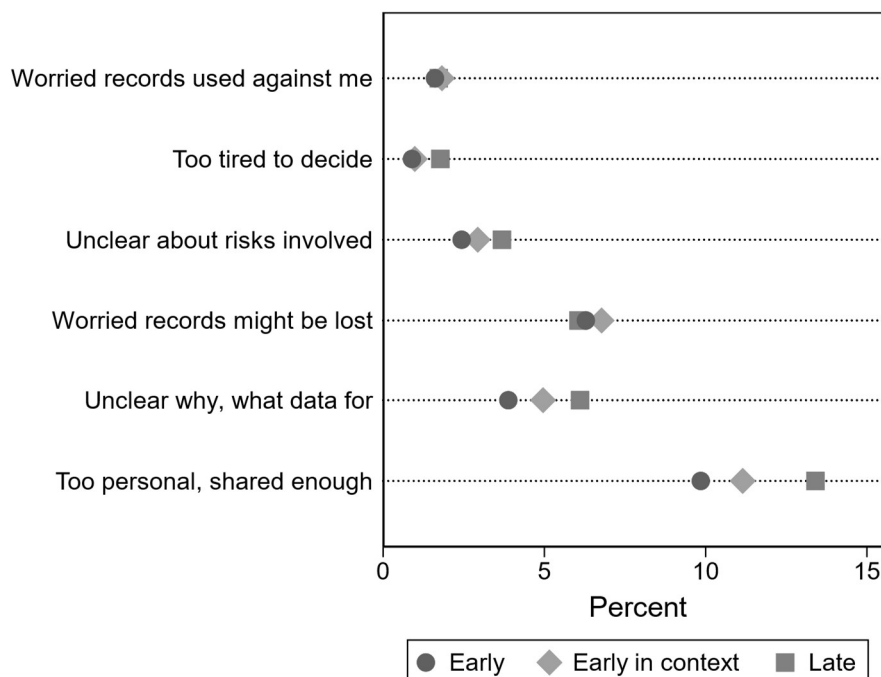
To examine why respondents are more likely to consent when the request is placed earlier in the questionnaire, we asked respondents about the reasons why they did or did not consent. The questions about reasons for not consenting included six response options that were the same for all three consent requests. The reasons for consenting similarly included five common response options. For both types of questions the response options were randomized and respondents could select all answers that applied to them (see Appendix A1).

For each consent request (health, registry, serology) and each reason for not consenting, we create an indicator variable with value 1 if the respondent mentioned this reason, and 0 if the respondent

either consented or did not consent but did not mention this reason.³ For each of these binary indicators we estimate a pooled logit regression of the probability that the respondent mentioned that reason for not consenting, regressed on the position indicator and a control for the consent question (health, registry, serology). Standard errors are adjusted for the clustering of observations in respondents, since each respondent contributes three observations, one for each consent request. Appendix Table C3 contains the predicted probabilities estimated from the logit models, for each reason for not consenting, by position of the consent request. Figure 1 graphically represents the predicted probabilities. The most common reason stated is that the request was too personal, the respondent had already shared enough information with the study.

There are significant differences between placement groups for four of the six reasons for not consenting. Respondents in the late group were significantly more likely to say they were too tired than respondents in the early or early in context group; they were also more likely to say the information was too personal, they had shared enough information, than respondents in either of the other groups. Respondents in the early group were less likely to say they were unclear about the risks involved than respondents in the late group, and less likely to say they were unclear why they were being asked and what would be done with the data than respondents in either of the other two groups. There are no differences between any of the three groups in the percentage of respondents who reported concern that their records might be used against them, or concern that their records might be lost, stolen, or hacked. If we restrict the analyses to non-consenters only, a slightly different pattern emerges (see Appendix Figure D1): those in the early group are more likely to say they were worried their records could be lost/stolen/hacked, and the differences in “too personal, shared enough” responses are smaller.

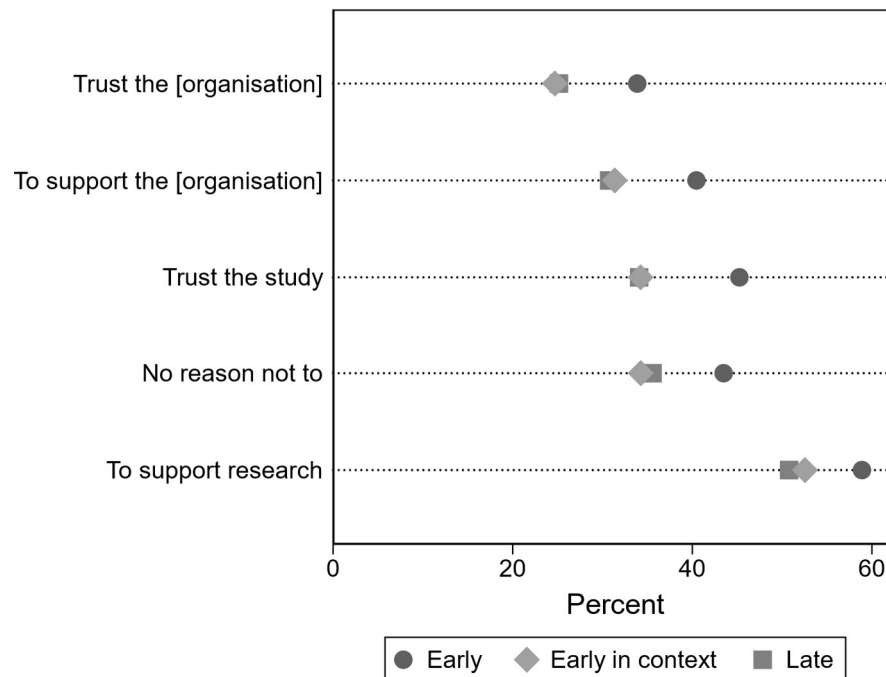
Figure 1: Predicted probabilities of each reason for not consenting, full sample



³ The reason for including all respondents in this analysis rather than only the non-consenters, is that the non-consenters are not necessarily comparable across the three placement treatment groups, since the non-consent rates differ between consent positions.

Figure 2 similarly represents the predicted probabilities of respondents in each placement group who gave each of the reasons for consenting. The most common reason given is to support research. For each of the five reasons for consenting, respondents in the early group were significantly more likely to report the reason than respondents in either of the other placement groups. In addition, there are no differences between the early in context and the late groups. The patterns are similar if the analysis is restricted to consenters only (see Appendix Figure D2).

Figure 2: Predicted probabilities of each reason for consenting, full sample



4. Discussion

We manipulated the placement and order of three consent requests to examine the effect on consent rates. We found that the placement of the request has a large effect on consent, with higher consent rates if asked early in the survey. This is consistent with several other recent studies (e.g. Eckman and Haas 2017; Jäckle et al. 2023; Sakshaug et al. 2019), suggesting this is a robust finding. We also found that the order in which these requests were made has no effect on consent rates. This is contrary to Beuthner et al. (2023) who asked seven consents across different domains, but consistent with Walzenbach et al. (2022) who asked five consents and did not find a consistent order effect. This suggests that an order effect may be found only when the number of requests is relatively large. Further replications are needed. We further found that the effect of placement can sometimes depend on the order, but this interaction is weak. Our final finding was, contrary to our expectation, putting the request in context – after a module about coronavirus vaccination – reduced consent rates compared to asking the questions slightly earlier in the survey, though this was confounded with being slightly later in the survey.

The most frequent reason chosen for not consenting was “too personal, shared enough information”. Our finding that this was selected as a reason for non-consent more frequently by those asked late in the survey points to some form of fatigue, as does the placement effect we find. More work is needed to understand the mechanism behind the consistent effect of placement on consent found in the literature.

The most common reason chosen for consenting was “to support research”. Our finding that trust – whether in the organization (the National Health Service) or the study (*Understanding Society*) – is selected more often as a reason for consenting by those who got the consent questions early suggests that the prevailing belief that trust is built up over the course of a survey (used as justification for asking questions like these at the end of the survey) is not empirically supported in our study. Again, understanding how trust evolves or changes over the course of a survey is an area ripe for further exploration.

References

- Bergmann, M., Franzese, F., and Schrank, F. (2022), "Determinants of Consent in the SHARE Accelerometer Study." SHARE Working Paper Series 78-2022, <https://doi.org/10.17617/2.3372782>.
- Beuthner, C., Weiß, B., Silber, H., Keusch, F. and Schröder, J., (2023), "Consent to Data Linkage for Different Data Domains—the Role of Question Order, Question Wording, and Incentives." *International Journal of Social Research Methodology*, published online first, <https://doi.org/10.1080/13645579.2023.2173847>.
- Dykema, J., DiLoreto, K., Croes, K.D., Garbarski, D., and Beach, J. (2017), "Factors Associated with Participation in the Collection of Saliva Samples by Mail in a Survey of Older Adults." *Public Opinion Quarterly*, 81 (1): 57-85, <https://doi.org/10.1093/poq/nfw045>
- Institute for Social and Economic Research (ISER) (2021), Understanding Society COVID-19 User Guide. Version 10.0, October 2021. Colchester: University of Essex.
<https://www.understandingsociety.ac.uk/documentation/covid-19/user-guide>
- Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F., and Walzenbach, S. (2023), "Survey Consent to Administrative Data Linkage: Wording and Format Experiments." Paper under review.
- Keusch, F., Struminskaya, B., Antoun, C., Couper, M.P., and Kreuter, F. (2019), "Willingness to Participate in Passive Mobile Data Collection." *Public Opinion Quarterly*, 83 (S1): 210-235, <https://doi.org/10.1093/poq/nfz007>.
- Kreuter, F., Haas, G.-C., Keusch, F., Bähr, S., and Trappmann, M. (2020), "Collecting Survey and Smartphone Sensor Data with an App: Opportunities and Challenges around Privacy and Informed Consent." *Social Science Computer Review*, 38 (5): 533-549, <https://doi.org/10.1177/0894439318816389>.
- Mneimneh, Z.N., McClain, C., Bruffaerts, R., and Altwaijri, Y.A. (2021), "Evaluating Survey Consent to Social Media Linkage in Three International Health Surveys." *Research in Social and Administrative Pharmacy*, 17 (6): 1091-1100, <https://doi.org/10.1016/j.sapharm.2020.08.007>.
- O'Doherty, K., Jaszczak, A., Hoffmann, J.N., You, H.M., Kern, D.W., Pagel, K., McPhillips, J., Schumm, L.P., Dale, W., Huang, E.S., and McClintock, M.K. (2014), "Survey Field Methods for Expanded Biospecimen and Biomeasure Collection in NSHAP Wave 2." *The Journals of Gerontology: Series B*, 69 (S2): S27-S37, <https://doi.org/10.1093/geronb/gbu045>.
- Sakshaug, J.W., Couper, M.P., and Ofstedal, M.B. (2010), "Characteristics of Physical Measurement Consent in a Population-Based Survey of Older Adults." *Medical Care*, 48 (1): 64-71, <https://doi.org/10.1097/MLR.0b013e3181adcdb3>.
- 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, <https://doi.org/10.1177/0049124112460381>.
- 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, <https://doi.org/10.1093/poq/nfz018>.
- 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,
<https://doi.org/10.1080/13645579.2014.899101>.

Sloan, L., Jessop, C., Al Baghal, T., and Williams, M. (2020), "Linking Survey and Twitter Data: Informed Consent, Disclosure, Security, and Archiving." *Journal of Empirical Research on Human Research Ethics*, 15 (1-2): 63-76, <https://doi.org/10.1177/1556264619853447>.

University of Essex, Institute for Social and Economic Research (2021), Understanding Society: COVID-19 Study, 2020-2021. [data collection]. 11th Edition. UK Data Service.
SN: 8644, 10.5255/UKDA-SN-8644-11.

Walzenbach, S., Burton, J., Couper, M.P., Crossley, T.F., and Jäckle, A. (2022), "Experiments on Multiple Requests for Consent to Data Linkage in Surveys." *Journal of Survey Statistics and Methodology*, published online first, <https://doi.org/10.1093/jssam/smab053>.

Appendix

A. Full wording of consent questions, and other questions used in analysis (follow up of non-consenters, objective understanding)

Health data linkage consent

To complement the information Understanding Society collects, we would like to find out more about your health and treatment from data held by the NHS (in the four countries of the UK). This would allow researchers to investigate how your health and wider circumstances interact. If you agree:

- We would like to link to information the NHS has about your health including:
 - data from hospital care records (including dates of admission and consultations, treatments received, and referrals made)
 - primary care records (including doctor and nurse consultations, diagnoses received, treatments given, and referrals made)
 - data on prescriptions
 - information on COVID-19 infection notification and test results
- We will send NHS data holders your personal identifiers (including name, address, sex and date of birth) so that they can identify the records they have about you.
- The NHS data holders will anonymise your records: they will contain an anonymous identification number but not your personal identifiers (name, address, sex, date of birth or NHS number).
- The anonymised NHS records will be added 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 NHS in any way.

Please read this [leaflet](#) and look at this [diagram](#) for further information.

Do you give permission for us to pass your personal identifiers (including name, address, postcode, sex and date of birth) to NHS data holders for this purpose?

1. Yes
2. No

(Link to leaflet: <https://www.understandingsociety.ac.uk/participants/health-records> and link to diagram: <https://www.understandingsociety.ac.uk/participants/data-linkage>)

Registry data linkage consent

We would also like to link to information about cancer registrations and the death registration records held by the General Registrars and Public Health bodies in each of the four countries of the UK.

Do you give permission for us to pass your personal identifiers (including name, address, postcode, sex and date of birth) to the NHS data holders for this purpose?

1. Yes
2. No

Serology consent

We would like to understand how many of our participants across the country have developed antibodies against the virus that causes COVID-19.

This is part of a national initiative where other research studies in the UK are also asking their participants to complete the same antibody test. Analysing the information from Understanding Society alongside these other studies will allow a greater understanding of the impact of COVID-19 on people's health and other aspects of life. You can take part even if you have recently been vaccinated.

- If you agree, we would like to send you a small kit which will allow you to take an antibody test.
- The test will involve pricking your finger with a small needle and collecting the blood in a small tube.
- You will then need to send the blood sample back in the pre-paid return envelope that will be included in the pack.
- The laboratory will test your blood for COVID-19 antibodies and, if you want it, you will receive a letter with your result within two weeks of posting your sample back to us.
- If you change your mind after you receive the kit, that is fine – you don't need to do anything.
- All the materials and full instructions will be included in the package.
- Your name and address will be used by Ipsos MORI to send you the kit, but no identifiable information will be sent to the laboratory.
- The results of your test will be added to your survey responses, and the data from the project will be available in **anonymised** format to other researchers, through the UK Data Service or other data repositories. **Anonymised** data and sample results may be shared with other bona fide scientists, including those from the Department of Health and Social Care (DHSC) and Public Health England (PHE).
- If you have certain health or medical conditions you will not be able to take part.

For more information, please click <here>.

To see a video of what is in the kit and what is involved, please click <here>.

May we send you the antibody testing kit through the post?

1. Yes
2. No
3. I have one of the health or medical conditions that means I cannot take part

(Link for additional information: www.understandingsociety.ac.uk/participants/serology)

Link to video: <https://www.youtube.com/watch?v=okTozcGMDIU>)

Why not consented to health data linkage

Can you tell us why you did **not** give us permission to add records collected by the National Health Service, or NHS, to the answers you have given in this study?

Please select all that apply.

1. Too personal, I've shared enough information with this survey
2. I am unclear about the risks involved / don't understand what would be done
3. I'm worried that my health records might be used against me
4. I'm worried that my health records might be lost / hacked / stolen
5. I was too tired / didn't want to make a decision
6. Unclear why I am being asked for this / what the data would be used for
7. The NHS doesn't have any data about me
8. Other reason

(Order of items 1-6 randomized to minimize response order effects.)

Why not consented to Registry data linkage

Can you tell us why you did **not** give us permission to add records held by the General Registrars and Public Health bodies to the answers you have given in this study?

Please select all that apply.

1. Too personal, I've shared enough information with this survey
2. I am unclear about the risks involved / don't understand what would be done
3. I'm worried that my records might be used against me
4. I'm worried that my records might be lost / hacked / stolen
5. I was too tired / didn't want to make a decision
6. Unclear why I am being asked for this / what the data would be used for
7. The General Registrars and Public Health bodies don't have any data about me
8. Other reason

(Order of items 1-6 randomized to minimize response order effects.)

Why not consented to serology

Can you tell us why you did **not** give permission for us to send you the COVID-19 antibody testing kit through the post?

Please select all that apply.

1. Too personal, I've shared enough information with this survey
2. I am unclear about the risks involved / don't understand what would be done
3. I'm worried that my COVID-19 test result might be used against me
4. I'm worried that my COVID-19 test result might be lost / hacked / stolen
5. I was too tired / didn't want to make a decision
6. Unclear why I am being asked for this / what the data would be used for
7. I don't want to know whether I have had COVID-19
8. I don't want to prick my finger
9. Takes too much time
10. Other reason

(Order of items 1-6 randomized to minimize response order effects. Order of items 7-9 also randomized)

Why consented to health data linkage

Can you tell us why you gave us permission to add records collected by the National Health Service, or NHS, to the answers you have given in this study?

Please select all that apply.

1. I trust the study not to mishandle information
2. I trust the NHS
3. To support the NHS
4. To support research for the greater good of society
5. I didn't see any reason not to / I have nothing to hide
6. Other reason

(Order of items 1-5 randomized to minimize response order effects.)

Why consented to Registry data linkage

Can you tell us why you gave us permission to add records held by the General Registrars and Public Health bodies to the answers you have given in this study?

Please select all that apply.

1. I trust the study not to mishandle information
2. I trust the General Registrars and Public Health bodies
3. To support the General Registrars and Public Health bodies
4. To support research for the greater good of society
5. I didn't see any reason not to / I have nothing to hide
6. Other reason

(Order of items 1-5 randomized to minimize response order effects.)

Why consented to serology

Can you tell us why you gave permission for us to send you the COVID-19 antibody testing kit through the post?

Please select all that apply.

1. I trust the study not to mishandle information
2. I trust the NHS
3. To support the NHS
4. To support research for the greater good of society
5. I didn't see any reason not to / I have nothing to hide
6. I wanted to know if I have had COVID-19
7. Other reason

(Order of items 1-5 randomized to minimize response order effects.)

Objective understanding of health data linkage

To help us understand whether the explanation we gave you about linking NHS 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.

Please select one answer per row.

The NHS will be able to see the answers you have given in this study. [FALSE]

The NHS data holder 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 the NHS data holder. [TRUE]

B. Background tables

Table B1: Cell sizes for Table 1

	Early	Early, in context	Late	Total
NHS data	3,586	3,637	3,534	10,757
Register data	3,580	3,636	3,535	10,751
Send serology kit	3,487	3,550	3,448	10,485
All three	3,473	3,548	3,447	10,468

Table B2: Cell sizes for Table 2

	Health asked first	Serology asked first	Total
NHS data	5,489	5,268	10,757
Register data	5,485	5,266	10,751
Send serology kit	5,348	5,137	10,485
All three	5,347	5,121	10,468

C. Full tables of results

Table C1: Break off rates by consent position

	N	Break off before closing module (%)	Break off in consent module (%)
Early	3,618	4.1 ^{AB}	2.6 ^{AB}
Early in context	3,651	2.7 ^A	1.0 ^{AC}
Late	3,609	2.1 ^B	0.3 ^{BC}
Total	10,878	3.0	1.3
		Pearson $\chi^2(2) = 24.61, P < 0.001$	Pearson $\chi^2(2) = 78.40, P < 0.001$

Notes: Break off in the consent module includes breakoffs in the health consent module, the serology consent module, or the module of consent follow up questions. ^{A B C} denote break off rates that are significantly different from each other in pairwise Pearson χ^2 tests ($P \leq 0.001$).

Table C2: Predicted probabilities of reasons for not consenting, by consent position

Reason for not consenting	Consent position	Predicted Probability	S.E.	95% C.I.	
Worried records used against me	Early	1.61	0.17	1.27	1.95
	Early in context	1.82	0.18	1.47	2.18
	Late	1.73	0.17	1.39	2.06
Too tired	Early	0.89	0.13	0.64	1.14 ^A
	Early in context	0.98	0.13	0.74	1.23 ^B
	Late	1.77	0.18	1.42	2.12 ^{AB}
Unclear about risks involved	Early	2.43	0.20	2.04	2.82 ^A
	Early in context	2.93	0.22	2.50	3.37
	Late	3.68	0.25	3.19	4.17 ^A
Worried records might be lost	Early	6.29	0.33	5.63	6.94
	Early in context	6.77	0.35	6.10	7.45
	Late	6.06	0.32	5.43	6.69
Unclear why	Early	3.88	0.26	3.38	4.39 ^{AB}
	Early in context	4.96	0.28	4.40	5.51 ^A
	Late	6.11	0.31	5.49	6.72 ^B
Too personal, shared enough	Early	9.85	0.42	9.03	10.67 ^A
	Early in context	11.15	0.44	10.28	12.02 ^B
	Late	13.40	0.47	12.47	14.33 ^{AB}

Notes: Predicted probabilities calculated from six separate logit models. N = 31,993 observations with valid answers to the consent request and reason for consenting or not consenting. Standard errors are adjusted for clustering of consent requests in respondents. For the first model, the dependent variable is the probability that the respondent stated “Worried records used against me” as a reason for not consenting. For the second model the dependent variable is the probability that the respondent stated “Too tired” as a reason, etc. The independent variables in all models are the consent domain (health data, registry data, serology test kit) and the position of the consent request (early, early in context, late). ^A ^B denote pairs of Confidence Intervals that do not overlap.

Table C3: Predicted probabilities of reason for consenting, by consent position

Reason for consenting	Consent position	Predicted Probability	S.E.	95% C.I.	
Trust the study not to mishandle information	Early	45.25	0.77	43.75	46.75 ^{AB}
	Early in context	34.24	0.71	32.84	35.64 ^A
	Late	34.07	0.73	32.64	35.50 ^B
Trust the NHS / General Registrars and Public Health bodies	Early	33.87	0.71	32.48	35.26 ^{AB}
	Early in context	24.70	0.64	23.44	25.95 ^A
	Late	25.14	0.66	23.86	26.43 ^B
To support the NHS / General Registrars and Public Health bodies	Early	40.46	0.72	39.04	41.87 ^{AB}
	Early in context	31.37	0.67	30.05	32.68 ^A
	Late	30.75	0.68	29.42	32.08 ^B
To support research for the greater good of society	Early	58.89	0.73	57.45	60.32 ^{AB}
	Early in context	52.54	0.73	51.12	53.97 ^A
	Late	50.78	0.74	49.33	52.22 ^B
Didn't see any reason not to / nothing to hide	Early	43.49	0.75	42.02	44.95 ^{AB}
	Early in context	34.26	0.70	32.88	35.63 ^A
	Late	35.56	0.72	34.15	36.98 ^B

Notes: Predicted probabilities calculated from six separate logit models. N = 31,993 observations with valid answers to the consent request and reason for consenting or not consenting. Standard errors are adjusted for clustering of consent requests in respondents. The independent variables in all models are the consent domain (health data, registry data, serology test kit) and the position of the consent request (early, early in context, late). ^{A B} denote pairs of Confidence Intervals that do not overlap.

D: Additional Figures

Figure D1: Predicted probabilities that respondent gave each of the reasons for not consenting, non-consenters only

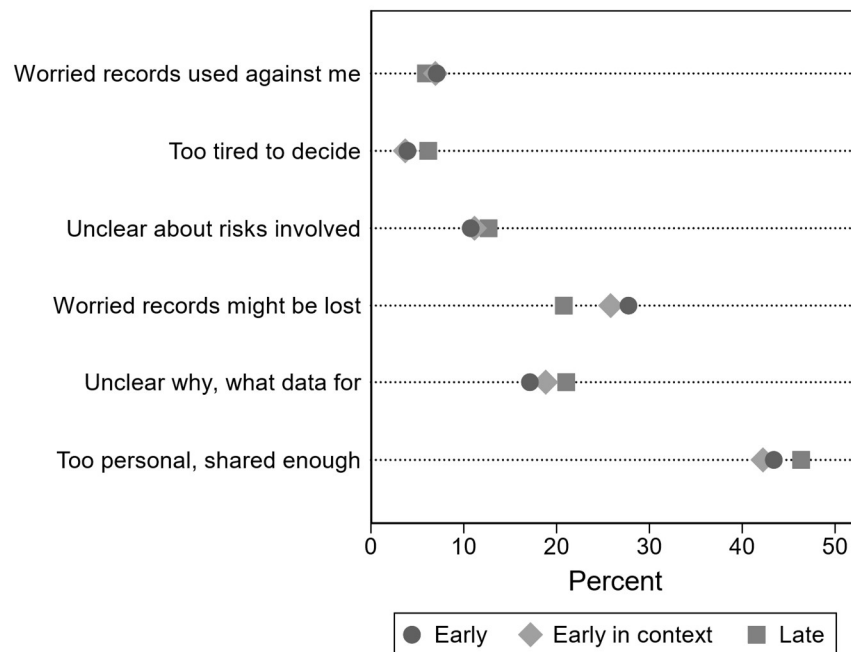


Figure D2: Predicted probabilities that respondent gave each of the reasons for consenting, consenters only

