Understanding Society

THE UK HOUSEHOLD LONGITUDINAL STUDY

How to ask for consent to data linkage: Things we've learnt

Annette Jäckle (University of Essex)

JPSM MPSDS Seminar (15/03/2023)

University of Essex

ISER

UKRI Economic and Social Research Council

An initiative by the Economic and Social Research Council, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public

Acknowledgements

- Collaborators

  Jonathan Burton (University of Essex)

  Mick P. Couper (University of Michigan)

  Thomas F. Crossley (European University Institute)

  Sandra Walzenbach (University of Konstanz)

- Funders

  Nuffield Foundation

  UKRI Economic and Social Research Council

Context

Understanding Society panel study

- UK Household Longitudinal Study
  - Similar to PSID
- All adults aged 16+ interviewed
  - Annually since 2009
- Modes of data collection
  - Web, CAPI, CATI
- Innovation Panel (IP)
  - Same design as main study
  - 1,500 respondent households
  - Fielded as separate survey
  - Experimental methods testing

IP competition for 2024 survey:
https://www.understandingsociety.ac.uk/innovation-panel-competition
Deadline: 16 April 2023

Data linkage
In Understanding Society

- Consents to data linkage

  Government administrative data: health, education, taxes, State benefits and pensions, household energy use

  Banking transactions: credit data

  Social media: Twitter, LinkedIn

- GDPR: legal basis for processing personal data

  'Public Task'

  Consent: for ethical reasons

- Consent rates web < FTF

  About -30 percentage points (Jäckle et al 2021)

  Similar in other studies

Why?
What can we do?

What we knew about consent
Key findings

- Correlates of consent inconsistent
  - Between studies / within studies over time (e.g. Peycheva et al 2021)
  - Respondent / interviewer characteristics (e.g. Sala et al 2012)
- Large variation in consent rates between
  - Data to be linked / topics (e.g. Sakshaug et al 2012)
  - Interviewers (e.g. Korbmacher & Schröder 2013)
  - Modes of interview (Thornby et al 2018)
- Little effect of question wording experiments, e.g.
  - Length of consent question text (e.g. Edwards & Biddle 2021, Singer & Frankel 1982)
  - Emphasising benefits to respondent / survey (e.g. Pascale 2011)
  - Loss framing – inconsistent results (e.g. Kreuter et al 2016, Sakshaug et al 2019)

What we knew about consent
Key findings

- Many respondents do not understand request (Das & Couper 2014)
- Asking earlier in interview (in context) increases consent (e.g. Sala et al 2014, Sakshaug et al 2013)

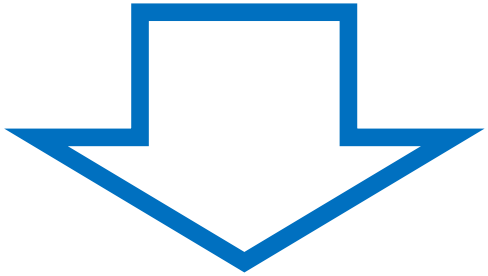
  → **Why?**
- Half of non-consenters say 'yes' if asked again (Weir et al 2014, Mostafa & Wiggins 2018)

  → **Decision is not fixed, can be influenced**
- Multiple consents in one interview: latent willingness to consent (Jenkins et al 2006; Mostafa 2016)
- But only weak latent willingness over time (Mostafa & Wiggins 2018)

  → **Situational factors important**

Aims

How do respondents decide whether to consent?



What can we do to

- **Reduce barriers to informed consent?**
- **Especially in web?**

Methods

- Qualitative in-depth interviews with IP sample members
  - What determines whether respondents consent?
- Conceptual framework
  - How respondents make the consent decision
  - Hypotheses
- Experimental testing

Conceptual framework
Based on

- Qualitative interviews with IP respondents

  Factors that influence consent decision (Beninger et al 2017)
- Cognitive model of survey response process

  How Rs answer survey questions (Cannell et al 1981, Tourangeau et al 2000)
- Survey methods literature

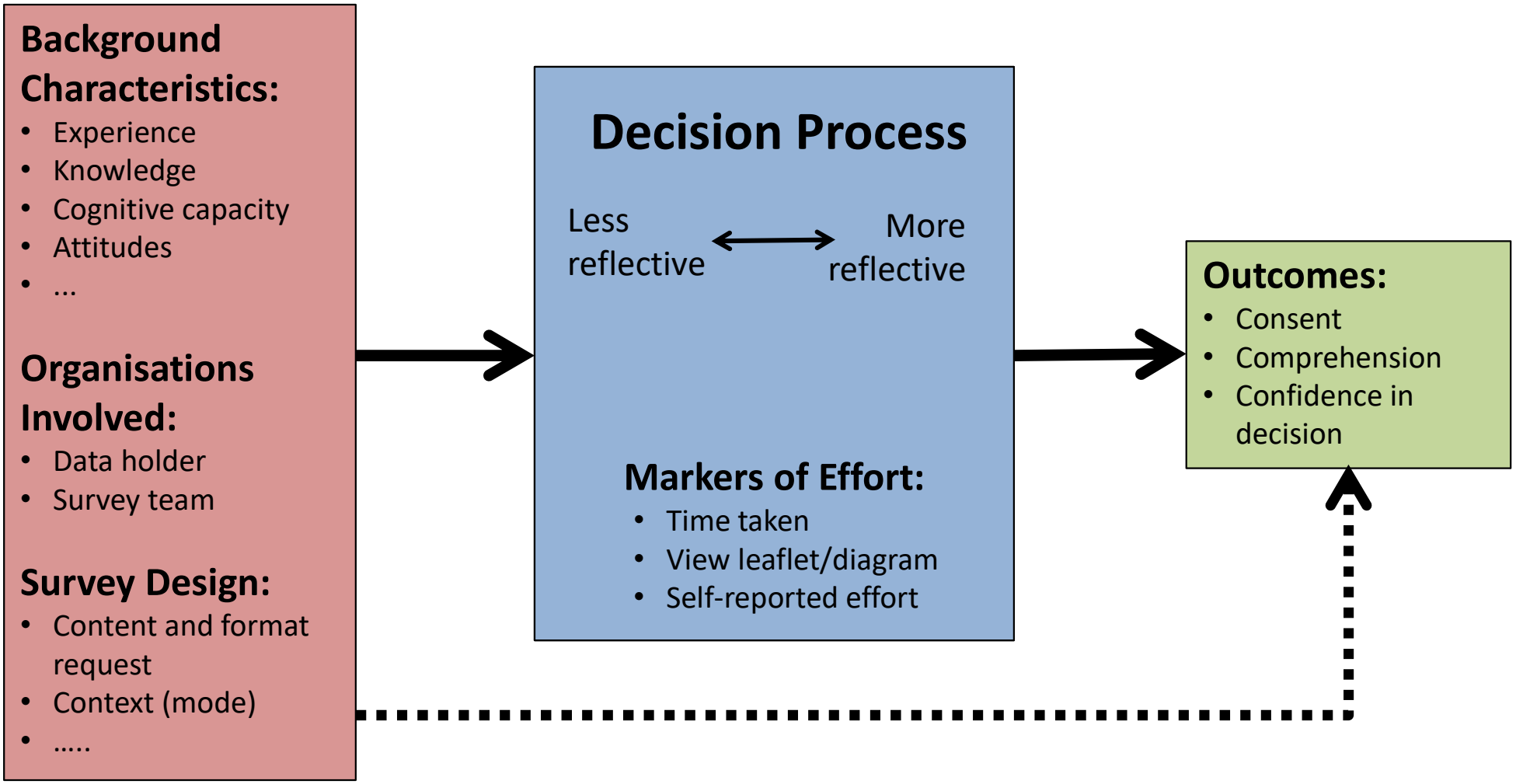
  Consent to data linkage, experiments
- Rational vs heuristic decision making

  System 1 vs system 2 processing (Petty and Cacioppo 1986, Kahneman 2011)
- Real-life decision making

  People reduce amount of information considered (Galotti 2007)

Conceptual framework

How respondents decide whether to consent



Burton et al (2021)

Today

Synthesis of research on these topics

(1) How do respondents decide whether to consent to data linkage?

(2) Why are respondents less likely to consent in web than CAPI surveys?

(3) How best to ask for multiple consents within a survey?

(4) Which wording and formats affect informed consent and why?

Experimental testing

General research design

Experimental testing
Samples

- Innovation Panel (IP)

  General population, Great Britain

  2018

Web	1,299
CAPI	1,363

- PopulusLive online access panel

  Quotas to match IP sample, Great Britain

  2018, 2019

Sample 1 wave 1	5,684
wave 2	1,634
Sample 2	3,850

- Understanding Society COVID-19 survey

  General Population, United Kingdom

  2021

Web	11,802

Experimental testing

Questionnaire content (not all used in all tests)

- Background questions, e.g.

  Socio-demos; attitudes to privacy, data security; data sharing; emotional state

- Consent request – either

  Single: Tax records OR health data

  Multiple: Tax, education, health, benefits & pensions, energy usage

- Follow-up questions

Self-reported decision process Subjective understanding of request Objective understanding (test Qs) Confidence in consent decision	Self-reported effort made Factors taken into account Reasons for consenting / or not

- Paradata: response times, clicks on links, interviewer observations
- Audio recordings (CAPI): behaviour coding

Experimental testing

Consent question treatments

Single consents:

- Readability
- Placement
- Trust priming
- Wording
- Mode of interview

Multiple consents:

- Order
- Format (single/joint)

1. How do respondents decide whether to consent?

- Consent to link to tax records (HMRC)

  Standard question text used in *Understanding Society*

  Question explains what, why, how

  "….Do you give permission for us to pass your name, address, sex and date of birth to HMRC for this purpose? (Yes/No)"

- Analysis replicated

  Innovation Panel: Web and CAPI

  Access Panel 3x

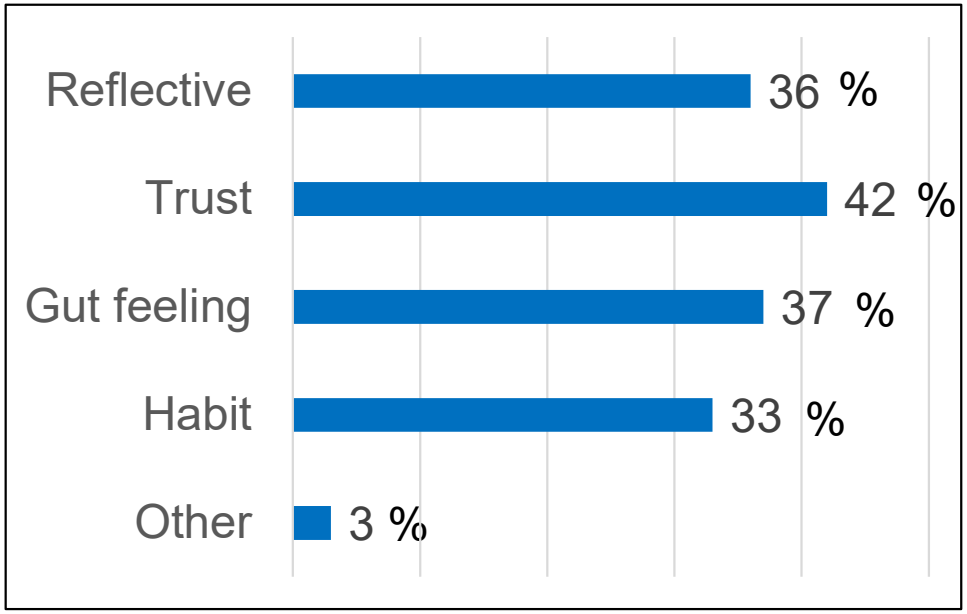
Burton et al (2021)

Consent decision process

How did you decide whether to say “yes” or “no” in response to the question about data linkage?

Please select all that apply

1. I thought about what would happen if I said “yes” or “no” **(reflective)**
2. Instinct or gut feeling **(gut)**
3. I said what I usually say when I’m asked for information that is very personal **(habit)**
4. I thought about how much I trust the organisations involved **(trust)**
5. Something else



Source: Access Panel 1.2

Only 1/3 respondents make a reflective decision

Self-reported decision process is corroborated by other indicators

- Decision based on ‘habit’ or ‘gut feeling’ vs. ‘reflective’

  Consent question answered more quickly (paradata)

  Less likely to click on links to leaflet or diagram explaining linkage (paradata)

  Lower self-reported effort to answer consent question (scale 0-10)

  Less likely to say the considered ‘information’ in making their decision, e.g.:

  What information the government has about me

  Less likely to say they considered ‘risks’, e.g.:

  How much I trust the organisations involved

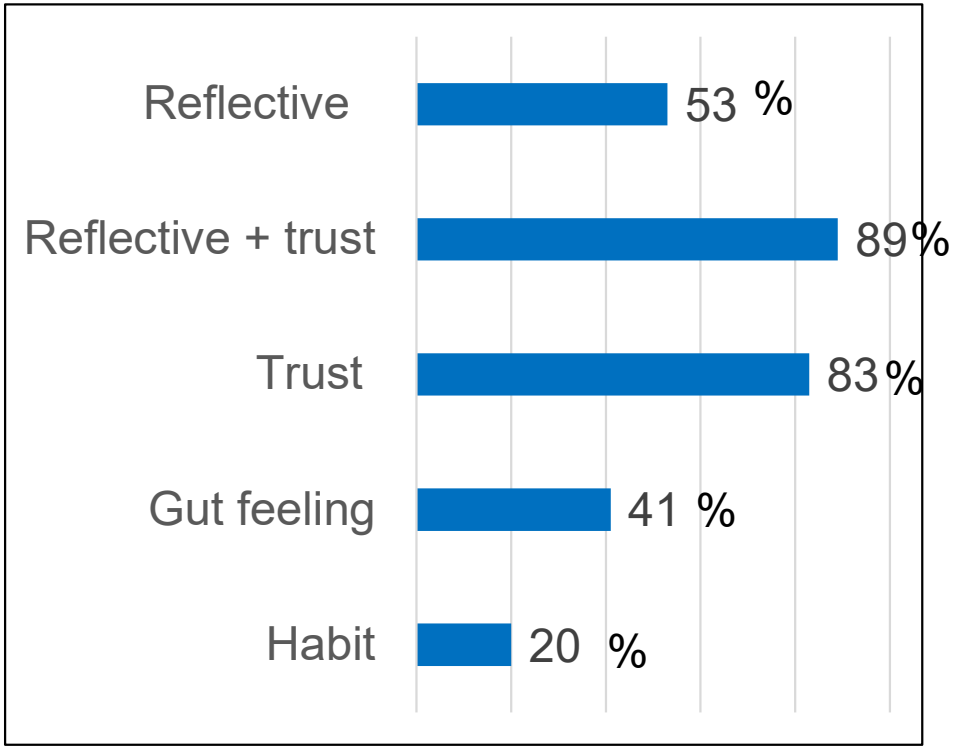
  Less likely to say they considered ‘benefits’:

  The benefits to society

Self-reported decision process measures genuine differences between respondents

Consent rates by decision process

- Respondents making habit-based decisions
  - Least likely to consent
  - Over time: most likely to report same decision process and same consent decision



Source: Access Panel 1.2

Decision process relates to consent outcome
But not necessarily causal effect

In sum…

- Respondents use different decision processes
  - Majority: gut feeling or habit
  - Minority: reflective
- Decision process related to consent
  - But not necessarily causal

Burton et al (2021)

2. Why are respondents less likely to consent in web than FTF?

Difference between modes in

1. Consent rate
2. Understanding of linkage request

Potential mechanisms

3. Respondent attitudes
4. Consent decision process
5. Device used to complete web survey
6. Interviewer behaviours

Jäckle et al (2022a)

Data

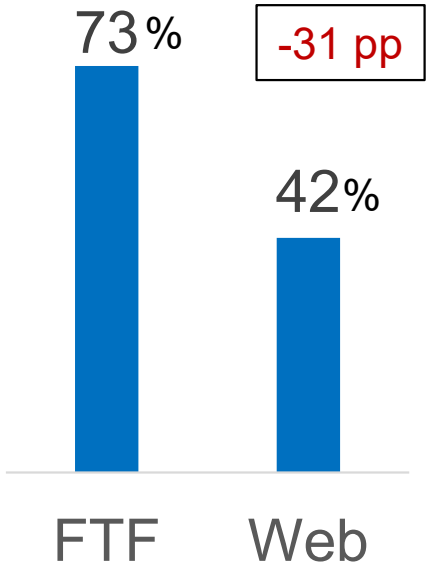
- Innovation Panel
- Mixed mode data collection

	Mode of interview		
Random allocation	FTF (%)	Web (%)	Total
FTF-first	93.5	6.5	1,032
Web-first	22.3	77.7	1,576

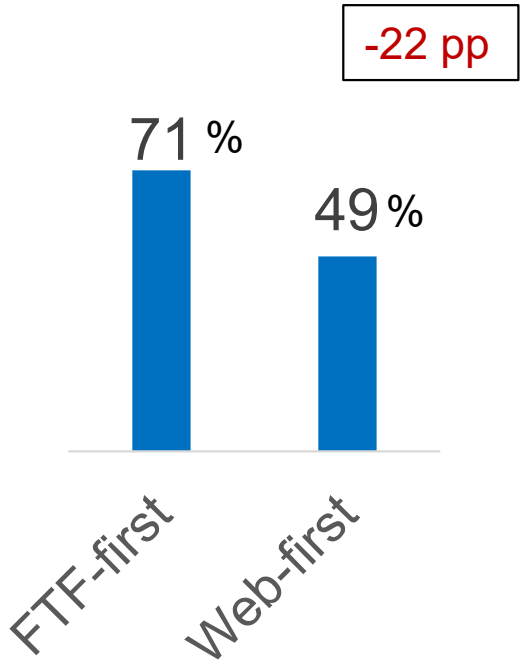
1. Difference between modes:
Consent rate

Linkage to tax records (HMRC)

Mode of interview
= as treated



Random allocation
= intention to treat (ITT)



Instrumental variable
= local average t. effect

-30 pp

ITT estimate
Proportion of web compliers

Effect of mode on answers, not selection

2. Difference between modes:
Understanding of linkage request

- FTF: questions asked in CASI
- Subjective understanding:

  "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?"

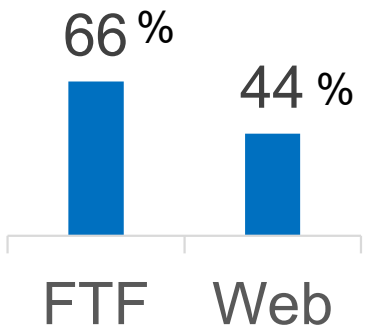
- Objective understanding:

  8 true/false statements

- Confidence in consent decision:

  "…. How confident are you about the decision you made?"

Completely/mostly understood:



Web: lower test scores

No difference

Respondents understand less when answering online

3. Potential mechanisms:
Respondent attitudes

- Privacy: very/somewhat concerned
  +8 pp in web
- Data security: very/somewhat concerned
  +5 pp in web

Web treatment effect (IV)

- HMRC tax data are (highly) sensitive
- Trust in survey organisation, university
- Trust in HMRC

No difference between modes

Respondents more concerned about privacy / data security when answering online

4. Potential mechanisms:
Consent decision process

- Decision process

  "How did you decide whether to say "yes" or "no" in response to the question about data linkage?" *(Select all)*

  1. I thought about what would happen if I said "yes" or "no" **(reflective)**
  2. Instinct or gut feeling **(gut)**
  3. I said what I usually say when I'm asked for information that is very personal **(habit)**
  4. Something else

- Web treatment effect (IV)

  Reflective decision: -9 pp

  Habit-based decision: +12 pp

  Response time: 2.2 x faster

  Read/clicked leaflet: -32 pp

  Information presented was "too much": +6 pp

Respondents answer consent Qs less thoroughly when answering online

5. Potential mechanisms:
Device used to complete web survey

- Devices used by web respondents:
  - 57% PC, laptop, notebook
  - 29% tablet
  - 14% smartphone
- Logit models to control for respondent characteristics

- No differences in
  - Consent
  - Objective understanding
  - Privacy / data security concerns
  - Consent decision process
  - Whether clicked leaflet/diagrams

Devices not driving mode effects

6. Potential mechanisms:
Interviewer behaviours

- Audio-recordings

  FTF interviews
- Coded what interviewers & respondents did

  Consent question

- Interviewers – rarely

  Emphasized confidentiality (4% cases)

  Offered additional information (14%)
- Respondents – rarely

  Expressed concern/uncertainty (5%)

  Asked questions (16%)

Interviewer behaviours not driving mode differences

In sum…

- Survey mode has a causal effect on
  - Probability of consent
  - Decision process
  - Understanding of consent request
  - Concerns about privacy and data security
- Audible interviewer behaviours do not explain mode effect

Jäckle et al (2022a)

3. Some question wording experiments

- Multiple consents
- Easier wording
- Trust priming
- Early / late placement in the questionnaire

Questionnaire design:
How best to ask multiple consents?

- Access panel
  - 5 consent questions
- Q format:
  - Single question per page
  - All questions on one page
- Q order:
  - Start with education
  - Start with taxes

No effect

Carry-over effects of starting with high consent Q?
But not replicated

- COVID-19 study
- Q order:
  - Health/register → test kit
  - Test kit → Health/register

No effect on consent

Order effects? See also Beuthner et al (2023)

❞ Walzenbach et al (2022)

❞ Burton et al (2023)

Questionnaire design:
Easier wording of consent request

- Innovation Panel & Access Panel
- Standard wording
  - UKHLS consent question (tax records)
  - Reading difficulty: Flesch-Kincaid Grade level 14.5
- Easy wording
  - Lower reading difficulty: 8.2
  - Shorter sentences & words, no passive voice
  - Results from qualitative interviews
  - Text broken up into bullet points

Increasing understanding did not increase consent
Easier wording did not reduce mode effect on consent

Jäckle et al (2022a, 2022b)

Questionnaire design:
Trust priming

- Access panel
- Introduction to consent question:

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

  HMRC is a trusted data holder

  ← Treatment

  1 Continue

- "Easy" tax data consent question

Jäckle et al (2022b)

Questionnaire design:
Trust priming

- Consent rate:

  +5 pp

Relatively large effect
Equivalent to 1/6th of mode gap

- No effects on

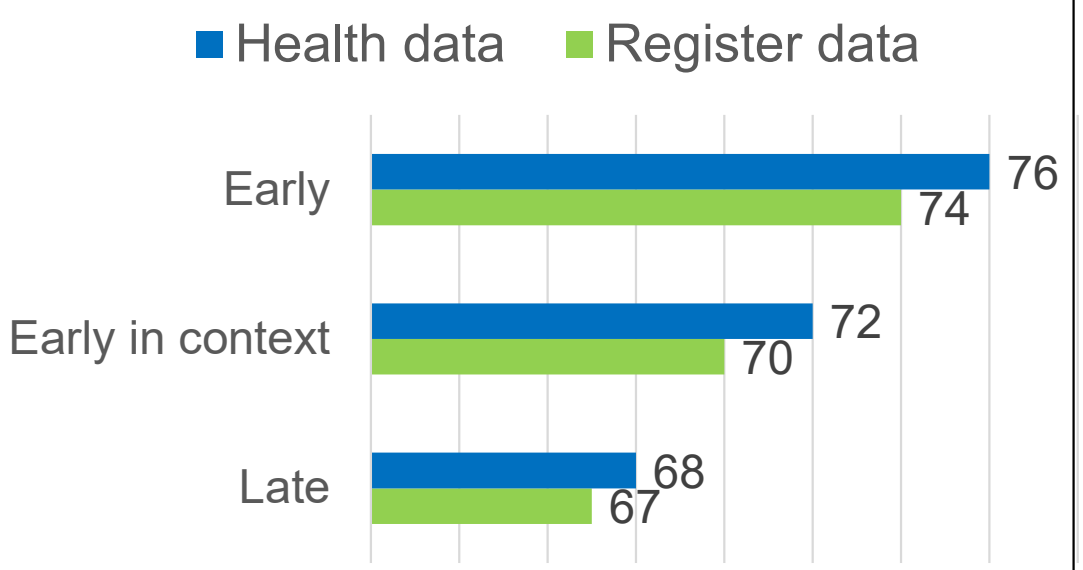
Understanding	Response time
Objective understanding	Whether clicks on leaflet/diagram
Confidence in decision	Trust in HMRC
Decision process	Factors respondents considered

Trust is important
Survey design / setting can affect trust
Sub-conscious?

Questionnaire design:

Early/late placement in the questionnaire

- COVID-19 study (Web)
- Consent rates (%):



- Innovation Panel
  - Similar results
  - But only with ‘standard’ question wording
  - No effect with ‘easy’ wording

Burton et al (2023)

Jäckle et al (2022b)

Questionnaire design:
Early/late placement in the questionnaire

- COVID-19 study
- Follow-up question for each consent

  Health

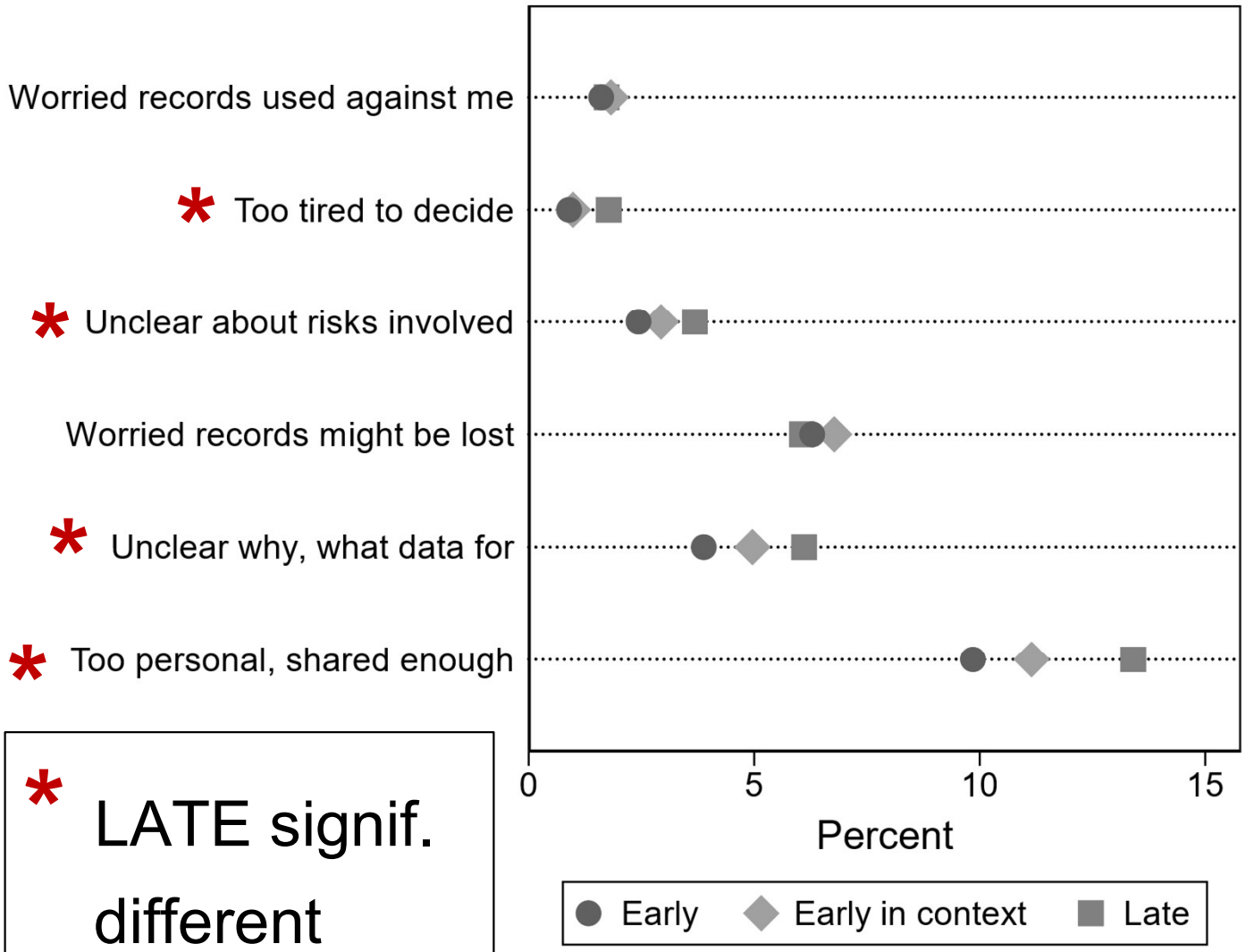
  Cancer & death register

  Antibody test kit

*Can you tell us why you did **not** give us permission to add records collected by the <data holder> to the answers you have given in this study?*

Burton et al (2023)

Questionnaire design:
Early/late placement in the questionnaire



- Cognitive fatigue?
  - But no difference in objective understanding by placement
  - Same in
    - ❞ Jäckle et al (2022b)

* LATE signif. different

“Giving fatigue”?

❞ Burton et al (2023)

Conclusions

- Examine how respondents decide whether to consent
- Most do not make a fully reflective decision

  Instead use heuristics, e.g.

  Trust

  Gut feeling

  Habit

- Many do not read additional information materials

Conclusions ctd.

- When Rs complete survey online rather than FTF
  - Much less willing to consent
  - Understand the linkage request less well
- Why is this?
  - Rs more concerned about privacy / data security when answering online
  - Rs process consent requests less carefully when answering online
  - Not due to devices used to complete web survey
  - Not due to interviewer verbal behaviours
- What can we do?
  - Providing additional information for web respondents does not help
  - Improving understanding of request does not help
  - Trust priming increases consent
  - Not tested: Social pressure with interviewer? Undecided more likely to agree?

Practical implications:
How best to ask for consent?

Ask for consent early in the interview, rather than at the end[3,6]

- 'Giving fatigue' more likely if ask at the end

Emphasize trust in the organisations involved in the data linkage [6]

Include all key information about the linkage in the consent question [5]

- Respondents unlikely to read information leaflets/diagrams

When asking multiple consent questions, start with the one that gets highest consent rate [7]

Simplify readability of consent question wording [6]

- e.g. using Flesh-Kinkaid readability score implemented in MS Word

Multiple consent questions can be shown separately or on a single page [7]

Outlook

- What is it interviewers do to increase consent?

  Experimental comparison of CAPI, CASI, web

- What do respondents want to know about the linkage?

  Information about linkage process / benefits and risks / value for science…?

- Is there a causal effect of the decision process on consent?

  Experimental manipulation of decision process

- Use consultation and opt out instead of individual consent?

  Who would respondents trust to review linkages? How to present information to respondents? Effect on survey participation?

- More generally:

How do respondents decide whether to do additional tasks for a survey?

References – Our Work

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

[2] ** Burton, J., Couper, M.P., Crossley, T.F., Jäckle, A. and Walzenbach, S. (2021) How do survey respondents decide whether to consent to data linkage?, *Understanding Society Working Paper 2021-05*, Colchester: University of Essex. https://www.understandingsociety.ac.uk/research/publications/547074

[3] Burton, J., Couper, M.P. and Jäckle, A. (2023) The effects of position, context, and order on consent to data linkage in a survey, *Understanding Society Working Paper 2023-07*, Colchester: University of Essex.

[4] Jäckle, A., Beninger, K., Burton, J. and Couper, M.P. (2021) Understanding data linkage consent in longitudinal surveys. Ch. 6 in P. Lynn (ed.) *Advances in Longitudinal Survey Methodology*. Chichester: Wiley. https://doi.org/10.1002/9781119376965.ch6.

[5] Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F. and Walzenbach, S. (2022a) How and why does the mode of data collection affect consent to data linkage?, *Survey Research Methods,* 16(3):387–408. https://doi.org/10.18148/srm/2022.v16i3.7933.

[6] ** Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F. and Walzenbach, S. (2022b) Consent to data linkage: Question wording and format experiments, *Understanding Society Working Paper 2022-05*, Colchester: University of Essex. https://www.understandingsociety.ac.uk/research/publications/547321.

[7] 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.* https://doi.org/10.1093/jssam/smab053.

**** Email aejack@essex.ac.uk for updated versions**

References – Literature Review I

Beuthner, Weiß, Silber, Keusch, and Schröder (2023) Consent to data linkage for different data domains–the role of question order, question wording, and incentives." *International Journal of Social Research Methodology,* preprint https://doi.org/10.1080/13645579.2023.2173847.

Cannell, Miller, and Oksenberg (1981) Research on interviewing techniques, *Sociological Methodology*, 12:389-437.

Das and Couper (2014) Optimizing opt-out consent for record linkage, *Journal of Official Statistics*, 30(3):479-497.

Edwards and Biddle (2021) Consent to data linkage: Experimental evidence on the impact of data linkage requests and understanding and risk perceptions. Ch. 8 in Advances in Longitudinal Survey Methodology, edited by P. Lynn. Chichester: Wiley.

Galotti (2007) Decision structuring in important real-life choices, *Psychological Science*, 18(4):320-325.

Jenkins, Cappellari, Lynn, Jäckle, and Sala (2006) Patterns of consent: Evidence from a general household survey, *Journal of the Royal Statistical Society Series A,* 169(4):701-722.

Kahneman (2011) *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Korbmacher and Schröder (2013) Consent when linking survey data with administrative records: The role of the interviewer, *Survey Research Methods,* 7(2):115-31.

Kreuter, Sakshaug, and Tourangeau (2016) The framing of the record linkage consent question, *International Journal of Public Opinion Research,* 28(1):142-52.

Mostafa (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.

Mostafa and Wiggins (2018) What influences respondents to behave consistently when asked to consent to health record linkage on repeat occasions? *International Journal of Social Research Methodology*, 21(1):119-134.

References – Literature Review II

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

Petty and Cacioppo (1986) The elaboration likelihood model of persuasion, *Advances in Experimental Social Psychology*, 19:123-205.

Peycheva, Ploubidis, and Calderwood (2021) Determinants of consent to administrative records linkage in longitudinal surveys: Evidence from Next Steps. Ch. 7 *in Advances in Longitudinal Survey Methodology,* edited by P. Lynn. Chichester: Wiley.

Sakshaug, Couper, Ofstedal, and Weir (2012) Linking survey and administrative records: Mechanisms of consent. *Sociological Methods & Research,* 41(4):535-69.

Sakshaug, Schmucker, Kreuter, Couper, and Singer (2019) The effect of framing and placement on linkage consent, *Public Opinion Quarterly*, 83(S1):289-308.

Sakshaug, Tutz, and Kreuter (2013) Placement, wording, and interviewers: Identifying correlates of consent to link survey and administrative data, *Survey Research Methods*, 7(2):133-44.

Sala, Burton, and Knies (2012) Correlates of obtaining informed consent to data linkage: Respondent, interview and interviewer characteristics, *Sociological Methods and Research*, 41(3):414-39.

Sala, Knies, and Burton (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.

Singer and Frankel (1982) Informed consent procedures in telephone interviews, *American Sociological Review*, 47(3):416–426.

References – Literature Review III

Thornby, Calderwood, Kotecha, Beninger, and Gaia (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.

Tourangeau, Rips, and Rasinski (2000) *The psychology of survey response*, Cambridge, England: Cambridge University Press.

Weir, Faul, and Ofstedal (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.

Thank you for listening

Annette Jäckle

University of Essex

aejack@essex.ac.uk