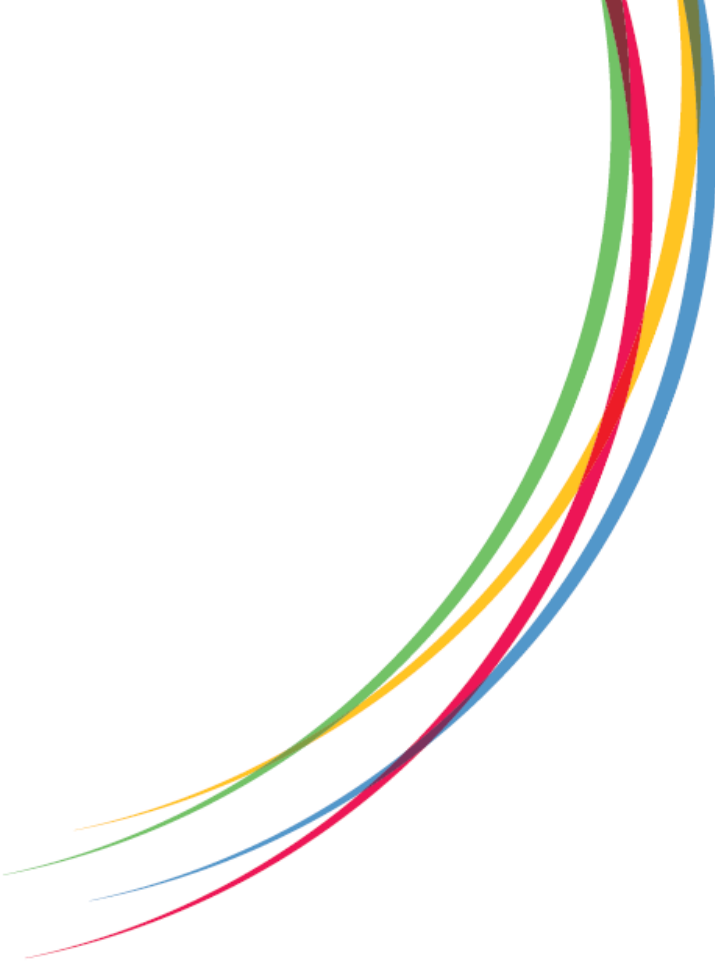




**Understanding Society
Working Paper Series
No. 2023 – 10
November 2023**

**Understanding Society Innovation Panel Wave 15:
Results from Methodological Experiments**

Jim Vine¹, Tarek Al Baghal¹, Jonathan Burton¹, Mick P. Couper², Thomas Crossley³,
Paul Fisher¹, Karon Gush¹, Daniel Horn¹, Ivelina Hristova⁴, Annette Jäckle¹, Curtis
Jessop⁵, Meena Kumari¹, Peter Levell⁶, Shujun Liu⁷, Hamish Low⁸, Sarah Parbury¹,
Paulo Serodio¹, Luke Sloan⁷, Sandra Walzenbach⁹

¹ Institute for Social and Economic Research, University of Essex, ² University of Michigan, ³ European
University Institute, ⁴ London School of Economics and Political Science, ⁵ NatCen Social Research,
⁶ Institute for Fiscal Studies, ⁷ Cardiff University, ⁸ University of Oxford, ⁹ University of Konstanz.



Non-technical summary

The *Understanding Society* survey includes what is known as an ‘Innovation Panel’ sample (IP). This sample of originally 1,500 households is used to test different methods for conducting longitudinal surveys in order to produce the highest quality data. The results from the Innovation Panel provide evidence about the best way to conduct a longitudinal survey which is of relevance for all survey practitioners as well as influencing decisions made about how to conduct *Understanding Society*. This paper reports the experiments and methodological tests carried out at wave 15 of the Innovation Panel in the summer of 2022.

IP15 employed a mixed-mode design; whilst IP13 and IP14 exclusively used web-first and telephone approaches due to COVID-19, IP15 included a return to some respondents being interviewed face-to-face. IP15 also continued ongoing experiments on the impact of incentives, and, as with prior waves, several other methodological experiments were included in the survey. Experiments were conducted on: how to ask respondents for consent to link data from their Twitter accounts to their survey responses; how to ask how much of a monetary windfall people would spend; how to find out whether respondents provide informal care; how to find out how much alcohol people drink; how the presence of an interviewer affects respondents’ likelihood of providing consent to data linkage; how to ask respondents for their mobile contact details and their consent to send questions by SMS; how to invite respondents to participate in additional data collection tasks conducted via mobile apps; different ways of collecting hip and waist measurements; and how to ask respondents about their national identity.

Understanding Society Innovation Panel Wave 15: Results from Methodological Experiments

Jim Vine¹, Tarek Al Baghal¹, Jonathan Burton¹, Mick P. Couper², Thomas Crossley³, Paul Fisher¹, Karon Gush¹, Daniel Horn¹, Ivelina Hristova⁴, Annette Jäckle¹, Curtis Jessop⁵, Meena Kumari¹, Peter Levell⁶, Shujun Liu⁷, Hamish Low⁸, Sarah Parbury¹, Paulo Serodio¹, Luke Sloan⁷, Sandra Walzenbach⁹

¹ Institute for Social and Economic Research, University of Essex, ² University of Michigan, ³ European University Institute, ⁴ London School of Economics and Political Science, ⁵ NatCen Social Research, ⁶ Institute for Fiscal Studies, ⁷ Cardiff University, ⁸ University of Oxford, ⁹ University of Konstanz

Abstract: This paper presents some preliminary findings from Wave 15 of the Innovation Panel (IP15) of *Understanding Society*: The UK Household Longitudinal Study. *Understanding Society* is a major panel survey in the UK. In June 2022, the fifteenth wave of the Innovation Panel went into the field. IP15 used a mixed-mode design, using on-line interviews, face-to-face interviews, and telephone interviews. This paper describes the design of IP15, the experiments carried and the preliminary findings from early analyses of the data.

Keywords: longitudinal, survey methodology, experimental design, question wording, questionnaire design.

JEL classification: C80, C81, C83

Acknowledgements: *Understanding Society* is an initiative funded by the Economic and Social Research Council (ESRC) and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. Wave 15 of the Innovation Panel was funded by ESRC grant ES/T002611/1. The research data are distributed by the UK Data Service.

Contact: Jim Vine (jim.vine@essex.ac.uk), Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK.

Contents

1. Introduction	1
References	2
2. Understanding Society: The UK Household Longitudinal Study	3
References	5
3. Innovation Panel Wave 15: Design	6
3.1 Call for experiments	7
3.2 Sample	7
3.3 Questionnaire design	8
3.4 Response outcomes	9
Longitudinal response outcomes	10
References	10
4. Experimentation in IP15	12
4.1 Impact of design and wording on Twitter data linkage	13
Introduction	13
Methods	13
Results	14
References	16
4.2 Eliciting the marginal propensity to consume in surveys	17
Background	17
Experimental design	18
Initial findings	19
References	21
4.3 Identifying informal carers: question wording and format	23
Experimental design	25

Results	27
Conclusion	28
References	29
4.4 Measuring alcohol consumption – an experiment	31
Experimental design.....	31
Results	31
Conclusion	33
References	33
4.5 Social pressure or nudging towards deeper cognitive processing? Experiments on the role of the interviewer in increasing consent to data linkage.....	35
Conceptual idea	35
First results.....	37
References	38
4.6 Improving the reporting of mobile contact details and consent to send survey questions by SMS	39
Introduction	39
Methods.....	40
Results	44
Conclusion.....	47
References	47
4.7 Protocols to invite survey respondents to data collection using mobile apps: Effects of survey mode, incentives and feedback on participation rates and bias .	49
Motivation.....	49
The BodyVolume app.....	49
Study design	49
Results	51

Summary	53
References	54
4.8 Can we rely on self-reported biomarker data? Benchmarking the accuracy and feasibility of self-reported hip and waist measurements using analogue and digital methods	55
Introduction	55
Methods.....	56
Results.....	58
Benchmarking self-measurement against reliable reference measurement.....	60
Conclusion.....	63
References	64
4.9 National identity choice and meaning over time: role of affective priming and question order	65
Experiment, methods and data	66
Discussion of results.....	67
Conclusions and next steps.....	72
References	73

1. Introduction

This paper presents early findings from the fifteenth wave of the Innovation Panel (IP15) of *Understanding Society: The UK Household Longitudinal Study*. *Understanding Society* is a major panel survey for the UK. The first thirteen waves of data collection on the main sample have been completed, and the fourteenth and fifteenth waves are currently in the field. The data from the first twelve waves of the main samples are available from the UK Data Service, and the thirteenth will be available towards the end of 2023. Data from a nurse visit to collect bio-markers from the general population sample and the British Household Panel Survey (BHPS) are also available. Data for all completed waves of the Innovation Panel are also available from the UK Data Service (University of Essex, Institute for Social and Economic Research 2023).

One of the features of *Understanding Society*, alongside the large sample size (40,000 households at Wave 1) and the ethnic minority boost sample and the collection of bio-markers, is the desire to be innovative. This has been a key element of the design of *Understanding Society* since it was first proposed. Part of this drive for innovation is embodied within the Innovation Panel (IP). This panel of 1,500 households was first interviewed in 2008. Refreshment samples of around 500 respondent households each have been added in waves 4, 7, 10, 11, and 14 of the Innovation Panel.

The design in terms of the fieldwork protocols, sample following rules, and questionnaire content are modelled on the main *Understanding Society* survey. The IP is used for methodological testing and experimentation that would not be feasible on the main sample. The IP is used to test different fieldwork designs, new survey questions and new ways of asking existing questions. The design of the Innovation Panel, fieldwork and outcomes, and the content of the survey at each wave are documented in the User Guide (Institute for Social and Economic Research 2023).

Working Papers covering the experiments carried out in all previous Innovation Panels are available from the *Understanding Society* website.¹

¹ <https://www.understandingsociety.ac.uk/research/publications/working-papers>

This paper describes the design of IP15, the experimental studies carried and some preliminary findings from early analyses of the data. Section 2 outlines the main design features of *Understanding Society*. Section 3 describes the design and conduct of IP15. Section 4 then reports on the experiments carried at IP15.

References

- Institute for Social and Economic Research (2023). Understanding Society – The UK Household Longitudinal Study, Innovation Panel, Waves 1-15, User Guide. Colchester: University of Essex, <https://www.understandingsociety.ac.uk/documentation/innovation-panel/user-guide>.
- University of Essex, Institute for Social and Economic Research. (2023). Understanding Society: Innovation Panel, Waves 1-15, 2008-2022. [data collection]. 12th Edition. UK Data Service. SN: 6849, <http://doi.org/10.5255/UKDA-SN-6849-15>.

2. Understanding Society: The UK Household Longitudinal Study

Understanding Society is an initiative of the Economic and Social Research Council (ESRC) and is one of the major investments in social science in the UK. The study is managed by the Executive Team (ET), based at the Institute for Social and Economic Research (ISER) at the University of Essex and includes topic experts from a number of institutions. The fieldwork and delivery of the survey data for the first five waves of the main samples were undertaken by NatCen Social Research (NatCen). Since Wave 6, Kantar Public has been the lead contractor. *Understanding Society* aims to be the largest survey of its kind in the world. The sample covers the whole of the UK, including Northern Ireland and the Highlands and Islands of Scotland. *Understanding Society* provides high quality, longitudinal survey data for academic and policy research across different disciplines. The use of geo-coded linked data enables greater research on neighbourhood and area effects, whilst the introduction of bio-markers and physical measurements (Waves 2 and 3) opens up the survey to health analysts.

The design of the main *Understanding Society* study is similar to other household panel studies around the world. In the first wave of data collection, a sample of addresses was issued. Up to three dwelling units at each address were randomly selected, and then up to three households within each dwelling unit were randomly selected. Sample households were then contacted by NatCen interviewers, and the membership of the household enumerated. Those aged 16 or over were eligible for a full adult interview, whilst those aged 10-15 were eligible for a youth self-completion questionnaire. Sample members are interviewed annually, although the fieldwork for each wave is spread over two years: for each wave, the sample is split into 24 monthly batches with one issued to the field every month.

Initially all adult interviews were conducted using computer-assisted personal interviewing (CAPI). Adults who participated in *Understanding Society* were also asked to complete a self-completion questionnaire, in which questions thought to be more sensitive were placed. The adult self-completions at Waves 1 and 2, and the youth self-completions, were paper questionnaires. From Wave 3 onwards the adult self-completion instrument was integrated into the interviewing instrument and the respondent used the interviewer's laptop to

complete that portion of the questionnaire themselves (Computer-Assisted Self-Interviewing, CASI). For the first seven waves, surveys of continuing sample members were interviewer-administered. Before Wave 7 was issued, a random 20% of households were designated as ring-fenced face-to-face and would only be issued to CAPI first. Except for this ring-fenced sample, households that had not responded at Wave 6 were issued at Wave 7 to a sequential mixed mode design in which adults in those households were initially invited to participate online, and then those who did not take part online were issued to interviewers ("Web-first"). At Wave 8, the proportion of households that were issued web-first increased to 40%; this included households that had not participated at Wave 7, as well as those households who were predicted to be most likely to complete online – except for those who were in the ring-fenced sample. The remaining 60% of households were issued to interviewers, with non-respondents at the reissue stage invited to complete online ("F2F-first"). In each successive wave the proportion of the sample invited to take part online first was increased until it reached a maximum of 80%. In mid-March 2020, face-to-face interviewing was suspended due to the lock-down associated with the COVID-19 pandemic. Adult sample members who were allocated to interviewers at that time were sent a letter, explaining the position with face-to-face interviewing, and giving them their log-in details so they could complete online. Those who did not complete online were followed up by interviewers who tried to conduct the interview by telephone. From the April 2020 monthly sample onwards, all adult sample members were issued web-first, with telephone as the follow-up mode.

In between each wave of data collection, sample members are sent short reports of early findings from the survey, and a change-of-address card, to allow them to inform ISER of any change in their address and contact details. Before each sample month is issued to field, each adult is sent a letter which informs them about the new wave of a survey, includes a token of appreciation in the form of a gift voucher and also includes a change-of-address card. Interviewers then attempt to contact households and enumerate them, getting information of any new entrants into the household and the location of anyone who has moved from the household. For households completing the survey online, the information about household members is collected as part of the online survey. New entrants are eligible for inclusion in the household. Those who move within the UK are traced and

interviewed at their new address. Those people living with the sample member are also temporarily eligible for interview. More information about the sampling design of *Understanding Society* is available in Lynn (2009). From Wave 2, the BHPS sample has been incorporated into the *Understanding Society* sample. The BHPS sample is interviewed in the first half of each wave.

References

Lynn, P. (2009). Sample Design for Understanding Society *Understanding Society Working Paper Series No. 2009 – 01*,
<https://www.understandingsociety.ac.uk/research/publications/514007>

3. Innovation Panel Wave 15: Design

IP15 comprised six samples: the original sample from IP1, and refreshment samples taken at IP4, IP7 IP10, IP11, and IP14.

Starting at IP5, the modes used were online (web, CAWI) and face-to-face (CAPI) interviewing. In IP5, a random selection of two-thirds of households was allocated to the mixed-mode design with the remaining third of households allocated directly to face-to-face interviewers. This sample allocation has been maintained at each wave. However, starting at IP8 a subgroup of households with a very low propensity to respond via the web in the web condition was assigned to CAPI to begin fieldwork. Very low web propensity was determined by modelling web-completion using previous IP data. Unlike previous refreshment samples, the IP11 refreshment sampled households were allocated to either face-to-face or web-first recruitment. The IP14 refreshment sample were all allocated to web only recruitment.

IP12 had a focus on health and three types of interviews were conducted: by nurse face-to-face first, or interviewer face-to-face first and web first sequential mixed-mode designs similar to past waves. Households were randomly allocated to one of these three modes.

Due to COVID-19, all households were issued to web-first for IP13 and again for IP14.

For IP15, there was a return to the mixed mode design used at IP11 and previous waves, with households allocated to face-to-face-first or web-first designs. Fieldwork took place between 15th June and 25th November 2022.

For the **web-first sample**, for the first five weeks of fieldwork households were initially invited to take part online. After those five weeks, any households and individuals that had not completed online were issued to an interviewer for a 14-week face-to-face period (although the web survey remained open to complete during this time). Finally, the 'mop up' phase ran for the last 3 weeks of fieldwork, with interviews conducted via the telephone.

All web-first sample members aged 16 or over were sent a letter on the first day of web fieldwork asking them to complete the survey online and providing the web address and

their login details for doing so. The letter also explained that if they were unable to complete the survey online an interviewer would contact them as usual. Adults were sent four reminder emails (if an email address was available) and two reminder letters if they had not completed online by the time these reminder mailings were being prepared. These reminders were sent during the initial five-week web-first fieldwork period before households were issued to an interviewer.

For the **face-to-face-first sample**, households were initially issued to an interviewer for face-to-face interviewing, conducted over a 14-week period; during that time respondents could complete via the web if they requested login details, but this was not actively offered to the face-to-face sample during that period. In the final 3-week ‘mop up’ phase for the face-to-face sample, households were offered to complete online alongside telephone interviewing.

Face-to-face-first sample members aged 16 or over were sent an advanced letter, shortly before the start of the fieldwork period. This explained that an interviewer would contact them soon.

3.1 Call for experiments

IP15 was the twelfth time the Innovation Panel was open for researchers outside the scientific team of *Understanding Society* to propose experiments. A public call for proposals was made and six proposals were received with two being accepted. The initial proposals were reviewed by a panel formed of members of the *Understanding Society* executive team and of the fieldwork agency. In addition to the accepted experiments, seven additional experiments were included by the executive team in order to develop methodology for the main *Understanding Society* survey. One further experiment that is core to the fieldwork design of the study — the mixed-mode design — was continued from past waves, having been on hiatus at IP13 and IP14 due to COVID-19.

3.2 Sample

There were six samples issued at IP15: the original sample from IP1 and refreshment samples issued at IP4, IP7, IP10, IP11, and IP14. Samples comprised those households that had responded at IP14, plus some households that had not responded at IP14. Households

that had adamantly refused or were deemed to be mentally or physically incapable of giving an interview were withdrawn from the sample. There were 681 original sample households, 319 IP4 refreshment sample households, 392 IP7 refreshment sample households, 299 IP10 refreshment sample households, 487 IP11 refreshment sample households and 788 IP14 refreshment sample households issued. All households were originally selected from the Postcode Address File (PAF) using the same methods (see Lynn 2009).

3.3 Questionnaire design

The questionnaire at IP15 followed the standard format used in the previous Innovation Panels as well as the main stage of *Understanding Society*. The questionnaires used at IP15 are available from the *Understanding Society* website.² The interview included the following sections with the corresponding target times for each:

- Household roster and household questionnaire: 15 minutes per household.
- Individual questionnaire: average 31 minutes for each person aged 16 or over.
- Adult self-completion: around 9 minutes, computer assisted self-interview (CASI). In web and telephone interviews this was administered in the same way as the individual questionnaire.
- Youth self-completion: 10 minutes for each child aged 10-15 years.
- Proxy questionnaire: 10 minutes for adults ages 16 or over who are not able to be interviewed.

Some features of the questionnaire were designed to facilitate participants completing it online. These were largely introduced at IP5, when the web design was first introduced, and are described in more depth in the working paper containing results from the experiments in IP5 (Burton et al., 2013, 6-7). Briefly, the features are as follows. Questions are worded to include interviewer instructions that may clarify the definition of the question, as required. Text is designed to be participant-focused rather than interviewer-focused. The first person in the household to log in to the web survey is asked to complete the household enumeration; on completing the household enumeration, a question about who is

² <https://www.understandingsociety.ac.uk/documentation/innovation-panel/questionnaires>

responsible for paying household bills is asked; the person or people indicated as responsible are routed first to the household questionnaire (if no one in the household has completed it yet) and then to the individual questionnaire, while non bill-payers are routed directly to their individual questionnaire.

If a participant starts to answer their questionnaire and leaves the computer for 10 minutes, they are automatically logged out. The participant is able to log back in, and they are taken to the place where they left the questionnaire. This also applies to those who close the browser mid-interview. A 'partial interview' marker is included about two-thirds of the way through the questionnaire, after the benefits section. If a participant reaches this stage, the response is considered to be a 'partial interview'. A participant logging off after reaching the partial interview marker can log back in and complete if they want, but otherwise they are not contacted by an interviewer. If a participant logs out without having reached this marker, they are sent an email overnight which thanks them for their work so far and encourages them to complete the survey, giving them the URL to click through to the survey. Again, they would start at the point where they had left off. In addition, those who have started the questionnaire but not reached the partial interview marker are, after the initial five weeks, issued to interviewers who are able to finish the survey with them, from where they had left off.

3.4 Response outcomes

The issued sample at the 15th wave consisted of 2966 households. Fieldwork for all sample members was split by CAPI-first and web-first allocations. There were 1797 interviewed households from the continuing samples, for a 60.6% overall household response rate. Within these households, 2686 people were interviewed, for a conditional individual response rate of 76.5%. Of those 2686 respondents, 71% responded via the web, 27% face-to-face, and 2% via telephone. Of the web respondents, 51% responded using a PC/laptop, with the remainder split approximately equally between smartphones and tablets. Further details of the response rates, by sample and interview mode, and the devices used for web responses, are reported in the Innovation Panel User Guide (Institute for Social and Economic Research 2023, tables 13.36, 13.37 and 13.38).

Longitudinal response outcomes

The individual re-interview rate is an important outcome in a longitudinal panel study since many analyses require pairs of observations to measure change. Re-interview rates are calculated as the percentage of eligible units responding at later waves who were also surveyed at the initial wave. For those in the original sample, the percentage is predicated on response at IP1, while for the various refreshment samples their respective initial waves were IP4, IP7, IP10, IP11, and IP14.

As with any longitudinal study, there has been attrition over time, decreasing the overall numbers for each sample. At IP15, the response rates for the samples ranged from 21.0% of initial sample members (for whom IP15 was their 15th wave) being re-interviewed up to 68.4% of the IP14 refreshment sample (for whom IP15 was their second wave), with the IP4 to IP11 refreshment samples all having response rates between 31.4% and 38.9%. Full longitudinal re-interview rates are reported in the Innovation Panel User Guide (Institute for Social and Economic Research 2023, table 13.39).

References

- Burton, J. (Ed.) with Auspurg, K., Burton, J., Cullinane, C., Delavande, A., Fumagalli, L., Iacovou, M., Jäckle, J., Kaminska, O., Lynn, P., Mathews, P., Nicolaas, G., Nicoletti, C., Ye, C. and Zafar, B. (2013). Understanding Society Innovation Panel Wave 5: results from methodological experiments *Understanding Society Working Paper Series No. 2013–06* at <https://www.understandingsociety.ac.uk/research/publications/522000>
- Institute for Social and Economic Research (2023). Understanding Society – The UK Household Longitudinal Study, Innovation Panel, Waves 1-15, User Guide. Colchester: University of Essex, <https://www.understandingsociety.ac.uk/documentation/innovation-panel/user-guide>.
- Lynn, P. (2009). Sample Design for Understanding Society *Understanding Society Working Paper Series No. 2009–01*, <https://www.understandingsociety.ac.uk/research/publications/514007>

Williams, J., Ward, L., Burton, J., Carpenter, H., Cole, K., Hayward, B., Nicolaas, G., Parutis, V., Seymour, B., Thornton, K., Wood, M., and Woodward, L. (2022) *Understanding Society Wave 14 Boost Trial: Experiments with methods of recruiting a probability online boost sample*, *Understanding Society Working Paper Series No. 2022–07*. Colchester: University of Essex,
<https://www.understandingsociety.ac.uk/research/publications/547448>.

4. Experimentation in IP15

The following sections contain summaries of the experimental studies fielded in IP15, describing the design of each experiment and some initial results from early analysis of the data. The analyses in this working paper were based on a preliminary dataset which contained all cases but did not have weights or derived variables. The authors and proposers of the experiment of each sub-section below are given in the headings.

4.1 Impact of design and wording on Twitter data linkage

Tarek Al Baghal, Curtis Jessop, Paulo Serodio, Luke Sloan and Shujun Liu

Introduction

Linking social media and survey data at the individual level has the potential to add evidence to a variety of research questions. For example, respondents can be followed over time, real-time ‘observed’ data collected, new measures generated, and possible adjustments to non-response made (Al Baghal et al. 2020). Several studies have explored the nature of such linkages, with particular focus to date on ethics (Sloan et al. 2020), consent rates within the survey context (Al Baghal et al. 2020; Mneimneh 2022), and the asymmetric nature of the two data sources (Al Baghal et al. 2021). Other research has used these linkages to identify the error in reported and actual social media use (Henderson et al. 2019; Guess et al. 2018).

Previous research has shown consent to link Twitter and survey data is lower in a web and higher in interviewer-administered designs (Al Baghal et al. 2020). Developing a framework to understand consent to link social media data to survey responses, Mneimneh (2022) found that those with greater privacy concerns are less like to consent. These findings are similar to other consent for data linkage requests (e.g. Jackle et al. 2021). However, replication of findings would improve understandings and generalizability. Similarly, understanding ways to improve questionnaire design to improve consent rates (for example by addressing privacy concerns) can inform future study design.

Methods

Consequently, we included questions about Twitter use and a request for consent to link Twitter data to survey responses at IP15 (University of Essex, Institute for Social and Economic Research 2023). Twitter consent was originally asked at IP10 (Al Baghal et al. 2020). The questions at IP15 were based on these with updates made as a result of cognitive and qualitative interviews which explored participants’ understanding of the consent questions and how they answered them. These interviews suggested that when participants were provided with a lot of information, although its presence was generally

appreciated, many were not engaging with it fully and were instead taking ‘short-cuts’ in their decision making, potentially affecting (informed) consent rates.

To identify if we could improve informed consent rates, we included an experiment on how this information was presented. In the first condition, help links were presented as part of the initial asking of the consent question, as was done previously. In the second condition, the help links were removed from the initial asking. Instead, an additional explicit option was offered in the initial consent question (along with ‘yes’ and ‘no’): “Not sure, I would like more information”. Those selecting this option were then shown a new page, which presented the same help links on a separate screen, and the consent request was made again.

Additional changes to the question wording were made for both arms of the experiment compared to that used in IP10, including more information about why we want to collect the information and references to GDPR.

Results

Table 4.1-1 presents Twitter usage overall and by mode of response. Given the small number of telephone respondents, analysis only includes web and face-to-face respondents. The overall percent having an account (25.2%) is higher than that found in IP10 (21.7%), conducted in 2017 (Al Baghal et al. 2020). Similar to what was found in IP10 data, web respondents are significantly more likely to report having a Twitter account than face-to-face respondents ($\chi^2_1 = 29.4, p < 0.0001$).

Table 4.1-1 Twitter usage, overall and by mode of response

	Overall	Web	Face-to-face
Has Twitter account	25.2%	28.1%	17.9%
	(n=663)	(n=531)	(n=132)
		$\chi^2_1 = 29.4, p < 0.0001$	

The 663 respondents who indicated having a Twitter account were asked if they would consent to link their survey responses to their Twitter data. Table 4.1-2 presents consent rates overall, by mode of response, and by the experimental information conditions.

Table 4.1-2 Twitter linkage consent, overall, by mode of response, and by information condition

		<i>Mode of response</i>		<i>Information condition</i>	
	Overall	Web	Face-to-face	Same page	Separate page
Twitter consent	28.7%	25.1%	43.2%	29.6%	27.6%
	(n=190)	(n=133)	(n=57)	(n=101)	(n=89)
		$\chi^2_1 = 17.0, p < 0.0001$		$\chi^2_1 = 0.3, p = 0.57$	

As in IP10, and in-line with wider literature on consent to data linkage, there is a strong mode effect on consent rates, with web respondents having a significantly lower consent rate ($\chi^2_1 = 17.0, p < 0.0001$). Although the overall 28.7% consent rate is slightly lower than the 30.6% in IP10 (Al Baghal et al. 2020), the consent rate is actually higher at IP15 in both modes (25.1%, 43.2%) than the corresponding rates across the same modes at IP10 (24.3%, 40.5%). The decline in overall consent rate is attributable to the much larger percentage of web respondents at IP15, shifting the consent rate towards that mode's outcome. Overall, these are in line with most studies attempting social media linkage (Mneimneh 2022).

No significant effect is identified for whether help links were presented on the same or separate pages ($\chi^2_1 = 0.3, p = 0.57$). Analysis looking whether there is any difference in this result across modes also shows non-significant differences. Very few of the respondents given the additional option of "Not sure, I would like more information" chose this rather providing a consent decision in the first instance (n=13). This finding may be suggestive that provision of help links makes no difference one way or the other, but ethical considerations require inclusion. However, separating these from the main consent question seems unnecessary.

References

- Al Baghal, T., Sloan, L., Jessop, C., Williams, M., and Burnap, P. (2020). Linking Twitter and Survey Data: The Impact of Survey Mode and Demographics on Consent Rates Across Three UK Studies. *Social Science Computer Review*, 38: 517-532
- Al Baghal, T., Wenz, A., Sloan, L., and Jessop, C. (2021). Linking Twitter and Survey Data: Quantity and Possible Biases. *EPJ Data Science*, 10:32.
<https://doi.org/10.1140/epjds/s13688-021-00286-7>
- Guess, A., Munger, K., Nagler, J., & Tucker, J.A. (2018). How Accurate Are Survey Responses on Social Media and Politics? *Political Communication*, 36, 241-258
- Henderson, M., Jiang, K., Johnson, M. & Porter, L. (2019) Measuring Twitter use: validating survey-based measures. *Social Science Computer Review*, 39(6)
- Mneimneh, A. (2022) Evaluation of Consent to Link Twitter Data to Survey Data, *Journal of the Royal Statistical Society Series A: Statistics in Society*, Volume 185, Issue Supplement_2: S364–S386
- Jäckle, A., Beninger, K., Burton, J. & Couper, M.P. (2021) Understanding and improving data linkage consent in surveys. In P. Lynn (ed.) *Advances in Longitudinal Survey Methodology*, Chichester: Wiley.
- 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*, 15:63-76
- University of Essex, Institute for Social and Economic Research. (2023). Understanding Society: Innovation Panel, Waves 1-15, 2008-2022. [data collection]. 12th Edition. UK Data Service. SN: 6849, <http://doi.org/10.5255/UKDA-SN-6849-15>.

4.2 Eliciting the marginal propensity to consume in surveys

Thomas Crossley (European University Institute, tfcrossley@gmail.com)

Peter Levell (Institute for Fiscal Studies, peter.l@ifs.org.uk)

Hamish Low (University of Oxford, hamish.low@economics.ox.ac.uk)

Paul Fisher (Institute for Social and Economic Research, pfishe@essex.ac.uk)

Background

The “marginal propensity to consume” (MPC) is a crucial policy variable. It measures how consumers would change their spending in response to transitory income changes, and hence how consumers would respond to policies, such as tax cuts or cash payments, aimed at stimulating demand and employment. A large literature has attempted to estimate consumers’ MPCs by asking individuals how they respond to hypothetical windfalls or losses. Such questions have now been included in a number of surveys, including surveys run by central banks, reflecting the strong policy interest in these questions. Examples include the Eurozone’s Consumer Expectations Survey (run by the ECB), the New York Fed’s Survey of Consumer Expectations and the Bank of England/NMG consulting survey. Similar questions were included in the Understanding Society COVID survey.

It has emerged that different studies have yielded very different estimates of the distribution of MPCs and these appear unlikely to be explained by only by differences in the economic environments that the questions were asked in. For example, Jappelli and Pistaferri (2014) and Jappelli and Pistaferri (2020) find that consumers would spend around 50% of each additional £1 they receive, while Fuster, Kaplan and Zafar (2020) find an average MPC of around 8%. One important possibility is that these differences reflect differences in the way questions are worded, rather than differences in the way consumers would behave.

A key difference in between the questions studied in the literature is whether they are a **direct question** or a **two-part question**. A **direct question** simply asks how much of a windfall consumers would spend. A **two-part question** asks, first, whether consumers would increase, decrease, or not change spending in response to a windfall, and then, if they indicated that would increase (or decrease) spending, they are asked (in the second

part) by how much. Studies using a direct question tend to find larger MPCs across different countries (e.g., Jappelli and Pistaferri 2014, 2020; Drescher et al. 2020; Christelis et al. 2020), while studies using two-part formats find smaller MPCs (Fuster, Kaplan and Zafar 2020; Crossley et al. 2021). We hypothesize that the differences are due to question format. The direct question may prime uncertain respondents to believe they should spend some part of a hypothetical windfall. It also constrains responses to be between 0 and 100%. The two-part question may be more neutral and allows responses that are outside that range. However, comparisons across past studies using alternative formats are confounded by differences in targeted population, economic conditions, and timing more generally. In this Innovation Panel Experiment we randomly assign panel members to either the direct or two-part format, and so isolate format effects.

Experimental design

We randomly assigned subjects to either the direct or two-part formats. The wording was as follows:

Direct question

Imagine you unexpectedly received a one-time payment of £[AMOUNT] today. How much of it would you spend over the next [DURATION] months?

Two-part format

Now consider a hypothetical situation where you unexpectedly receive a one-time payment of £[AMOUNT] today. We would like to know whether this extra income would cause you to change your spending, borrowing and saving behaviour in any way over the next [DURATION] months.

If you received the one-time £[AMOUNT] payment: [CHOOSE AMONG THE FOLLOWING 3 OPTIONS]

Over the next [DURATION] months, I would spend more than if I hadn't received the £[AMOUNT]

Over the next [DURATION] months, I would spend the same as if I hadn't received the £[AMOUNT]

Over the next [DURATION] months, I would spend less than if I hadn't received the £[AMOUNT]

You indicated that you would {increase/ reduce} your spending over the next [DURATION] months following the receipt of the £[AMOUNT] payment. How much {more / less} would you spend than if you hadn't received the £[AMOUNT]?

Previous literature has also documented difference in elicited MPCs by the amount of the windfall, and the time-horizon for subsequent spending (Fuster, Kaplan and Zafar, 2020; Fagereng et al., 2021). For this reason, we also randomly allocated respondents to time periods (how much they would change spending over the next 3 months, or how they would respond over the next 12 months) and windfalls (how they would respond to either a £500 or a £2500 windfall.)

The three treatments were fully interacted, so the Innovation Panel was randomly allocated to eight equally sized groups, and respondents within each group asked a different question variant. Randomisation took place at the household level and was implemented by block where blocks were defined by survey design variables (incentive, mode and sample origin), and for question variant and windfall, the randomised time period. Our final sample consists of 2737 individuals. (University of Essex, Institute for Social and Economic Research 2023)

The group sizes can be found in column 7 of Table 4.2-1.

Initial findings

For each of the 8 treatment groups, Table 4.2-1 reports a statistical summary of the elicited MPCs. The first column gives the unconditional mean, followed by the prevalence of a strictly positive MPC, the mean of the MPC conditional positive, and the 25th, 50th and 75th percentile of the distribution. Two points are immediately obvious, both consistent with our hypothesis that the direct question primes the respondent to report positive expenditure. First, the direct question results in much higher reported MPCs than the two-part format. Across time-horizons and windfall sizes, the mean MPC elicited by the direct question ranges from 0.51 to 0.79, while the MPC elicited with the two-part format range from 0.11 to 0.13. Second, much of the difference is on the extensive margin. A respondent randomly allocated to the direct question is much more likely to report a positive MPC (80-87%) than

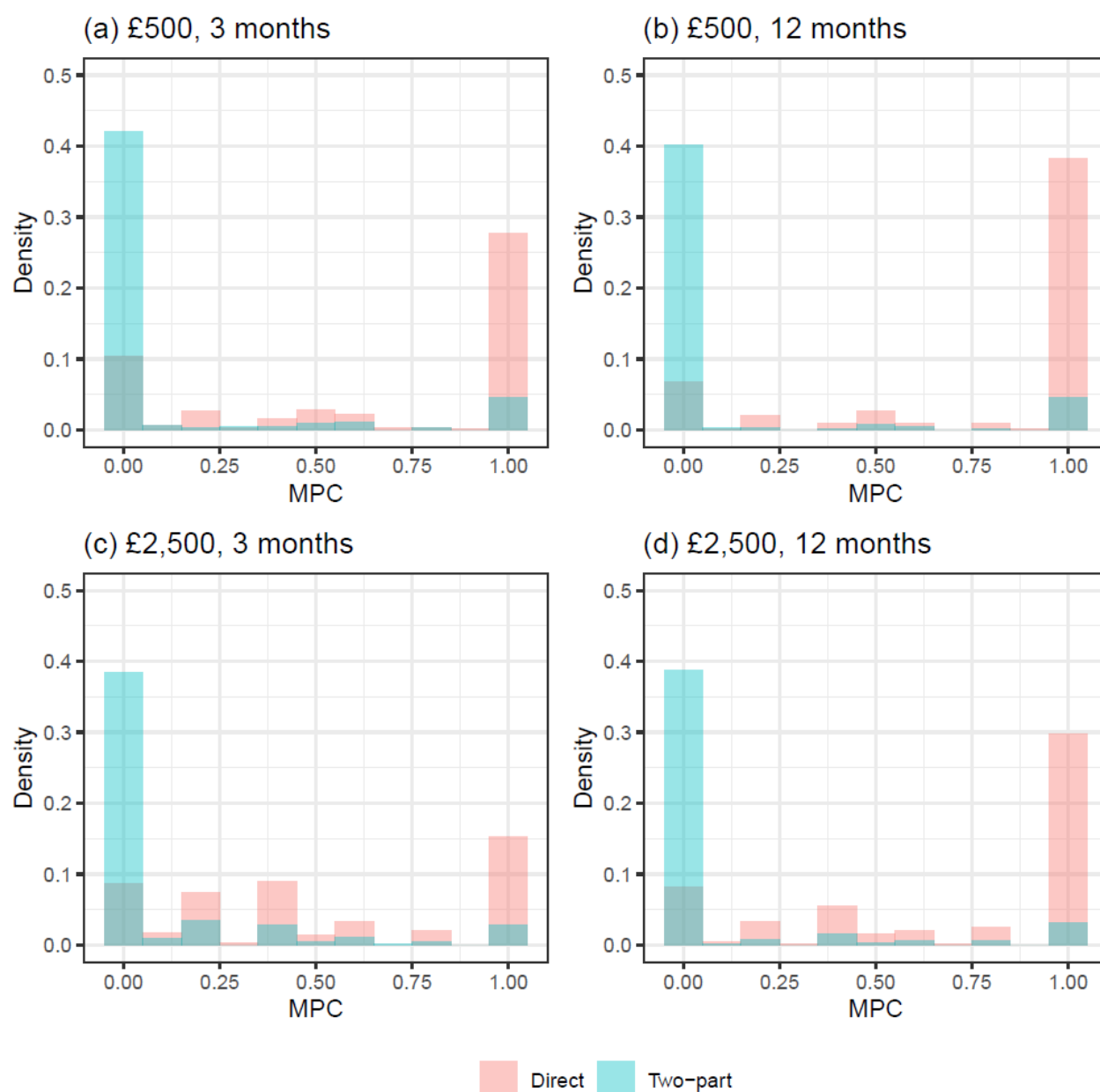
a respondent randomly allocated to the two-part format (15-25% with a positive MPC). Across the two treatments, there is much less difference in mean MPCs conditional on positive (column 3).

Table 4.2-1 Distribution of MPCs in different treatments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mean	Prob MPC>0	Mean MPC>0	p25	Median	p75	N
Direct q., £500, 3 months	0.66	0.80	0.83	0.20	1.00	1.00	334
Direct q., £500, 12 months	0.79	0.87	0.91	0.60	1.00	1.00	366
Direct q., £2,500, 3 months	0.51	0.85	0.60	0.20	0.40	1.00	331
Direct q., £2,500, 12 months	0.69	0.86	0.80	0.40	1.00	1.00	353
Two-part, £500, 3 months	0.13	0.17	0.73	0.00	0.00	0.00	353
Two-part, £500, 12 months	0.12	0.15	0.81	0.00	0.00	0.00	326
Two-part, £2,500, 3 months	0.12	0.25	0.49	0.00	0.00	0.00	345
Two-part, £2,500, 12 months	0.11	0.16	0.67	0.00	0.00	0.00	329

Figure 4.2-1 displays these findings graphically. While we observe many MPCs of 0 or 1, interestingly we don't find the heaping at 0.5 that has been reported in some previous studies (Jappelli and Pistaferri 2014).

Figure 4.2-1 MPCs by question format, time-horizon and windfall size



References

- Bunn, Philip, Jeanne Le Roux, Kate Reinold, and Paolo Surico. "The consumption response to positive and negative income shocks." *Journal of Monetary Economics* 96 (2018): 1-15.
- Christelis, Dimitris, Dimitris Georgarakos, Tullio Jappelli, and Geoff Kenny. "The Covid-19 crisis and consumption: survey evidence from six EU countries." *Available at SSRN 3751097* (2020).

- Crossley, Thomas F., Paul Fisher, Peter Levell, and Hamish Low. "MPCs in an economic crisis: Spending, saving and private transfers." *Journal of Public Economics Plus* 2 (2021): 100005.
- Drescher, Katharina, Pirmin Fessler, and Peter Lindner. "Helicopter money in Europe: New evidence on the marginal propensity to consume across European households." *Economics Letters* 195 (2020): 109416.
- Fagereng, Andreas, Martin B. Holm, and Gisle J. Natvik. "MPC heterogeneity and household balance sheets." *American Economic Journal: Macroeconomics* 13, no. 4 (2021): 1-54.
- Fuster, Andreas, Greg Kaplan, and Basit Zafar. "What would you do with \$500? Spending responses to gains, losses, news, and loans." *The Review of Economic Studies* 88, no. 4 (2021): 1760-1795.
- Jappelli, Tullio, and Luigi Pistaferri. "Fiscal policy and MPC heterogeneity." *American Economic Journal: Macroeconomics* 6, no. 4 (2014): 107-136.
- Jappelli, Tullio, and Luigi Pistaferri. "Reported MPC and unobserved heterogeneity." *American Economic Journal: Economic Policy* 12, no. 4 (2020): 275-97.
- University of Essex, Institute for Social and Economic Research. (2023). Understanding Society: Innovation Panel, Waves 1-15, 2008-2022. [data collection]. 12th Edition. UK Data Service. SN: 6849, <http://doi.org/10.5255/UKDA-SN-6849-15>.

4.3 Identifying informal carers: question wording and format

Karon Gush and Sarah Parbury

Informal care plays a vital role in meeting the needs of those requiring care and support. Current debates around social care funding and the recent COVID pandemic have thrown light on just how fragile and fragmented the formal UK social care system is and its reliance on informal care work in shoring up many of its shortcomings. Census 2021 data shows that 5 million people in England and Wales identified as informal carers, a decrease from the 5.8 million in 2011 (ONS 2023). This decrease is surprising as many were predicting increases (e.g., Guzman-Castillo et al. 2017; Carers UK 2019; Carers UK 2020). However, the 2021 Census took place during the pandemic and it is not yet clear how mortality rates and lockdowns affected the practice of unpaid caring, nor is it clear how changes in the way the unpaid caring question was asked in the 2011 and 2021 censuses affected reporting (Petrillo and Bennett 2023).

Measurement of informal care is problematic for several reasons. Caregiving is a complex concept and ontologically vague. Linguistically speaking, 'care' can refer to emotions and/or practices and, in common parlance, 'caregiving' tends to embody a nebulous set of activities (McCance et al. 1997). Often considered as including support with *tangible* tasks (personal care, bathing, dressing, food preparation, transportation, shopping, housework, etc.), whilst forms of *intangible* support (emotional support, communication, companionship and socialising, supervision, advocacy, etc.) are not always considered (Dumont et al. 2010). At an abstract level, caring is doing things that the recipient cannot do for themselves and that which goes 'beyond the normal reciprocities between adults' (Twigg and Atkin 1994: 8). However, the generalised reciprocities within families, particularly in co-residential relationships, can make it difficult to distinguish between caring and what Waerness (1984) calls 'servicing' work, i.e. activities in filial and spousal relationships either as part of kinship obligation or of household labour specialisation. That is to say, the social relations under which the task is performed are as important as the actual task being performed.

Informal care is often defined as unpaid, non-contractual or that which happens outside of formal employment, which may seem straightforward, but this can also pose measurement challenges where boundaries of paid/unpaid, job/non-job-related are blurred. These include

'home share' schemes where more-able adults are matched with those with care needs and the former receives subsidised accommodation in return for care work; spill-overs from the formal care worker/service user relationship where care tasks are performed in unpaid time; what Ungerson (1995: 33) calls 'symbolic' payments where notional amounts are paid to volunteers or the 'shadowy' caring labour market (Ungerson 1995: 37) where small cash/gifts are given to friends/family in recognition of received care services.

In addition to the lack of a clear definition of what constitutes informal care, public perceptions and personal identities shape whether individuals recognise themselves as unpaid carers and/or recipients. Some do not associate the term 'care' with the support they give/receive, either because it does not occur to them it is a form of care or because they reject the notion that what they provide is 'care' (Heron 1998; Henderson 2001; Stalker 2002; Corden and Hirst 2011). Part of this may be due to the increasing visibility and professionalisation of the formal care sector, whereby unpaid carers see what they do as distinct and different from that in the paid sector (Lloyd 2006). Whilst it may be clear that the term 'care' is probably not ideal terminology to use in survey questionnaires, it is also not clear how much of an improvement the phrases typically used are (e.g., 'providing support', 'looking after', 'special help') as these are also subject to the same issues of conceptual vagueness and overlap with everyday exchanges between friends, family and neighbours.

To investigate the effect of question wording and format on reporting of informal care provision, at IP15 we experimented with different ways of identifying informal carers (University of Essex, Institute for Social and Economic Research 2023). Our main research questions were:

RQ1. To what extent do clarifications of terminology around informal caring matter in question wording?

RQ2. What is the effect of asking respondents to report the informal care tasks they perform rather than asking them to report generalised unpaid care work?

We hypothesised that adding clarifications would increase the number of respondents reporting as carers as this would more clearly convey the full range of what should be

included. We also hypothesised that adopting a task-based approach would further increase the number of respondents who could be identified as informal carers given this is about reporting specific behaviours and less about subjective identities, the latter of which informal carers have been shown to sometimes reject.

Experimental design

Respondents were allocated randomly at the household level into one of three groups: a control group and two treatment groups.

Control group: Respondents answer existing set of informal care questions that have been carried at every wave of the Understanding Society main survey.

- Is there anyone living with you who is sick, disabled or elderly whom you look after or give special help to (for example, a sick, disabled or elderly relative, husband, wife or friend etc)?
- Do you provide some regular service or help for any sick, disabled or elderly person not living with you? Please exclude any help that you provide as part of your job.

Treatment Group A: Respondents are presented with a preamble explanatory text and are then asked whether they provide any 'unpaid support' to anyone.

- The next few questions are about any unpaid support you may give. By 'unpaid support' we mean helping someone who could be finding it hard to manage because of mental or physical illness, needing extra help as they grow older or because they have a physical or learning disability. This could be for anyone that you know (e.g. family, a friend, neighbour, colleague etc.)
Your support might include shopping, helping to find or arrange care or support, helping with managing money, giving regular emotional support, helping with transport, picking up prescriptions or providing hands on care (e.g. help with bathing, dressing etc.).
- Do you give unpaid support to anyone because they have long-term physical or mental health conditions or illnesses, or problems related to old age?

Treatment Group B: Respondents are first asked if any household members have health problems or disabilities and, if so, then asked about help provided with Activities of Daily

Life (ADL) / Instrumental Activities of Daily Life (IADLS). Next, respondents are asked about provision of specific support tasks to non-household members.

- Do you provide any of the following kinds of help to <NAME> because of their health problems or disabilities? Please select all that apply.
 - Getting up and down stairs or steps
 - Getting around the house (except for any stairs)
 - Getting in and out of bed
 - Cutting their toenails
 - Bathing, showering or washing all over
 - Using the toilet, including getting up and down
 - Eating, including cutting up food
 - Washing their face and hands
 - Dressing and undressing, including putting on shoes and socks
 - Taking the right amount of medicine at the right times
 - Going outdoors and walking down the road
 - Shop for food, including getting to the shops, choosing items, carrying items home and unpacking them
 - Doing routine housework, e.g. preparing a hot meal, doing personal laundry
 - Doing paperwork or pay bills
 - Something else
 - None of these
- Apart from any paid work you do as part of your job, do you provide any of the following kinds of support or help to anyone not living with you who has a physical, mental, emotional or memory problem? Please select all that apply.
 - Personal care, e.g. dressing, bathing or showering, eating, getting in or out of bed, using the toilet, taking medication
 - Practical tasks, e.g. shopping for groceries, preparing a hot meal, personal laundry, help getting out and about, transportation.
 - Help with paperwork, e.g., filling out forms, paying bills or settling legal matters.
 - Emotional support, e.g. visiting them, keeping company, someone to talk to, listening to problems

- Something else
- None of the above

Results

Table 4.3-1 shows the proportions of respondents in each of the three experimental groups reporting they provide informal care. In the case of Treatment Group B, this is the provision of any of the specific tasks.

Table 4.3-1 Identification of informal carers across experimental groups

	Control Group	Treatment A	Treatment B
Carer	15.5% (143)	19.2% (170)	24.1% (209)
Non-carer	84.5% (778)	80.8% (716)	75.9% (209)
Total	100% (921)	100% (886)	100% (869)

$\chi^2(2) = 2.8, p < 0.001; n = 2,676$ (cell sizes in parentheses)

These results support both the expectation that adding more clarification about what respondents should include as caring increases the number reporting they provide it; and that using a specific task-based questioning approach further identifies informal carers. When asking in a non-specific way with minimal clarification, 15.5% of respondents report providing informal care (Control Group) however when clarification is added, or when a task-based approach is used, 19.2% and 24.1% of respondents respectively report providing informal care.

These findings largely hold true when examining whether informal care is provided to co-residents or not, though the picture is a little more mixed. Table 4.3-2 shows that the task-based questioning approach identifies higher proportions of informal carers for those who perform care only for those inside the household or only for those outside the household. For those caring both inside *and* outside the household the task-based approach identified slightly fewer carers (1.8%) compared to that in the Control Group (2.1%). Further analysis is required to understand this better. The proportion caring both inside and outside the

household is understandably small. Caring can be very demanding and to do so across multiple locations is likely to pose additional challenges. Equally caring can be very personally rewarding and thus motivations for providing care can be complex.

Table 4.3-2 Identification of intra- and extra-household informal carers across experimental groups

	Control group	Treatment A	Treatment B
Carer – for those in the household only	6.6% (61)	4.6% (41)	7.4% (64)
Carer – for those outside the household only	6.8% (63)	13.9% (123)	14.8% (129)
Carer – for both those inside <u>and</u> outside the household	2.1% (19)	0.7% (6)	1.8% (16)
Non-carer	84.5% (778)	80.8% (716)	75.9% (209)
Total	100% (921)	100% (886)	100% (869)

$\chi^2(6) = 45.6, p < 0.001; N = 2,676$, (cell sizes in parentheses)

Conclusion

Informal care is a highly nebulous concept. Despite being a hugely important and much discussed social good, there is wide concern about the difficulties in measuring it. These results suggest that giving respondents more information about what should be included when they are formulating their responses aids the identification of informal carers.

Furthermore, if this is done by asking respondents to report provision of specific care-tasks rather than ‘unpaid support’, ‘special help’, or ‘look[ing] after’ more generally, the identification is further improved.

References

- Carers UK (2019) Juggling work and unpaid care. A growing issue. London.
- Carers UK (2020) The rise in the number of unpaid carers during the coronavirus (COVID-19) outbreak. *Carers Week 2020 Research Report*. London.
- Corden, A. & Hirst, M. (2011) Partner care at the end-of-life: identity, language and characteristics. *Ageing and Society*, 31, 217-242.
- Dumont, S., Jacobs, P., Turcotte, V., Anderson, D. & Harel, F. (2010) Measurement challenges of informal caregiving: a novel measurement method applied to a cohort of palliative care patients. *Social science & medicine (1982)*, 71, 1890-1895.
- Guzman-Castillo, M., Ahmadi-Abhari, S., Bandosz, P., Capewell, S., Steptoe, A., Singh-Manoux, A., Kivimaki, M., Shipley, M., Brunner, E. & O'flaherty, M. (2017) Forecasted trends in disability and life expectancy in England and Wales up to 2025: a modelling study. *The Lancet. Public health*, 2, 307-313.
- Henderson, J. (2001) 'He's not my carer—he's my husband': Personal and policy constructions of care in mental health. *Journal of Social Work Practice*, 15, 149-159.
- Heron, C. (1998) *Working with Carers*, London, Jessica Kingsley.
- Lloyd, L. (2006) Call us carers: Limitations and risks in campaigning for recognition and exclusivity. *Critical Social Policy*, 26, 945-960.
- McCance, T., McKenna, H. & Boore, J. (1997) Caring: dealing with a difficult concept. *International journal of nursing studies*, 34, 241-248.
- ONS (2023) *Health, disability and unpaid care. Census 2021*. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/bulletins/unpaidcareenglandandwales/census2021> (accessed August 2023).
- Petrillo, M. & Bennett, M. (2023) *Valuing Carers 2021: England and Wales*. London: Carers UK

- Stalker, K. (2002) Carers: An Overview of Concepts, Developments and Debates. *In*: STALKER, K. (ed.) *Reconceptualising Work with Carers: New Directions for Policy and Practice*. London: Jessica Kingsley Press.
- Twigg, J. & Atkin, K. (1994) *Carers Perceived: Policy and Practice in Informal Care*, Buckingham, UK., Open University Press.
- Ungerson, C. (1995) Gender, Cash and Informal Care: European Perspectives and Dilemmas. *Journal of Social Policy*, 24, 31-52.
- University of Essex, Institute for Social and Economic Research. (2023). Understanding Society: Innovation Panel, Waves 1-15, 2008-2022. [data collection]. 12th Edition. UK Data Service. SN: 6849, <http://doi.org/10.5255/UKDA-SN-6849-15>.
- Waerness, K. (1984) Caring as women's work in the welfare state. *In*: HOLTER, H. (ed.) *Patriarchy in a Welfare State*. Oslo: Universitetsforlaget.

4.4 Measuring alcohol consumption – an experiment

Sarah Parbury and Tarek Al Baghal

AUDIT is a comprehensive 10 question alcohol harm screening tool, developed by the World Health Organisation (WHO) and modified for use in the UK (Saunders et al. 1993; Office for Health Improvement and Disparities 2020). It is used in a variety of health and social care settings including the NHS. A subset of three questions asks about consumption: frequency of alcohol consumption, amount consumed on a typical drinking day, and frequency of binge drinking. In the Audit-C version currently used by the NHS, people are asked about amounts of alcohol consumed in terms of **units**. Units are defined in help text as a ‘1/2 pint of beer, a glass of wine or a single measure of spirit or liqueur’. These questions ask about how many units are drunk on a typical day (question AUDITC4) and how many times more than 6 (for women) or 8 (for men) units were consumed on a single occasion in the past year (AUDITC5). In Understanding Society, historically the same questions have been asked, but AUDITC4 was asked in terms of number of **drinks** (not units), with AUDITC5 asked in terms of units, as in the NHS version. The question arises as to whether respondents distinguish between drinks and units when asked differently.

Experimental design

We used a split-sample design to compare these two measures, with data from the IP15 wave of Understanding Society (University of Essex, Institute for Social and Economic Research 2023). Half of the sample received the NHS units question wording, and the other half received the drinks wording. If respondents distinguish between the two, we would expect a divergence, with drinks reports being fewer than units reports.

Results

The base for analysis is respondents who have consumed at least one alcoholic drink in the last 12 months ($n = 2,122$). Looking first at consumption on a typical day when drinking, reporting varied by experimental group (see Table 4.4-1). Those in the ‘drinks group’ were more likely to answer 1-2 drinks, compared with those in the ‘units group’ (61% and 48%

respectively). In contrast, those in the 'units group' were more likely to report higher numeric values for alcohol consumption. 12% of those in the 'units group' said they consumed 5-6 units of alcohol on a typical day when they were drinking, compared to only 8% of those in the 'drinks group'. This distribution is indicative that in the aggregate, respondents do differentiate when asked about drinks or units. Given units are generally smaller measures, fewer drinks should be reported than units, and that is what is observed.

Table 4.4-1 AUDITC4 Alcohol consumption, typical day, by experimental group

	Number of units/drinks on typical day when drinking					
	1-2	3-4	5-6	7-9	10+	<i>Base</i>
Units	48%	28%	12%	5%	5%	1,092
Drinks	61%	25%	8%	3%	2%	1,030
	1,157	567	213	84	74	2,122

Note: Differences are statistically significant $p < 0.001$

Now looking at binge drinking, that is, drinking heavily over a short space of time, we also see variation by experimental group (see Table 4.4-2). A higher proportion of those in the 'drinks group' responded 'never', compared with those in the 'units group' (52% and 41% respectively). For the other categories, respondents asked in units were more likely to say monthly or weekly, compared with respondents asked in drinks (13% vs 9% and 7% vs 4% respectively).

Table 4.4-2 AUDITC5 Frequency of binge drinking (in the last year) by experimental group

	Frequency of binge drinking – 6/8 plus units/drinks					
	Never	Less than monthly	Monthly	Weekly	Daily or almost daily	<i>Base</i>
Units	41%	36%	13%	7%	2%	1,092
Drinks	52%	34%	9%	4%	1%	1,030
	987	740	232	123	30	2,122

Note: Differences are statistically significant $p < 0.001$

Survey questions are designed with the principle of maximising response. Questions should ask for information in a way that respondents can access and recall that information.

Questions asking about units of alcohol consumed require respondents to have both

knowledge of the units in each measure of drink they had, as well as remembering and totalling this information. Some respondents may find that cognitive task difficult or may not think about their alcohol intake in that way. For some, it may be easier to remember how many drinks they consumed.

Help text in the question does clarify what a 'unit' is, if that information is requested. We don't have information on how many respondents asked for such clarification, but the results do suggest that respondents can and do differentiate drinks and units, to some degree, at least. The distributions for both questions using units skews towards higher numbers than when the same questions are asked with drinks, as is the expectation. There is some (albeit minor) evidence that the units questions are more unclear than drinks. Whilst the number of cases are small, there were more 'don't know' (DK) responses to the units-based question about consumption than there were to the drinks-based question (ten DK vs. six DK respectively at AUDITC4). There was no difference in 'don't know' responses at the binge-drinking question (two DK for the units-based question and two DK for the drinks-based question at AUDITC5).

Conclusion

Alcohol consumption can be measured in different ways. Respondents can be asked to report the number of drinks or units consumed. This experiment aimed to compare these two different approaches. Given that the results suggest that respondents **do** differentiate between drinks and units, Understanding Society will look to use units as in the standard NHS assessment.

References

- Office for Health Improvement and Disparities (2020). Alcohol use screening tests.
[website]. Updated 30 October 2020. GOV.UK (London, UK).
<https://www.gov.uk/government/publications/alcohol-use-screening-tests>
- Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993).
Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO
Collaborative Project on Early Detection of Persons with Harmful Alcohol

Consumption--II. *Addiction (Abingdon, England)*, 88(6), 791–804.

<https://doi.org/10.1111/j.1360-0443.1993.tb02093.x>

University of Essex, Institute for Social and Economic Research. (2023). Understanding Society: Innovation Panel, Waves 1-15, 2008-2022. [data collection]. 12th Edition.

UK Data Service. SN: 6849, <http://doi.org/10.5255/UKDA-SN-6849-15>.

4.5 Social pressure or nudging towards deeper cognitive processing? Experiments on the role of the interviewer in increasing consent to data linkage

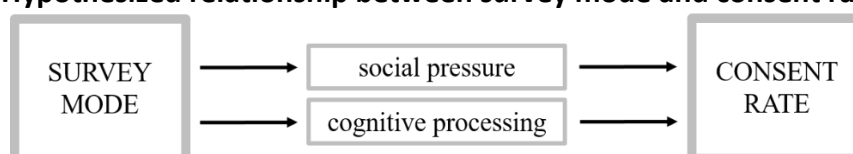
Annette Jäckle, Jonathan Burton, Mick P. Couper, Sandra Walzenbach

Conceptual idea

These experiments were designed to examine the mechanisms behind the large differences that we typically observe between survey modes, when asking respondents for consent to link their survey data to administrative records. One of the few consistent findings in the literature on data linkage consent is that consent rates are much higher in personal interviews than in web surveys: previous studies have found differences of 20-40 percentage points (Jäckle et al 2021; Sakshaug et al 2017; Thornby et al 2018). This mode effect is somewhat surprising — even more so as studies that have behaviour-coded respondent-interviewer interactions have found that interviewers do not actively do much to encourage consent, apart from reading out details of the linkage and the actual consent request (Jäckle et al 2022; Sakshaug et al 2021). In fact, Burton et al (2014) found that respondents who asked questions about the consent request were less likely to consent than respondents who did not ask any questions.

So far, previous research could not disentangle possible mechanisms through which the presence of the interviewer might affect consent (Figure 4.5-1): are the observable differences in consent a consequence of interviewer presence inducing socially desirable responding or does the process of interviewers reading out the relevant information nudge respondents into processing the request more thoroughly, which in turn leads to higher consent rates?

Figure 4.5-1 Hypothesized relationship between survey mode and consent rate



We implemented two experiments in IP15, to empirically assess this question (University of Essex, Institute for Social and Economic Research 2023). We made use of the mixed mode design, whereby sample members were randomly allocated to web-first or CAPI-first. For the consent question (asking for permission to link to government tax records), respondents who completed the survey with a face-to-face interviewer were randomly assigned to one of three groups:

CAPI: the interviewer read out the consent question and recorded the respondent's answer [high social pressure, nudge towards deeper cognitive processing]

Partial-CASI: the interviewer read out the consent question but asked the respondent not to tell them their answer; instead the respondent entered the answer in the interviewer's laptop, as the first question in the self-completion module [low social pressure, nudge towards deeper cognitive processing]

CASI: the respondent read and answered the consent question as part of the self-completion module of the survey, which respondents complete on the interviewer's laptop [low social pressure, no nudge towards deeper processing]

Respondents who completed the survey online were randomly assigned to one of two groups:

Web: a control group that was asked the standard self-completion consent question [no social pressure]

Web photo: a condition with a photo and personal message from the study director: "Hello, I am Professor Michaela Benzeval, Director of Understanding Society. Linking your survey answers with administrative records while preserving the confidentiality of your data increases the value of the information you provide. Please read the following carefully." [attempt to induce social pressure]

First results

Our preliminary analyses replicate the mode difference that previous research has found (see Table 4.5-1 and Table 4.5-2): at 77.3%, the consent rate is much higher in a personal interview (CAPI) than the web consent rate (41.6%), a difference of 35.6 percentage points. Additional analyses suggest that a relatively small part of the effect is driven by non-compliance with the randomly allocated survey mode, that is, by self-selection of different groups of respondents into different modes. Using allocated survey mode as an instrument for realized survey mode still suggests a highly significant mode difference of 30.1 percentage points.

In line with theoretical expectations, adding a personal note in the *web photo* condition compared to the standard *web* format significantly increased consent rates by 5.4 percentage points, from 41.6% to 47.0% with $p=0.02$ according to a mean comparison test (see Table 4.5-2).

However, there are some surprising results when we look at the three respondent groups that completed the survey with an interviewer present (see Table 4.5-1). The attempt to decrease the social pressure of the *CAPI* condition in the *partial-CASI* group, did not affect consent rates at all. Changing from the supported processing in *partial-CASI* to unsupported processing in *CASI*, increased consent rates by 7.6 percentage points ($p=0.045$).

Table 4.5-1 Consent rates for face-to-face conditions

	consent rate (%)	n
CAPI	77.3	220
Partial-CASI	77.3	198
CASI	84.9	219

Table 4.5-2 Consent rates for web conditions

	consent rate (%)	n
Web	41.6	903
Web photo	47.0	878

Further analyses will exploit a range of covariates that were collected in IP15 to shed light on these findings. Among these are a self-reported measure of the depth of cognitive processing, an objective measure of understanding the linkage process based on knowledge

test questions, self-reported feelings of social pressure as well as response times and audio recordings for part of the sample.

References

- Burton, J., Sala, E., and Knies, G. (2014). "Exploring role of interviewers in collecting survey respondents' consent to link survey data to administrative records." Presented at the International Panel Survey Methods Workshop, Ann Arbor, Michigan. 20-21 May 2014. <https://www.researchgate.net/publication/266532481>
- Jäckle, A., Beninger, K., Burton, J., and Couper, M.P. (2021). "Understanding data linkage consent in longitudinal surveys." Pp. 122-150 in Lynn, P. (Ed.) *Advances in longitudinal survey methodology*. Chichester: Wiley.
<https://doi.org/10.1002/9781119376965.ch6>
- Jäckle, A., Burton, J., Couper, M. P., Crossley, T. F. and Walzenbach, S. (2022) "How and Why Does the Mode of Data Collection Affect Consent to Data Linkage?", *Survey Research Methods*, 16(3), pp. 387–408.
<https://doi.org/10.18148/srm/2022.v16i3.7933>.
- Sakshaug, J.W., Hülle, S., Schmucker, A., and Liebig, S. (2017). "Exploring the effects of interviewer- and self-administered survey modes on record linkage consent rates and bias." *Survey Research Methods*, 11(2) pp. 171-88.
<https://doi.org/10.18148/srm/2017.v11i2.7158>
- 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>.
- University of Essex, Institute for Social and Economic Research. (2023). *Understanding Society: Innovation Panel, Waves 1-15, 2008-2022*. [data collection]. 12th Edition. UK Data Service. SN: 6849, <http://doi.org/10.5255/UKDA-SN-6849-15>.

4.6 Improving the reporting of mobile contact details and consent to send survey questions by SMS

Jim Vine, Daniel Horn, Jonathan Burton, Annette Jäckle

Introduction

Understanding Society, the UK Household Longitudinal Survey, is developing a programme of event-triggered data collection, to facilitate the collection of data close in time to significant life events, between the main annual waves (Jäckle et al., 2023). Panel members are initially asked a single question, about whether they have experienced any of the relevant life events in the past calendar month. This question is implemented as a web survey: respondents are sent invitations with personalised links by email and/or SMS, depending on the contact details they have provided. Respondents who report one or more events are routed into modules of follow-up questions about those events.

To increase coverage amongst sample members who use the internet infrequently or not at all, we have considered sending the initial filter question by SMS. Instead of including the link to the web survey in the SMS, we would include the question itself and ask respondents to reply by SMS. Consequently, the event-triggered data collection project is dependent on having accurate mobile phone numbers for panel members and benefits from having their informed consent to send survey questions via SMS.

In this section, we report on the findings of a non-experimental request made of all eligible respondents at IP15, seeking their consent to be sent questions via SMS (University of Essex, Institute for Social and Economic Research 2023). We also report on the reasons provided by those respondents who declined to give such consent.

In addition, we report on an experiment carried at IP15 that tested alternative ways of requesting respondents' contact details to improve the reporting of mobile phone numbers.

Methods

SMS question consent request

Most respondents were already asked for consent to send questions via SMS at IP13 (Jäckle et al 2021).

At IP15 all respondents who reported having a mobile phone were asked the same SMS consent question, if they had not already given consent at IP13. See Box 4.6-1.

Box 4.6-1: SMS consent question

Consent to send questions by text

UNIVERSE: Has a mobile phone and did not consent in IP13

TEXT: May we send you an occasional text message which contains a quick question?

OPTIONS:

- 1 Yes
- 2 No
- 1 Don't know
- 2 Refused
- 8 Inapplicable
- 9 Missing

The eligible pool of respondents included both respondents who had previously been asked for consent and not provided it and respondents who had not previously been asked for consent:

Asked at IP13 for consent to send questions via SMS but not provided it.	• Answered “No” to the consent question. • Answered “Don’t know” to the consent question. • Refused to answer the consent question or where the answer to that question is missing.
Had not previously been asked the consent question.	• New panel members who have joined since IP13 (including a refreshment sample added at IP14). • IP13 non-respondents including panel members with responses via proxy interview. • IP13 respondents who were not eligible to answer the consent question because they did not have a mobile phone at IP13.

Respondents who had provided consent at IP13 were not eligible for being re-asked at IP15³.

Our analysis presents separately the consent propensities for those who had previously been asked the consent question and declined to provide it, and those who were asked the question for the first time at IP15.

Those who declined to consent were asked a follow-up question seeking their reasons for not providing consent. See Box 4.6-2.

³ Two respondents who had provided consent at IP13 were re-asked, outside of the eligibility criteria. As they were not eligible, their answers have been excluded from our analysis. One said “Yes”, the other said “No”.

Box 4.6-2 SMS consent follow-up question to non-consenters

Follow-up question if respondent answered “no” to being sent questions by text message

UNIVERSE: Has a mobile phone and did not consent in IP13 and did not consent to questions by text

TEXT: To help us improve this study, can you tell us why you did not give us permission to send you an occasional text message which contains a quick question?

OPTIONS (select all that apply):

- 1 I don't use text messaging (much)
- 2 I would have to pay to answer the text message
- 3 I don't want to answer additional questions
- 4 I don't understand what this would involve
- 5 Unclear what the purpose of this is
- 97 Other reason [text]
- 1 Don't know
- 2 Refused
- 8 Inapplicable
- 9 Missing

Experimental test of contact detail requests

Mobile phone numbers are collected / checked towards the end of the interview at each wave. Prior to IP12, respondents were asked to check multiple contact details (home landline, personal mobile, work phone, email address) within the same question (version A). At IP12, an experiment was conducted to compare version A to a version that prioritised checking the mobile phone number by itself first, before checking the other details (version B). At IP13, version B was used for all relevant respondents.

The experiment at IP15 compared two variants of the approach first tested at IP12, both prioritising checking the mobile phone number before other contact details, testing version B against a new version C. The initial question asking whether the mobile number was incorrect varied for respondents who had not indicated an expected residential move, varying whether the question text included “To help us keep in contact” or not. (For respondents who did expect a move, the two versions were the same.) See Box 4.6-3.

Box 4.6-3 Versions of question checking whether mobile phone number is incorrect (underlining added to draw attention to differences)

<u>Version tested at IP12 (version B)</u>	<u>New version tested at IP15 (version C)</u>
UNIVERSE: Has provided at wave t-1 a mobile phone number and allocated to mobile phone questions version <u>B</u> TEXT (web): [Earlier you said that you expect to move in the coming year. <u>To help us keep in contact</u> {if DEMOGRAPHICS.XPMOVE = 1}} we would just like to confirm the following details we have on record for you. Please select any details that are INCORRECT.	UNIVERSE: Has provided at wave t-1 a mobile phone number & allocated to mobile phone questions version <u>C</u> TEXT (web): [Earlier you said that you expect to move in the coming year. {if DEMOGRAPHICS.XPMOVE = 1}} <u>To help us keep in contact</u>, we would just like to confirm the following details we have on record for you. Please select any details that are INCORRECT.
OPTIONS (web) 1 Personal mobile 2 No change needed -1 Don't know -2 Refused -8 Inapplicable -9 Missing	

Note: The difference between the two versions is whether the phrase "To help us keep in contact" is contained within the conditional statement only displayed to those moving or displayed to everyone seeing the question.

After the question checking whether the mobile number was correct was a question requesting the correct mobile number. This was used both if the respondent had indicated that the number previously held was incorrect and for respondents who had reported having a mobile phone but for whom that number was not held (including new sample members). This question varied based on the same allocation as the question checking whether existing mobile numbers were correct. See Box 4.6-4.

Box 4.6-4 Versions of question asking for new mobile phone number if previous one incorrect (underlining added to draw attention to differences)

<u>Version tested at IP12 (version B)</u> UNIVERSE: Has a personal mobile phone AND is a new entrant/never interviewed or has no fed-forward mobile phone number or fed-forward mobile number is incorrect AND allocated to mobile phone questions version <u>B</u> TEXT: Can I please have your personal mobile phone number? TEXT (web): Please enter your personal mobile phone number. Please use only numbers, no spaces or dashes, in your answer.	<u>New version being tested at IP15 (version C)</u> UNIVERSE: Has a personal mobile phone AND is a new entrant/never interviewed or has no fed-forward mobile phone number or fed-forward mobile number is incorrect AND allocated to mobile phone questions version <u>C</u> TEXT: <u>Earlier you said that you have a mobile phone.</u> Can I please have your personal mobile phone number <u>to help us keep in contact?</u> TEXT (web): Earlier you said that you have a mobile phone. Please enter your personal mobile phone number to help us keep in contact. Please use only numbers, no spaces or dashes, in your answer.
RESPONSE 1 [string] -1 Don't know -2 Refused -8 Inapplicable -9 Missing	

Results

SMS question consent request

Of those IP15 respondents who had not previously been asked the SMS consent question at IP13, 73% provided consent. Of the respondents who had been asked and declined consent at IP13, 53% provided consent when re-asked at IP15. See Table 4.6-1.

Table 4.6-1 IP15 SMS consent propensities, by whether the respondent had not been asked for consent at IP13 vs. having been asked and declined consent at IP13

	IP15 consent to send questions by SMS text message		Total
	Yes	No	
Not asked at IP13	768 (73%)	286 (27%)	1054
IP13 SMS non-consenters	230 (53%)	201 (47%)	431

Note: In this table, the non-consenters at IP13 and the “No” respondents at IP15 include all those who were asked the consent question and declined to provide it, so these numbers include those who answered “Don’t know” or refused to answer the question.

Of the respondents who said they would not consent to be asked questions by SMS, the most common reasons were that they did not want to answer additional questions (39%), that they don't use text messaging (33%), and that they were unclear what the purpose was (15%). See Table 4.6-2.

Table 4.6-2 Reasons provided by respondents who did not consent to SMS questions

	Number mentioning	% of respondents mentioning (N = 460)
I don't want to answer additional questions	181	39%
I don't use text messaging (much)	151	33%
Unclear what the purpose of this is	67	15%
I don't understand what this would involve	21	5%
I would have to pay to answer the text message	19	4%
[Other coded] too busy	12	3%
[Other coded] can't be bothered/don't want to	11	2%
[Other coded] prefer to be contacted another way	7	2%
[Other coded] can't have phone with me all the time	7	2%
[Other coded] don't have a phone	3	1%
[Other coded] health issues	3	1%
[Other coded] language/literacy problems	2	<1%
[Other coded] privacy concerns	2	<1%

Note: In this table, only 'active' non-consenting respondents (those who answered "No") were asked for their reasons, so this follow-up question was asked of fewer non-consenters (460) than those represented in Table 4.6-1 (487). Respondents could select more than one of the main response options, while 'other' responses were each coded to a single option.

Experimental test of contact detail requests

For respondents who had not indicated an expectation of a residential move, the prevalence difference of flagging their mobile number as incorrect was 10 percentage points higher in version C than version B (95% confidence interval: 6pp—14pp). See Table 4.6-3.

Table 4.6-3 Respondents flagging or not flagging their mobile number as incorrect, by question version or whether they had indicated an expected residential move

	Details incorrect wording version			
	Version B	Version C	Expect to move	Total
Mobile number flagged as incorrect	165 (18%)	246 (28%)	66 (32%)	477 (24%)
N	897	878	209	1984

For those respondents who flagged that the mobile number held was incorrect, the prevalence of providing a new number in both versions was 92%, (prevalence difference <-1 percentage point [95% confidence interval: -6pp—4pp]). See Table 4.6-4.

Table 4.6-4 Respondents providing or not providing a mobile number having flagged the existing one as incorrect

	Wording version		
	Version B	Version C	Total
Provided mobile number	170 (92%)	258 (92%)	428 (92%)
N	184	281	465

Note: The total number in this table differs from the number in the previous table who flagged an incorrect mobile number as there were cases where respondents had said their mobile number was incorrect but were not asked for the correct one, where they had indicated elsewhere in the survey (question [MOBUSE]) that they did not have a mobile phone.

The variants of the question requesting a new mobile number were also asked for respondents for whom an old mobile number was not held. We analyse these separately as the propensities of people providing a mobile number for the first time might differ from those of people correcting a previously-held mobile number. For these respondents, there was a 6 percentage point [95% confidence interval: -2pp—13pp] prevalence difference in providing a mobile phone number. See Table 4.6-5.

Table 4.6-5 Respondents providing or not providing a mobile number if no previous number held

	Wording version		
	Version B	Version C	Total
Provided mobile number	192 (66%)	245 (72%)	437 (69%)
N	289	340	629

Conclusion

The non-experimental repetition of the question seeking consent to send questions via SMS confirmed that if respondents decline to consent when first asked, many of them will provide consent if asked again at a subsequent wave.

The experiment examining different variants of questions asking (non-moving) respondents about their mobile number found respondents were more likely to flag their existing mobile number as incorrect in version C, which essentially only added a few words of justification to the request: “To help us keep in contact...”.

Having achieved this increased ‘flagging’ in the first question, version C of the second question requesting the new number sustained a similar prevalence of numbers being provided to version B (prevalence difference <-1 percentage point).

For the subset of respondents who were asked for their mobile numbers when one was not previously held, the point estimate was also for version C resulting in a higher prevalence of numbers being provided, albeit with a confidence interval that did not exclude an effect of the opposite sign (6pp [95% confidence interval: -2pp—13pp]).

References

- Jäckle, A., Burton, J., Couper, M. P., Vine, J. (2021). Event Triggered Data Collection. In *Understanding Society Innovation Panel Wave 13: Results from Methodological Experiments*, Tarek Al Baghal (ed.). University of Essex.
<https://www.understandingsociety.ac.uk/research/publications/547058>

Jäckle, A., Burton, J., Couper, M. P., Vine, J., & Horn, D. (2023). *Trialling event-triggered data collection in the Understanding Society Innovation Panel: Response and measurement quality* (Understanding Society Working Paper 2023-09,). University of Essex. <https://www.understandingsociety.ac.uk/research/publications/557841>

University of Essex, Institute for Social and Economic Research. (2023). Understanding Society: Innovation Panel, Waves 1-15, 2008-2022. [data collection]. 12th Edition. UK Data Service. SN: 6849, <http://doi.org/10.5255/UKDA-SN-6849-15>.

4.7 Protocols to invite survey respondents to data collection using mobile apps: Effects of survey mode, incentives and feedback on participation rates and bias

Annette Jäckle, Jonathan Burton, and Mick P. Couper

Motivation

Our interest is in the use of mobile applications to collect additional data on respondents in a general population panel study. In this study we examine different aspects of the fieldwork protocols for inviting survey respondents to complete additional data collection tasks. The overall aim is to identify fieldwork protocols that increase participation rates and reduce non-participation bias. We use data from an app study implemented in the 2022 Innovation Panel survey, a probability panel of households in Great Britain (University of Essex, Institute for Social and Economic Research 2023).

The BodyVolume app

Respondents were asked to install and use the BodyVolume app, as part of a set of questions about body measurements. Within the app respondents were asked to answer profile questions (age, sex, height, weight, activity level) and take two photos of themselves (front and side view). The app converted the photos into outlines of body shape from which it calculated body fat, visceral body fat, waist-hip ratio, and the lengths and circumferences of body parts. See the Appendix for a selection of screenshots from the app. The app was developed by Select Research and customized and branded for our purposes. It was available for Android and iOS smartphones and tablets through the corresponding app stores.

Study design

Sample: All respondents in the 2022 Innovation Panel survey who had completed at least one previous interview (n=2,536) were invited to install the BodyVolume app. The Innovation Panel is part of *Understanding Society*: the UK Household Longitudinal Study. All

members (aged 16 or over) of sample households are interviewed annually about their socio-economic situation including education, employment, income, health, housing, household membership and relationships. The design of the Innovation Panel is based on that of the main survey, but fielded separately as a platform for methods testing and experimentation.

Experiments: Our study involves three experiments. Treatment allocations were at the household level, such that all members of a household received the same treatment:

- (1) Survey mode of the annual interview: about 2/3 of the sample were randomly allocated to web-first and 1/3 to CAPI-first, with non-respondents followed up by web, CAPI and CATI.
- (2) Feedback on the app body measurements: in the invitation to the app study respondents were either promised feedback on their body fat, feedback on their visceral body fat (fat on the inside of the abdomen around organs), or feedback was not mentioned (1/3 of sample allocated to each group). The actual feedback within the app was the same for everyone and included the measured body fat, visceral body fat and waist-hip ratio.
- (3) Respondent incentives: half the sample were allocated to a conditional £5 incentive for completing the BodyVolume task and half had £5 added to the unconditional incentive sent with the advance letter for the annual interview, with the text highlighting that the extra incentive was because of the additional health-related tasks.

The allocations for (2) were stratified by allocations to (1); allocations for (3) were stratified by allocations to (1) and (2).

Invitation to the BodyVolume app study: in the introduction to the app task, respondents were instructed on what we wanted them to do: “install the BodyVolume app and use it to take two pictures of yourself: one from the front and one from the side”. They were told that the app would convert the pictures into outlines of their body shape and that these outlines would be uploaded to a server in the UK to calculate some measures based on their body shape. The body shape images would then be deleted, the original photos would not be uploaded. They were then shown an example of a body shape (silhouette) image and given an information leaflet (Institute for Social and Economic Research 2023, p53) or link to

an FAQ page (<https://www.understandingsociety.ac.uk/bodyvolume>). Respondents were told to complete this task after the annual interview, that it would take just a few minutes, and that they could delete the app as soon as they had completed the task. Depending on the allocated treatment groups, they were then told about the incentives and feedback. Respondents were given a unique 8-digit alpha-numeric access code for the app.

We varied the instructions for finding the app depending on how the respondent was completing the annual interview (based on the user agent string detected at this point in the interview: using an iOS device, an Android device, a different device or with a CATI or CAPI interviewer). The respondent was either shown a link to the app in the corresponding app store, or shown QR codes that would take them to the app in the Apple or Google app stores, or asked to search for the app in their app store. Respondents were then asked whether they managed to install and log in to the app. If yes, how they installed it (using the link, QR code or search in app store). If not, why not and, if they tried unsuccessfully, where in the process they dropped out.

After the annual interview all respondents for whom we had a valid email address and who had not declined to take part in the BodyVolume study were sent a reminder with information on how to access the app and their personal access code.

Results

Respondents who completed the annual interview with a face-to-face interviewer were most likely to install the app during the interview (48.8%), followed by those who completed the survey online (32.6%, Table 4.7-1). Telephone respondents were unlikely to install the app (7.8%). However, web respondents were most likely to actually use the app (19.8%), followed by face-to-face respondents (15.7%). These estimates show clear differences between the mode of interview and whether respondents participate in the additional task; however, they are not causal estimates due to the self-selection of respondents into modes. Using the randomised allocation of sample members to web-first versus CAPI-first as an instrumental variable for the mode of interview suggests that respondents were 18.2 percentage points more likely to install the app if they completed the annual interview in CAPI than if they completed it online (S.E. = 3.12, $p < 0.001$), there was however no

significant effect of the mode on whether respondents used the app ($\beta = 0.8$, S.E. = 2.52, $p = 0.756$).

Table 4.7-1 Whether installed app during interview / successfully used app, by mode of annual interview

Mode of annual interview	% Installed app	% Used app	N
CAPI	48.8	15.7	689
CATI	7.8	5.9	51
Web	32.6	19.8	1,796
Total	36.5	18.4	2,536

The feedback treatments varying whether or not feedback was promised and the type of feedback offered led to differences in the percentage of respondents who installed or used the app, however, the differences were not significant at conventional levels (Table 4.7-2).

Table 4.7-2 Whether installed app during interview / successfully used app, by feedback treatment

	% Installed app	% Used app
Total body fat feedback	34.5	17.5
Visceral body fat feedback	36.6	18.5
Feedback not mentioned	38.6	19.3
Pearson Chi2(2)	2.956, $p = 0.228$	0.956, $p = 0.620$

Compared to the unconditional incentive, the conditional incentive for using the app increased the percentage of respondents who installed the app during the interview by 4 percentage points ($p = 0.035$) and the percentage who used the app by 7.2 percentage points ($p < 0.001$, Table 4.7-3).

Table 4.7-3 Whether installed app during interview / successfully used app, by incentive treatment

	% Installed app	% Used app
Unconditional £5 incentive	34.5	14.8
Conditional £5 incentive	38.5	22.0
Pearson Chi2(1)	4.438, $p = 0.035$	21.987, $p < 0.001$

Initial analyses examining the differences between app users and the full sample suggest that the incentive treatments had a small effect on the bias. Based on the sex of the respondent, their age group, and the weight they reported in the annual interview, the

average absolute bias in the sub-set of app users is 2.6 in the conditional incentive treatment group and 2.9 in the unconditional incentive treatment group. Comparing the feedback treatment groups suggests that the extent of bias was slightly smaller in the group promised feedback on visceral body fat (2.3) compared to the group promised feedback on total body fat (2.9) and the group not promised feedback (2.8).

Summary

Respondents were more likely to install the app during the interview if they completed the survey with a face-to-face interviewer than if they completed the survey online, but the mode of interview had no causal effect on whether or not respondents actually used the app.

Across all treatment groups, 36.5% of eligible respondents said they installed the app during the interview, however we only received data from the app for 18.4% of respondents. When we asked respondents who tried but did not manage to install and log into the app during the interview how far they got, 64.5% said they successfully found the app in the app store, but only 31.4% said they installed the app. Unfortunately we do not know the reason for this drop off; whether respondents tried to install the app but failed for technical reasons, or whether they did not like the description of the app in the app store and therefore did not try to install it.

What we told respondents about feedback they would get in the app led to different rates of downloading the app and returning app data, however these differences were not significant. The type of incentive had a small effect on app download during the interview: 38.5% with the conditional incentive, 34.5% with the unconditional; and a large effect on whether we received app data: 14.8% with the unconditional, 22.0% with the conditional incentive. Initial analyses suggest small differences between the experimental treatment groups in the extent of non-participation bias.

Further analyses will examine where in the process of installing and logging in to the app we lost respondents, as well as the reasons respondents gave for not wanting to participate in the app study.

References

Institute for Social and Economic Research. (2023). Understanding Society IP15 Interviewer Materials. UK Data Service.

https://doc.ukdataservice.ac.uk/doc/6849/mrdoc/pdf/6849_ip_wave_15_interviewer_documents.pdf.

University of Essex, Institute for Social and Economic Research. (2023). Understanding Society: Innovation Panel, Waves 1-15, 2008-2022. [data collection]. 12th Edition.

UK Data Service. SN: 6849, <http://doi.org/10.5255/UKDA-SN-6849-15>.

4.8 Can we rely on self-reported biomarker data? Benchmarking the accuracy and feasibility of self-reported hip and waist measurements using analogue and digital methods

Paulo Serodio, Tarek Al Baghal, Annette Jäckle, Jonathan Burton and Meena Kumari

Introduction

Biomarkers play a crucial role in health research, offering an invaluable tool for objectively and precisely measuring various health aspects, particularly in early disease stages.

However, the use of biomarkers in longitudinal surveys presents significant challenges, particularly when collecting data outside of clinical settings. These challenges become even more pronounced when measurements are taken within a participant's home environment, as established protocols for biomarker data collection become difficult to implement.

Nonetheless, the advantages of using biomarkers are substantial and should not be overlooked. Certain measurements, such as waist circumference and waist-to-hip ratio, offer relatively inexpensive and non-invasive ways to detect abdominal obesity, unconfounded by other factors such as height and age, thus providing valuable insights into an individual's risk of developing chronic diseases such as heart disease, stroke, type 2 diabetes, and certain cancers.

Moreover, these measurements can effectively monitor changes in body fat distribution over time, making them a powerful tool for assessing the effectiveness of interventions. Consequently, despite the challenges associated with data collection, the utility of biomarkers in evaluating health risks across diverse demographics is undeniable. The primary challenge lies in finding the balance between ensuring good quality data (perhaps shy of clinical accuracy) and obtaining a representative sample that satisfies the requirements of scientific research.

With surveys increasingly shifting away from face-to-face interviews in many countries, the viability of regularly collecting biomarker data in clinical settings or through home visits for large population samples is being called into question. The adoption of mixed-mode designs

that incorporate online data collection by longitudinal surveys is an example of this shift. This raises concerns about how biomarkers can be collected in such contexts while maintaining consistent measurement properties across different modes and over time.

To address these challenges and prepare for the collection of biomarkers in Wave 16, IP15 (University of Essex, Institute for Social and Economic Research 2023) served as a field trial for potential additions and validations in the mainstage wave. One promising tool that underwent testing during IP15 is a Body Volume Index (BVI) app. This innovative app uses a smartphone's camera to capture data on a person's body shape and employs various algorithms to estimate indicators of body shape. These metrics include measures of body fat percentage, visceral fat content, and waist/hip ratio. Additionally, IP15 incorporated two other modes of collecting the same indicators, allowing for the identification of appropriate and cost-effective approaches to obtaining high-quality biomarker data following three different data collection protocols: by social survey interviewers (McFall et al. 2014), by participants, and via an app equipped with advanced computer vision algorithms.

However, a shift towards self-collection necessitates careful evaluation to ensure that new methods can provide comparable data over time, considering the well-documented biases associated with self-reported height and weight data (Gunnell et al. 2000; Uhrig 2012). Likewise, while leveraging technology for measurements can alleviate respondent burden, it is crucial to benchmark the accuracy of these measurements against objective data obtained from established data collection protocols.

The IP pilot study sheds light on both of these issues by directly measuring biomarker indicators while also incorporating self-measurement within the same interview. This evaluation allows us to assess the merits of the three different protocols for collecting biomarker data and provides valuable insights for future data collection efforts.

Methods

The request for the collection of biomarkers in IP15 was not conducted experimentally but as a pilot to assess the feasibility and quality of different modes of data collection and to validate indicators for other measures in the study.

Consequently, all the biomarker measures mentioned earlier were obtained through direct measurements conducted by interviewers during face-to-face interviews and through self-reported measures from all respondents.

In the initial invitation letter sent prior to the survey, both CAPI (face-to-face) and CAWI (web) respondents received a metric-only tape measure (150cm long) along with instructions and a request to measure their waist and hips themselves, using the tape measure, and record the dimensions on a provided card.

During the interview, CAPI respondents were asked whether they had completed these measurements and, if so, to report their results. Additionally, respondents were asked to measure their waist and hips again under the guidance of the interviewers during the interview. The interviewers did not physically touch the respondents but provided instructions and visually inspected and recorded the results. In previous waves (2 and 3), nurses had collected these measurements but, in this case, interviewers performed the collection.

Respondents were also invited to download a BVI app and use it to capture these metrics themselves. Face-to-face respondents were instructed to download the app during the interview but were asked to use it after the interview or in the absence of an interviewer.

Instructions were provided to the respondents for taking two pictures of themselves, one from the front and one from the side, and uploading them to the app. The app would then use computer vision algorithms to convert the pictures into body shape outlines. These outlines were uploaded to a UK-based server, where proprietary algorithms combined the recreated 3D version of the respondent's body shape with their information (age, sex, height, weight, and activity level) to calculate waist-hip ratio, waist circumference, total body fat, visceral fat, lengths of body parts, and the Body Volume Index (BVI).

In this report, we use the waist and hip circumference measures obtained by the interviewers as reliable reference measurements against which we compare both the participation (compliance) and the properties of waist and hip self-measurements taken by the respondents themselves, using either the metric tape or the app.

While other indicators could be used to compare interviewer-led measurements with app-collected measurements, such as height, weight, fat percentage, and water percentage, this report focuses solely on waist and hip circumference.

Results

Feasibility of self-measurement

The feasibility of conducting biomarker data collection through the web or using an app largely depends on respondent compliance. Regardless of the quality and accuracy of the measurements, if the uptake is significantly lower compared to interviewer-led measurements, we will not achieve the necessary population coverage for meaningful analysis.

When considering self-measurement for waist and hip circumferences, compliance at IP15 could be readily determined by examining the proportion of respondents who received the tape measure by mail and provided the requested measurements. For respondents instructed to use the app, compliance encompasses two aspects: firstly, whether they downloaded the app, and secondly, whether they used it correctly to submit the required measurements. However, even if respondents completed both steps accurately, the receipt of their measurements is contingent on the app's computer vision algorithm successfully generating their body shape based on the captured photos.

Table 4.8-1 presents the compliance rates across the three modes of data collection. The preliminary findings reveal a notably higher compliance rate (93.4%) for self-measurement in the presence of an interviewer, surpassing the rates observed for self-measurement alone by 37 percentage points and the BVI app by 69 percentage points.

Table 4.8-1 Respondent compliance across three waist and hip measurement modules.

	Self-measurement	Self-measurement w/ interviewer	BVI app
Compliance rate	56.4%	93.4%	23.9%
[waist]	(N=2137)	(N=318)	(N=1986)
Compliance rate	54.6%	93.4%	8.9%
[hip]	(N=2137)	(N=318)	(N=1986)

Out of the 1986 respondents who were invited to download and use the BVI app, 34.2% (n=679) reported successful installation and login. However, of those who installed the app we only received waist data from 70% (n=475), and hip data from 26% (n=176). This can be attributed to one of two scenarios: either respondents installed and logged in but failed to upload two body photos, or they did upload the photos, but the quality did not meet the minimum threshold required for the servers to calculate the measurements.

Measurement properties

In Table 4.8-2, we present the average measurement (along with standard deviations) in centimetres for waist circumference, hip circumference and waist-to-hip ratio, obtained through the three modes of data collection: (1) participant self-measurement using a tape measure, (2) measurements conducted with the help of an interviewer, and (3) participant self-measurement using the BVI app.

The BVI app data provides us two sets of measurements: (1) measurements calculated within the app using the uploaded photo of the respondent, and (2) measurements calculated on the server using a 3D avatar generated from the uploaded photo. The differentiation between these two sets of measures is significant, as we find no missing data in app-based measures, while encountering substantial missing data (around 70%) in server-side measures.

In Table 4.8-2 we present ***app-based*** estimates for waist circumference and ***server-side*** estimates for hip circumference, as the calculation of hip circumference solely occurs in the server and is not displayed within the app.

Overall, the BVI app data appears to provide less precise estimates for both waist and hip circumferences when compared to self-measurement, regardless of the presence of an interviewer (the standard deviation for waist is twice as large in the BVI app compared to the other two methods).

The most striking difference is in **waist circumference**: the average waist measurement in the BVI app sample is approximately 23cm wider than the average for respondents who conducted measurements with the interviewer, and almost 30cm wider than those

respondents who conducted self-measurements unaided. These differences cannot be attributed to variations in weight across the samples, as the average weight is 79.3kg for the self-measurement group, 79.4kg for the BVI app group and 79kg for the interviewer-led group, with virtually identical standard deviations. The statistical insignificant differences in average weight across groups for different modes of data collection remains consistent when considering non-participants in both self-measurement and BVI app modules, indicating that self-selection is unlikely to explain the disparities in reported measurements.

Hip circumference exhibits fewer differences across the data collection methods. The BVI app sample still shows slightly higher averages, but the differences are much smaller: approximately a 2cm difference in average hip circumference between the BVI app and interviewer-led samples, and a 6cm difference between the BVI app and self-measurement samples. The standard deviations also do not exhibit significant differences.

Table 4.8-2 Average (and Std. Dev.) waist and hip measures across modules

	Self-measurement	Self-measurement with interviewer	BVI app
Waist circumference in cm	93 (15)	97 (15)	118 (23)
Hip circumference in cm	102 (13)	106 (13)	108 (16)
Waist-to-hip ratio	0.90 (0.11)	0.91 (0.09)	0.93 (0.13)

One important consideration in the BVI data is the presence of a larger number of outliers compared to the other two groups, which could contribute to variations in the reported means. In Table 4.8-2, the average waist circumference for the BVI sample was calculated after excluding respondents whose reported waist circumference exceeded 180cm (n=9). This step was taken to mitigate the potential impact of extreme values on the overall average.

Benchmarking self-measurement against reliable reference measurement

Table 4.8-3 presents the average differences in measurements between each pair of data collection modes for the three measures of interest: waist circumference, hip

circumference, and waist-to-hip ratio. The average differences are reported in centimetres specifically for waist and hip circumferences.

For the purpose of benchmarking against a reference measure, we consider interviewer-aided self-measurement as a reliable reference case for evaluating the quality of data collected in the other modes. While it may not reach the same level of accuracy as measurements taken by nurses in previous survey waves, it represents the best alternative when conducting measurements in a non-clinical setting.

Consistent with earlier findings, the BVI app appears to produce the least accurate data among the three modes of measurement. The average difference in waist circumference between app users and respondents who received assistance from an interviewer is 18cm. Additionally, the average difference in waist circumference between respondents who self-measured using a tape measure and those who used the BVI app is approximately 27cm. Although the sample size for the former comparison is small (60 respondents), it is not the case for the latter (N=292).

Similar discrepancies are observed in hip measurements between the BVI app and tape self-measurements. The average difference in hip circumference is 9.1cm when an interviewer **is involved** and 11.2cm **without** the presence of an interviewer.

Regarding waist-to-hip ratios, the bias introduced by the BVI app is twice as high as that of self-measurements without an interviewer present when compared to interviewer-led measurements (0.09 vs. 0.04).

Table 4.8-3 Average differences (in cm) between self-measurement with tape, self-measurement with interviewer (with tape), and BVI app measurements. Number of respondents in brackets.

Waist			Hip			Waist/Hip ratio		
	Self	Intvr		Self	Intvr		Self	Intvr
Self	-	-	Self	-	-	Self		
Intvr	3.1 (158)	-	Intvr	3.4 (155)	-	Intvr	0.04 (155)	
BVI	26.5 (292)	18.4 (60)	BVI	11.2 (118)	9.1 (34)	BVI	0.12 (288)	0.09 (60)

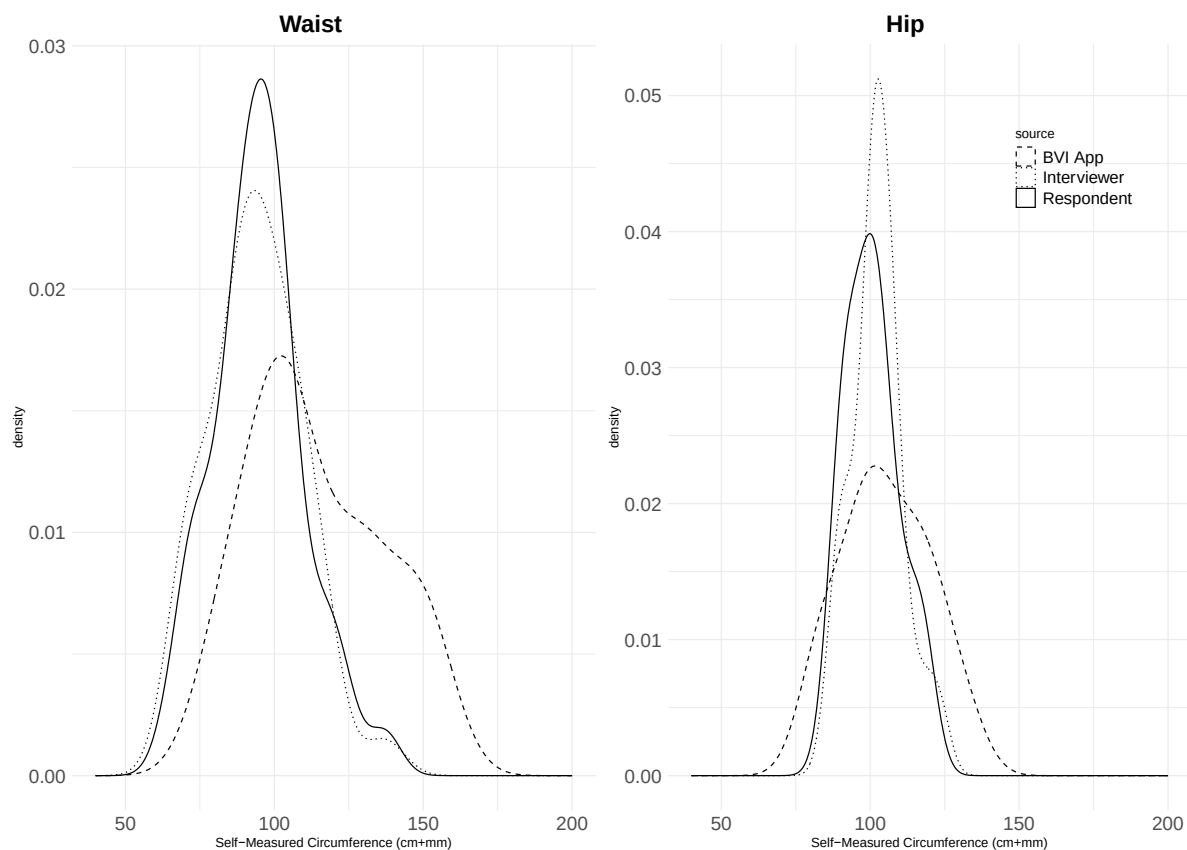
Note: Waist/hip ratio data from the BVI app was available for some respondents for whom hip measurements themselves were not available. Hip measurements were only available from the subset of app users where the server-side image processing succeeded, but waist/hip ratios were also available for some app users from in-app (i.e., non-server) assessments.

It is important to note that the comparisons presented earlier use the entire sample of respondents who participated in each mode of data collection. This approach may introduce potential biases as the sample mean can be affected by differential sample sizes across groups, which may artificially inflate some of the observed differences.

However, in Figure 4.8-1, we address this issue by plotting the distributions of waist and hip circumferences (in cm) specifically for respondents who participated in all three modes of data collection and from whom we were able to obtain usable data (N=41 for waist and 21 for hip).

Examining the distributions, we observe that for both hip and waist circumferences, the measurements from the BVI app display a larger variance compared to the other two modes of data collection. This difference is more pronounced for the waist distribution. This observation further supports a hypothesis that measurements estimated by the server, benefiting from higher processing power and more sophisticated algorithms, may tend to be more precise than those calculated and reported within the app itself.

Figure 4.8-1 Distribution of waist and hip circumferences for respondents who participated in all three modes of biomarker data collection



Conclusion

In conclusion, the BVI app displays noticeable issues regarding data accuracy, particularly in waist measurements. Despite the fact that 679 respondents were able to successfully install and log into the app, no data was returned from about a third of this group. A consistent bias has been observed in self-measurements compared to the app data, with the BVI app showing higher inconsistencies, displaying a bias three times higher for waist than for hip measurements.

It is important to highlight that these inaccuracies are notably affected by the low quality of the photos uploaded onto the app, possibly attributable to factors such as suboptimal picture taking conditions. A majority of the respondents failed to capture images under the best possible conditions, contributing to the lower quality data. The debrief questions asked in the app, after respondents took the photos of themselves, suggest that a mere 17.3% of the valid BVI app data sample took pictures with an empty, white wall in the background (as

per the instructions), a factor that might significantly influence the accuracy of the measurements.

This reflects the additional challenge of implementing this sort of study in a population survey, where there is less control over the environment in which the app is being used. This app may perform well in clinical situations, or where the researchers have greater control over the location of the study, or in cases where the person using the app has sought out and chosen to use the app for their own purpose. As part of a larger study where users have been asked to carry out an additional task, and are not intrinsically motivated in that task, and having to use the app in their own home, it appears to be less successful.

Moving forward, it will be crucial to address these issues to improve the reliability and effectiveness of app-based measurements.

References

- Gunnell, D., Berney, L., Holland, P. et al. (2000). How accurately are height, weight and leg length reported by the elderly, and how closely are they related to measurements recorded in childhood? *International Journal of Epidemiology* 29: 456– 464.
- Kumari, M., and Benzeval, M. (2021). "Collecting biomarker data in longitudinal surveys." *Advances in Longitudinal Survey Methodology* (2021): 26-46.
- McFall, S., Conolly, A., and Burton, J. (2014). Collecting biomarkers using trained interviewers. Lessons learned from a pilot study. *Survey Research Methods* 8: 57– 65.
- Uhrig, S. (2012). Understanding panel conditioning: an examination of social desirability bias in self reported height and weight in panel surveys using experimental data. *Longitudinal and Life Course Studies* 3: 120– 136.
- University of Essex, Institute for Social and Economic Research. (2023). Understanding Society: Innovation Panel, Waves 1-15, 2008-2022. [data collection]. 12th Edition. UK Data Service. SN: 6849, <http://doi.org/10.5255/UKDA-SN-6849-15>.

4.9 National identity choice and meaning over time: role of affective priming and question order

Ivelina Hristova

Evidence suggests strong and deepening sense of English and Scottish identity that does not always exist alongside a British identity (Bond 2015; NatCen 2013; Wyn Jones et al. 2012). While national identity needs to be periodically redefined to reflect past experiences, present social reality and future aspirations (Parekh 1995), longitudinal quantitative research is limited.

Survey data collection is essential to provide measurement of such changes and meanings. Questions in domains such as nationality need a careful matching between the intended concept and the question wording (Burton, Nandi, and Platt 2010). For example, changing the response order in the national identity question in the 2021 Census in England such that British was listed before English might account for a large part of the reported increase in British-only identity and decrease in English-only identity, suggesting a reported change that exceeds true change (ONS 2022). An ascribed interchangeability in the meaning of these two identities can co-exist with a true identity shift. However, better understanding of the matching of the intended national identity concept and the survey question is needed for a more robust interpretation.

Combining different identity measures and positioning of the questions might help provide deeper insight into the meaning respondents ascribe to listed identity (Burton et al. 2010). For example, placing the national identity question before the ethnic group question in the UK 2011 Census aimed to give ethnic minority respondents an initial opportunity to assert their Britishness and ascribe a meaning to their ethnic group beyond the Britishness aspect. Similarly, self-reported national identity measures might be combined with affective measures to be able to differentiate between a strictly legal and a more affective conceptualisation of being British at the analytical stage (Georgiadis and Manning 2013). Yet there is limited research on the interaction of affective and reported national identity measures in the UK context.

A random split-ballot experiment is best suited to address these gaps (Saris, Satorra, and Coenders 2004; Thau et al. 2021; van de Walle and van Ryzin 2011). Using a combination of cross-sectional and longitudinal analysis, I test whether affective priming and question order of a set of national identity measures result in differences in reported identity choice and ascribed meaning.

Experiment, methods and data

Households were randomly allocated to two equally sized groups before the start of fieldwork. As the measurement error decreases with an increasing sample size, the identity questions were asked of the whole sample, except new entrants in IP15. The control group was asked a multiple-choice national identity question with options English, Welsh, Scottish, Northern Irish, British, Irish and Other, followed by a question on the importance of being British measured on a scale between 0 and 10 (Not at all important – Extremely important). The intervention group was asked of the importance of being British before being asked the self-reported national identity question. Evaluating the importance of being British for oneself reveals a value judgement and attachment and is considered an affective measure.

Using Pearson's χ^2 and one-sided Fisher's tests, I compare the reported identity distributions between the two groups in the split-ballot experiment in IP15 and a refreshment sample in IP11, within the respective time points. The absence of analytical weights in the early data release and the sensitivity of the statistical tests to sample size might affect the results. I perform bivariate logistic and multinomial regressions to test how well the experimental group allocation predicts the reported national identity, adding robustness to the results. The cross-sectional analysis in IP15 provides information on the role of affective priming on self-reported identity choice and meaning between the two experimental groups within the same time point. The analysis in IP11 compares the two groups before the experiment was introduced.

Finally, I perform Pearson's χ^2 and one-sided Fisher's tests and a set of repeated measures bivariate logistic regressions to compare each experimental group between IP15 and IP11. The longitudinal setting allows us to test whether the intervention influences the reported identity and the associated ascribed meaning over time.

I define UK-country identity as an identity that relates to one of the four UK nations, and British identity as an identity that relates to Britain. In the context of the UK as a multi-nation state, all of these identities are national identities.

I restrict the sample to those respondents who reported no more than one UK-country identity regardless of reporting British identity alongside it or not, accounting for evidence that UK-country and British identities might co-exist (Bond 2015; NatCen 2013; Wyn Jones et al. 2012). In IP15, I exclude Northern Irish and Irish identities because the groups sizes are too small to allow for a comparison.

IP11 is best suited for the longitudinal analysis. IP11 has a relatively big sample size with respect to the national identity variables, is recent enough to expect the sample attrition to allow for a longitudinal analysis, and the data is collected four years before the IP15 data providing a time span for identity change to happen (Jäckle *et al.*, 2021; University of Essex, Institute of Social and Economic Research 2023).

In IP11, I restrict the sample to those new entrants who participated in the experiment and gave valid national identity responses in IP15. Due to sample attrition and the resulting small cells, I further exclude Welsh and Scottish identities. The intervention (N=1,157 in IP15; N= 136 in IP11) and control (N=1,211 in IP15; N= 131 in IP11) groups are slightly different in size due to random non-response. Non-response is not a major concern for the measurement of identity (Nandi and Platt 2017).

I construct separate measures for English, Welsh, Scottish and British identities in IP15, and English and British identities in IP11. The sample size allows the construction of a three-category measure that captures combinations of these identities only in IP15, with options UK-country only, British and UK-country, and British only. Binary measures account for group allocation and wave.

Discussion of results

Priming with importance of being British results in a lower share and likelihood of reported English and Welsh identities and a higher share and likelihood of reported Scottish identity in the intervention group, compared to the control group in IP15 (Table 4.9-1). In line with

previous observations with respect to option order (ONS 2022) and question order (Burton et al. 2010), the experimental results show that enquiring of the British identity before another related identity changes the way people report that second identity.

Table 4.9-1 National identities: introduction of priming
IP15 PANEL A: Statistical tests

	English	Welsh	Scottish	British	UK- country only	British & UK- country	British only
G1: Intervention							
(%)	64%	3%	7%	49%	51%	23%	26%
(n)	739	39	81	568	589	270	298
G2: Control							
(%)	68%	4%	6%	43%	57%	21%	22%
(n)	827	45	73	525	686	259	266
Pearson chi2							
value	5.158	0.206	0.921	7.844		8.197	
p-value	0.023	0.650	0.337	0.005		0.017	
One-sided Fisher's exact, p-value	0.013	0.366	0.190	0.003			
IP15 PANEL B: Regression analysis							
G1: intervention	-0.197	-0.101	0.160	0.231	1	0.194	0.266
p-value	0.023	0.650	0.338	0.005		0.061	0.009

Source: IP15. Notes: Analytical weights not applied. Pearson's Chi-square degrees of freedom: 1 for single UK-country and British identities, 2 for the composite measure. Shares of single UK-country and British national identities add to more than 100% because some respondents reported UK-country identity together with a British identity.

With respect to UK-country identities, these results are only significant for the English identity. The lack of statistical significance for Welsh and Scottish identities is inconclusive. The samples might be too small for the results to be statistically significant, there might be no significant difference between the experimental groups, or both. However, regardless of whether the results for Scottish identity are significantly different between the two experimental groups with an opposite sign to the results for English and Welsh, or not significant at all compared to the high significance level for the English, it seems a plausible interpretation that being Scottish has a different relationship to being British, compared to

being English or Welsh. Similarly, Welsh identity might be different from English identity in that respect. Therefore, adding the affective Britishness question before the reported national identity question reveals nuanced differences between the separate UK-country identities and their relationship to Britishness. This is in line with the suggestion made by Georgiadis and Manning (2013) that combining affective and nominal identity measures might provide grounds for more in-depth substantive understanding of the content of national identities.

Priming with importance of being British also leads to a higher share and likelihood of reporting British identity. The sample is highly dominated by the English, in line with the geographical split of the British population. Within a limited scope of uncertainty, this implies that respondents in the intervention group treated English and British identities as interchangeable. Similarly, respondents in the intervention group seem more likely to report either a combination of their UK-country and British identities, or a British only, compared to a single UK-country only identity. The higher percentage difference, magnitude and significance level for reporting British only identity compared to the combination of both further suggests that English and British identities can be interchangeable, at least to some extent. Therefore, positioning the importance of being British question before the reported national identity question has a similar effect as the change of choice order in the latest census question (ONS 2022).

The difference in the distribution of reported identities between the two experimental groups might also imply that respondents in the two groups ascribe different meanings to the national identity concept being measured. Indeed, the cross-sectional analysis at IP11 showed that the experimental groups were not significantly different with respect to the reported national identities before the introduction of the intervention (Table 4.9-2). Therefore, priming respondents to think about the importance of being British before asking them to list their national identities might have affected the meaning that they ascribe to these identity categories. In such a case, the distributions of the two experimental groups reflect the distributions of identity concepts that carry different contents. Such a conclusion further reiterates the need for a careful matching between the survey questions and intended concept, as discussed in Burton, Nandi, and Platt (2010). Clarity about that matching is essential to ensure that, first, the survey questions measure the intended

national identity concept, and second, such measurement is correctly interpreted at the stage of analysis.

Table 4.9-2 National identities: comparison before priming

IP11 PANEL A: Statistical tests

	English	British
G1: Intervention		
(%)	63%	53%
(n)	86	72
G2: Control		
(%)	62%	59%
(n)	81	77
Pearson chi2		
value	0.056	0.922
p-value	0.813	0.337
One-sided Fisher's exact, <i>p</i> -value	0.456	0.201

IP11 PANEL B: Regression analysis

G1: Intervention	0.060	-0.237
<i>p</i> -value	0.813	0.337

Source: IP11. Notes: Analytical weights not applied. Pearson's Chi-square degrees of freedom: 1. Shares of identities add to more than 100% because some respondents reported English identity together with a British identity.

For example, the results from the longitudinal analysis indicate that respondents in both groups report a shift towards English identity over time (Table 4.9-3). That shift is statistically significant for both experimental groups, with a stronger percentage change and coefficient magnitude, and a higher level of significance for the control group. Therefore, priming respondents with the importance of being British slightly lessens a potentially true shift towards reporting English identity but does not prevent it. Such a trend is in line with previous research (NatCen 2013; Wyn Jones et al. 2012) and in the opposite direction to the reported shift in the latest census data (ONS 2022). Comparing the two experimental groups in the innovation panel data reiterates the conclusion that the change in order option in the census question might explain a large part of the reported shift there. Moreover, building on the discussion above, the differently phrased questions in the two datasets might have prompted respondents to ascribe different meanings to the intended national identity concepts. Therefore, even though the two datasets suggest two opposite trends with

respect to English identity, the respective data might reflect true notions of national identities, albeit such that carry different meanings and are not directly comparable.

Table 4.9-3 National identities: priming and change over time
IP11-15 PANEL A: Statistical tests

Experimental group	English		British	
	G1: Intervention	G2: Control	G1: Intervention	G2: Control
IP15				
(%)	71%	76%	51%	46%
(n)	739	827	534	502
IP11				
(%)	62%	60%	53%	59%
(n)	91	81	78	79
Pearson chi2				
value	4.880	15.300	0.191	7.6404
<i>p</i> -value	0.027	<0.001	0.662	0.006
One-sided Fisher's				
exact, <i>p</i> -value	0.019	<0.001	0.364	0.004
IP11-15 PANEL B: Regression analysis				
IP15	0.351	0.709	-0.127	-0.494
<i>p</i> -value	0.038	<0.001	0.440	0.004

Source: IP11 and IP15. Notes: Analytical weights not applied. Pearson's Chi-square degrees of freedom: 1.

Shares of single UK-country and British national identities add to more than 100% because some respondents reported UK-country identity together with a British identity.

Finally, respondents in both experimental groups are less likely to report a British identity in IP15 compared to IP11. Again, there is a smaller percentage change and coefficient magnitude for the intervention group compared to the control group, suggesting that the shift of identity is more pronounced in the absence of the intervention. However, the shift in the intervention group is not statistically significant, while the shift in the control group is. At the same time, the cross-sectional analysis discussed above suggested that the

intervention results in a higher likelihood of reporting British identity, compared to the control group. Therefore, priming respondents to think of the importance of being British seems to offset the decline of identification with Britain to an extent that there is no statistically significant difference between the two time points for the intervention group. Such a finding only further reiterates the conclusion that affective priming and question order help shape the meaning that respondents ascribe to the pre-defined national identity categories. Therefore, the longitudinal findings reinforce the case for a careful matching between the intended concept and the measure (Burton et al. 2010).

Conclusions and next steps

Preliminary results support the hypothesis that affective priming and question order result in differences in reported identity between groups and over time, suggesting differences in the meaning respondents ascribe to pre-defined categories. This is in line with observations made by Burton, Nandi, and Platt (2010) and Georgiadis and Manning (2013). Moreover, the experimental results strongly support the need for a careful matching between the measure and the intended concept (Burton et al. 2010). As shown above, such considerations are essential to distinguish at the analytical stage whether a shift towards or away from reporting a specific national identity reflects a change in data collection, a true shift towards or away from a previously defined notion of that identity, a change in the meaning of that concept, or a combination of these. To truly capture the ways society redefines national identities to reflect past and current social realities and future aspirations (Parekh 1995), a clear understanding of the matching between the measure and intended national identity concept is essential.

The results from the identity experiment can inform future data collection, for example the design of the rotating identity modules in the Understanding Society dataset. The results can also support researchers in the interpretation of self-reported identity categories, feeding into evidence for public policy decision making.

Alongside standard demographic and socio-economic characteristics, the identity experiment included questions on the importance of ethnic background, political affiliation and voting behaviour. Using analytical weights, future research will focus on a deeper

understanding how affective priming and question order help shape and reveal the content respondents ascribe to reported national identities.

References

- Bond, Ross. 2015. "National Identities and the 2014 Independence Referendum in Scotland." *Sociological Research Online* 20(4):92–104.
- Burton, Jonathan, Alita Nandi, and Lucinda Platt. 2010. "Measuring Ethnicity: Challenges and Opportunities for Survey Research." *Ethnic and Racial Studies* 33(8):1332–49.
- Georgiadis, Andreas, and Alan Manning. 2013. "One Nation under a Groove? Understanding National Identity." *Journal of Economic Behavior and Organization* 93:166–85.
- Jäckle, Annette, Tarek Al Baghal, Jonathan Burton, Olena Kaminska, and Peter Lynn. 2021. *Understanding Society The UK Household Longitudinal Study Innovation Panel, Waves 1-13, User Manual*.
- Nandi, Alita, and Lucinda Platt. 2017. "Are There Differences in Responses to Social Identity Questions in Face-to-Face versus Telephone Interviews? Results of an Experiment on a Longitudinal Survey." *International Journal of Social Research Methodology* 20(2):151–66.
- NatCen. 2013. *British Social Attitudes 30 | Trends in National Identity*. London.
- Office for National Statistics (ONS). 2022. "National Identity, England and Wales: Census 2021." Retrieved May 26, 2023 (<https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/nationalidentityenglandandwales/census2021>).
- Parekh, Bhikhu. 1995. "The Concept of National Identity." *Journal of Ethnic and Migration Studies* 21(2):255–68.
- Saris, Willem E., Albert Satorra, and Germà Coenders. 2004. "A New Approach to Evaluating the Quality of Measurement Instruments: The Split-Ballot MTMM Design." *Sociological Methodology* 34(1):311–47.
- Thau, Mads, Maria Falk Mikkelsen, Morten Hjortskov, and Mogens Jin Pedersen. 2021. "Question Order Bias Revisited: A Split-ballot Experiment on Satisfaction with

Public Services among Experienced and Professional Users.” *Public Administration* 99(1):189–204.

University of Essex, Institute for Social and Economic Research. (2023). Understanding Society: Innovation Panel, Waves 1-15, 2008-2022. [data collection]. 12th Edition. UK Data Service. SN: 6849, <http://doi.org/10.5255/UKDA-SN-6849-15>.

van de Walle, Steven, and Gregg van Ryzin. 2011. “The Order of Questions in a Survey on Citizen Satisfaction with Public Services: Lessons From a Split-Ballot Experiment.” *Public Administration* 89(4):1436–50.

Wyn Jones, Richard, Guy Lodge, Ailsa Henderson, and Daniel Wincott. 2012. *The Dog That Finally Barked: England as an Emerging Political Community* | IPPR.