



SURVEY FUTURES

SURVEY DATA COLLECTION
METHODS COLLABORATION

Working Paper 18:

Assessment of a knock-to-nudge recruitment strategy to improve participation in a telephone survey: Evidence from the National Survey for Wales

Olga Maslovskaya¹, Cristian Domarchi¹,
Peter W. F. Smith¹

¹ University of Southampton

June 2026

www.surveyfutures.net

Survey Futures is an Economic and Social Research Council (ESRC)-funded initiative (grant grant ES/X014150/1) aimed at bringing about a step change in survey research to ensure that high quality social survey research can continue in the UK. The initiative brings together social survey researchers, methodologists, commissioners and other stakeholders from across academia, government, private and not-for-profit sectors. Activities include an extensive programme of research, a training and capacity-building (TCB) stream, and dissemination and promotion of good practice. The research programme aims to assess the quality implications of the most important design choices relevant to future UK surveys, with a focus on inclusivity and representativeness, while the TCB stream aims to provide understanding of capacity and skills needs in the survey sector (both interviewers and research professionals), to identify promising ways to improve both, and to take steps towards making those improvements. *Survey Futures* is directed by Professor Peter Lynn, University of Essex, and is a collaboration of twelve organisations, benefitting from additional support from the Office for National Statistics and the ESRC National Centre for Research Methods. Further information can be found at www.surveymfutures.net.

Prior to citing this paper, please check whether a final version has been published in a journal. If so, please cite that version. In the meanwhile, the suggested form of citation for this working paper is:

Maslovskaya O, Domarchi C & Smith PWF (2026) 'Assessment of a knock-to-nudge recruitment strategy to improve participation in a telephone survey: Evidence from the National Survey for Wales', *Survey Futures Working Paper* no. 18. Colchester, UK: University of Essex. Available at <https://surveymfutures.net/working-papers/>.

Contents

| | |
|--|----|
| Abstract | 6 |
| 1 Introduction | 7 |
| 2 Data | 8 |
| 2.1 The National Survey for Wales | 8 |
| 2.1.1 Fieldwork | 9 |
| 2.1.2 Knock-to-nudge protocol | 9 |
| 2.1.3 Additional tasks: Online modules for sensitive questions | 10 |
| 2.2 Dataset description | 10 |
| 3 Research questions and hypotheses | 11 |
| 3.1 RQ1: Is the inclusion of the knock-to-nudge stage associated with improved survey response in the National Survey for Wales? | 11 |
| 3.2 RQ2: Is the knock-to-nudge stage associated with improved sample composition in the National Survey for Wales? | 12 |
| 3.3 RQ3: Do participants recruited via knock-to-nudge differ from those recruited during the initial stage in their substantive responses? | 12 |
| 3.3 RQ4: Do participants recruited via knock-to-nudge provide data of comparable quality to those recruited at the initial stage? | 12 |
| 3.4 RQ5: Are participants recruited via knock-to-nudge less likely to participate in an additional task (online modules)? | 13 |
| 4 Methods | 14 |
| 4.1 Response rates | 14 |
| 4.2 Sample composition | 15 |
| 4.2.1 Comparison with issued sample | 15 |
| 4.2.2 Comparison with population benchmarks | 16 |
| 4.3 Substantive data | 17 |
| 4.4 Item non-response | 17 |
| 4.5 Participation in the additional online modules | 18 |
| 5 Results | 18 |
| 5.1 Response rates | 18 |
| 5.2 Sample composition | 20 |
| 5.2.1 Comparison with issued sample | 20 |
| 5.2.2 Comparison with population benchmarks | 21 |
| 5.3 Substantive data | 23 |
| 5.4 Item non-response | 25 |
| 5.5 Participation in online modules | 28 |

| | | |
|---|-----------------|----|
| 6 | Discussion..... | 29 |
| | References..... | 30 |
| | Appendix..... | 34 |

Assessment of a knock-to-nudge recruitment strategy to improve participation in a telephone survey: Evidence from the National Survey for Wales

*Olga Maslovskaya*¹ (corresponding author: om206@soton.ac.uk), *Cristian Domarchi*¹, *Peter W.F. Smith*¹

¹ Department of Social Statistics and Demography, University of Southampton, University Road, Southampton, SO17 1BJ, United Kingdom.

Authors

Olga Maslovskaya is Professor of Survey Methodology and Social Statistics at the Department of Social Statistics and Demography, University of Southampton, University Road, Southampton SO17 1BJ, UK.

Cristian Domarchi is a Research Fellow at the Department of Social Statistics and Demography, University of Southampton, University Road, Southampton SO17 1BJ, UK.

Peter W. F. Smith is Professor of Social Statistics at the Department of Social Statistics and Demography, University of Southampton, University Road, Southampton SO17 1BJ, UK

Acknowledgements

The work was supported by the UK Research and Innovation (UKRI)/Economic and Social Research Council (ESRC) research grant ES/X014150/1 for “Survey data collection methods collaboration: securing the future of social surveys”, known as Survey Futures. Survey Futures is directed by Professor Peter Lynn, University of Essex, and is a collaboration of twelve organisations, benefitting from additional support from the Office for National Statistics and the ESRC National Centre for Research Methods. Further information can be found at www.surveyfutures.net. We would also like to extend our sincere gratitude to the Welsh Government for facilitating access to the data.

Abstract

Knock-to-nudge (KtN) is a recruitment method in which interviewers visit sampled households and encourage participation in a later non-face-to-face survey. Originally introduced during the COVID-19 pandemic, KtN has since been adopted by several UK surveys in the post-pandemic context. Limited evidence suggests that it increases response rates and improves sample composition. However, its effects on data quality, substantive variables, participation in additional tasks, and the optimal number of KtN visits remain underexplored. Using data from the 2022-2023 National Survey for Wales, we examine KtN visits to non-responding addresses. We find that KtN stage in recruitment significantly increases response rates and improves sample composition by recruiting younger participants, respondents from non-White British backgrounds, renters, individuals without formal educational qualifications, and residents from larger households. These gains are concentrated in the first KtN visit, suggesting diminishing returns from subsequent visits and that a single visit may be sufficient. KtN respondents differ in substantive answers, indicating improved representativeness but show higher item non-response, more “don’t know” responses and lower participation in additional online modules, suggesting lower engagement. Overall, KtN may be an effective recruitment tool if gains in participation and representativeness outweigh increased costs, operational demands and potential reductions in data quality.

Keywords: knock-to-nudge, telephone survey, survey recruitment, sample composition, data quality

1 Introduction

Declining response rates, increasing survey costs, and increased internet penetration and mobile device ownership have contributed to reducing the predominant role of face-to-face interviewing in survey practice. The shift from face-to-face to mixed-mode surveys, including modes without field interviewers, has been encouraged by survey methodologies (Dillman, 2020; Luijkx et al., 2021; Wolf et al., 2021), and the COVID-19 pandemic reinforced the need for innovative approaches to adapt to rapidly changing data collection landscape and unforeseen events (Eurostat, 2020; Kastberg & Siegler, 2022).

In the UK, most high-quality cross-sectional surveys employ address-based sampling frames, which contain no information about individual household residents. Recruitment typically involves sending advance letters to sampled addresses, inviting potential respondents to participate. For surveys without field interviewers (including online and telephone surveys), the initial letters and subsequent reminders provide details on how to access the questionnaire, usually via web links, QR codes, or telephone helplines. This recruitment method involves no in-person contact between an interviewer and a potential respondent. This lack of in-person contact is often mentioned as one of the factors contributing to lower response rates observed in surveys without field interviewers, compared to in-person surveys (Daikeler et al., 2020; Millar & Dillman, 2011).

Knock-to-nudge (KtN) methods were implemented in response to the COVID-19 pandemic and the inability to conduct in-person interviews inside respondents' homes. KtN can be broadly defined as a contact method in which trained interviewers visit sampled households and encourage respondents at the doorstep to participate in a non-face-to-face survey at a later date (Domarchi et al., 2025). Either an appointment is made for the survey, which is conducted subsequently, usually by telephone, video interview, or a combination of these modes, or respondents are asked to complete an online survey (Cornick et al., 2022; Domarchi et al., 2025; Kunz et al., 2024). In early implementations during the COVID-19 pandemic, surveys that shifted from face-to-face to telephone administration instructed KtN visits to all sampled addresses in order to encourage residents to participate in the survey. However, the more common approach among surveys still using KtN is to send the usual invitation and reminder letters and then conduct KtN visits only for some or all non-responding addresses (Domarchi et al., 2025).

Although evidence in the literature remains limited, preliminary findings suggest that KtN methods are effective in increasing response rates and improving sample composition (Cleary, 2023; Domarchi et al., 2025; Elis et al., 2024; Kastberg & Siegler, 2022; Kunz et al., 2024). This is supported by recent research from Maslovskaya et al. (2025), using data from the UK National Readership Survey (PAMCo), which reports that KtN recruitment improves participation and representativeness, particularly by incorporating groups with lower educational levels and lower socio-economic status. While they also found that participants recruited via KtN show higher levels of item non-response than those recruited during the initial stages of the survey, suggesting lower data quality, there is a clear indication that KtN respondents provide significantly different responses to substantive questions. This suggests that KtN is effective in including participants from varied backgrounds who may otherwise be excluded from surveys and as a result improving inclusivity (Maslovskaya et al., 2025).

As KtN remains a novel recruitment approach for surveys without field interviewers, beyond Maslovskaya et al. (2025) analysis, detailed analyses of KtN remain limited. There is therefore

an urgent need for further research to better understand the effectiveness of this novel recruitment method. Specifically, there is still limited evidence of KtN on survey participation, sample composition, data quality, and substantive results in surveys in general and specifically in surveys like the NSW, where participants do not have a choice of mode for each question (although different modes (telephone and online) may be used for different modules of the questionnaire). Similarly, while Maslovskaya et al. (2025) report an aggregate increase in response rates following KtN implementation, the lack of information on non-respondents means these gains cannot be disaggregated to indicate which geographic areas or socio-demographic groups were most affected. There is no indication of the number of KtN visits required to maximise improvements in response and sample composition, while optimising fieldwork costs. Finally, there is no evidence on whether KtN influences the likelihood of participating in additional tasks.

This paper addresses these gaps in evidence and expands the literature on KtN recruitment method. In this paper, we analyse the potential effectiveness of a KtN strategy for contacting participants in the 2022-23 National Survey for Wales (NSW), a survey conducted by telephone followed by additional online modules, which respondents are encouraged to complete on their own within the next three days (Jenkins & Forbes, 2024). The NSW contact protocol involved an initial mailing of advance letters and survey leaflets, followed by a single reminder postcard sent to each sampled address. Participants were asked to provide their telephone number via an online portal, accessed using a unique code from the mailed materials. Where no telephone number could be obtained, KtN visit(s) were carried out: field interviewers visited sampled addresses to request a contact number and collect relevant paradata (see Section 2.1.3) for the fieldwork process. This paper assesses the association of the KtN strategy with response rates and sample composition. It also evaluates differences between nudged and non-nudged respondents in terms of substantive responses, item non-response, and participation in additional online modules containing sensitive questions.

The remainder of the paper is structured as follows. Section 2 outlines the design and sampling strategy of the NSW as well as details of the data collection. Sections 3 and 4 present the research questions and the analytical strategy and methodology used in the analysis. Section 5 details the findings, and Section 6 discusses their significance, implications for survey practice, and limitations, concluding with recommendations for future research.

2 Data

2.1 The National Survey for Wales

The NSW is a survey commissioned by the Welsh Government. The survey consists of over 50 main modules designed to cover a wide range of public policy and quality-of-life topics. Key themes include Welsh language use, digital access, employment, education, housing, and material deprivation, along with extensive demographic information on identity, ethnicity, and household composition. The survey also includes modules on health and wellbeing, including diet, physical activity, healthcare access, civic engagement, and participation in cultural, sporting, and community activities. More information can be found in Jenkins and Forbes (2024).

The NSW is a repeated cross-sectional survey conducted annually. It is designed to be representative of all adults aged 16 or over living in private households in Wales. The annual

sample is therefore designed as a stratified, single-stage random selection of addresses across Wales. The sample is drawn from the Royal Mail Small Users Postcode Address File (PAF) and is stratified by local authority using an allocation method designed to ensure a minimum effective sample size in each local authority, based on estimated response rates. Some areas were over-sampled (e.g., Powys) relative to their population size to achieve samples large enough to ensure robust estimates at different geographical levels.

Within each sampled household, an individual interview was attempted with one adult aged 16 or over. Where a household contained more than one adult, a single adult was selected using the “next birthday” method (Nicolaas, 2002; Salmon and Nichols, 1983). Jenkins and Forbes (2024) report that in 92.2% of the cases the respondent was selected correctly using this method.

2.1.1 Fieldwork

The 2022-23 NSW was conducted mainly via telephone with some sensitive questions asked in additional online modules. Fieldwork started at the beginning of April 2022 and ended on 31 March 2023 and was conducted by the UK Office for National Statistics (ONS). Advance letters and survey leaflets were sent by the ONS approximately 10 days prior to the start of each monthly field period. Field interviewers were responsible for sending out follow-up postcards to the addresses in their assigned quotas. This ensured that delivery of the materials could be timed to coincide closely with when interviewers planned to visit this address, and the postcard could be personalised with the name and telephone number of each interviewer (Jenkins and Forbes, 2024).

Several methods were used to obtain telephone numbers from sampled addresses. The communication materials encouraged respondents to provide their number on an online portal which they could enter using a unique access code provided in their advance letter. In addition, the follow-up postcard often contained contact details (name and telephone number) for the interviewer, so that participants could potentially call this number and provide their phone number directly to the interviewer. Finally, the whole sample was also tele-matched, a process by which telephone numbers were obtained from a commercially available source.

Where a telephone number for an address was obtained, interviewers called at different times of day and on different days of the week, including weekends and evenings. Interviewers aimed to make a minimum of six calls during a mainstage month, with at least two calls on a weekday evening (after 6pm) or at the weekend, before cases could be coded as non-contact. There was no maximum limit on the number of telephone calls with interviewers being instructed to continue to call while they felt there was still a chance of making contact.

2.1.2 Knock-to-nudge protocol

Where a telephone number had not been obtained or contact had not been established, KtN was carried out. In this process, interviewers visited the sampled addresses to make contact with potential respondents to obtain a telephone number. KtN also allowed interviewers to collect relevant fieldwork paradata during their visit. Specifically, interviewers were instructed to undertake the following activities at each address:

1. To check whether the address was ineligible or unoccupied.
2. If occupied, to confirm with a resident that they had called at the correct address.
3. To record the number of dwelling units at the address. If more than one dwelling was present (for example, if the property had been converted into flats), interviewers informed the office to update the sampling frame.
4. To record the number of households. If more than one household was present, a Kish Grid was used to randomly select one.
5. To select a respondent from the household using the “next birthday” method.
6. To gain parental permission if the selected respondent was aged 16 or 17.
7. To record the respondent’s preferred interview language (Welsh or English).

If an appointment for a telephone interview was made during the visit, the interviewer would leave an appointment card at the address. If no contact was achieved, a “called today” card was left instead. After three KtN visits, addresses were recorded as non-responding.

2.1.3 Additional tasks: Online modules for sensitive questions

Following the completion of the telephone survey, interviewers provided the respondents with a link to online modules with sensitive questions, along with a unique 12-digit access code. The respondents were expected to respond to the additional short online component, which was about 10 minutes long, within the three days. They could also request to be asked the questions for the online component via telephone (only 13.1% of those who responded to the online component did so by phone). The online component included seven additional modules, covering topics such as local services and facilities, climate change behaviours, recycling, a mental wellbeing scale, and attitudes towards gambling and volunteering.

The interviewer remained on the telephone until the respondent confirmed they had received the link and access code. If the respondent did not have access to the internet, they were automatically offered the online questions over the telephone. The incentives were only provided after this additional section had been completed.

Interviewers were informed daily about respondents who had not completed the online survey. Initially, interviewers followed up with non-respondents by text (two days after the telephone interview), then by email (three days after the telephone interview), and finally by telephone (five days after the interview). In December 2022, this fieldwork procedure was modified to allow interviewers to choose the order in which they contacted online component non-respondents, based on their own feedback.

2.2 Dataset description

Our analysis uses two linked datasets:

- 1) an internal administrative fieldwork dataset provided by the Welsh Government, containing variables that describe the KtN fieldwork procedures (“*administrative dataset*”);

- 2) and the publicly available 2022-23 NSW dataset, containing full responses for all responding addresses (“NSW dataset”).

The *administrative dataset* includes the full list of 32,464 addresses in the issued sample for the 2022-23 NWS. For each address, it contains a detailed record of response status. This allows identification of whether the address was eligible to respond and, if eligible, whether a full or partial response was obtained. If the household did not respond, the dataset provides a full breakdown of the reasons for non-response. Additionally, the administrative dataset indicates the year and month each address was issued, along with its local authority and whether it was located in an urban or rural area. Each address is also classified by an output area classification (OAC), which groups geographic areas with similar socio-economic and demographic characteristics based on census data. The OAC included in the dataset is based on data from the 2011 Census of England and Wales (Office for National Statistics, 2011). The administrative dataset also includes variables related to the KtN process: specifically, whether each address was part of KtN and, if so, the number of visits conducted. We use the full administrative dataset (32,464 addresses) in our analysis of unit non-response (see Table 1).

The *NSW dataset* is publicly available for download from the UK Data Service website, along with the relevant data dictionaries and technical reports (Office for National Statistics & Welsh Government, 2024). The NSW dataset contains 11,139 observations from NSW respondents and includes an identifier that allows linkage to the administrative dataset. We successfully linked 11,128 NSW dataset respondents with the administrative dataset. We excluded 90 responses due to missing observations in their KtN status, sex, age, ethnicity, highest educational qualification, economic activity, marital status, household size, household tenure or Welsh language proficiency. This resulted in a final analytical sample of 11,049 observations, representing 99.3% of the successfully linked cases.

3 Research questions and hypotheses

We analyse the effectiveness of the KtN strategy implemented as part of the 2022-23 NSW across five dimensions: response rates, sample composition and responses to substantive variables, data quality, and participation in an additional task (online modules), addressing the research questions formulated below.

3.1 RQ1: Is the inclusion of the knock-to-nudge stage associated with improved survey response in the National Survey for Wales?

We assess the increase in response rates, both overall and disaggregated by local authority, OAC, and urban/rural locations, and examine whether it is associated with the implemented KtN strategy. To address this research question, we formulated the following hypotheses:

H1.1: The implementation of the KtN phase has a positive impact on the overall response rate.

H1.2: The increase in response rates varies across local authorities, OACs, and urban/rural locations.

Increasing contact attempts with potential survey participants generally improves response rates, as each additional effort increases the opportunity to reach and persuade an individual to take part (Maslovskaya et al., 2025). However, this relationship is nonlinear, and gains become marginal after a certain point (Groves, 2006; Groves & Couper, 1998). Evidence indicates that participants who require greater recruitment effort are closer to being true

non-respondents than to respond, as they are harder to convert and they are less likely to respond even after repeated attempts. Moreover, after several contacts, remaining cases are increasingly difficult to reach or more inclined to refuse, making further attempts less productive. Based on these findings, we hypothesise that a KtN strategy would yield similar outcomes. Specifically, we formulate the following hypothesis:

H1.3: The largest increases in response rates occurs after the first KtN visit, with diminishing returns from subsequent visits.

3.2 RQ2: Is the knock-to-nudge stage associated with improved sample composition in the National Survey for Wales?

Evidence suggests that, beyond merely improving survey response rates, in-person recruitment strategies, including KtN, contribute to improving sample composition by incorporating harder-to-reach and less represented groups into survey samples (Cavazos et al., 2023; Cleary, 2023; Dodge & Chapman, 2018; Domarchi et al., 2025; Elis et al., 2024). In the UK, Maslovskaya et al. (2025) demonstrated that a KtN strategy was effective in boosting participation among traditionally harder-to-survey groups, including adults aged 70 or over, individuals without formal qualifications, ethnic minorities, single-person households, renters, and people with disabilities. These improvements result in a final sample that more accurately represents the target population. However, similar to RQ1, we expect the KtN visits to exhibit diminishing returns in terms of sample composition, owing to the lower effectiveness of the recruitment process during the second and third visits. Based on this, we propose the following hypotheses:

H2.1: The KtN stage is associated with recruitment of a higher proportion of individuals from lower socio-demographic backgrounds, with lower education attainment, and from ethnic minority groups, compared to the initial recruitment stage.

H2.2: Improvements in sample composition are greater after the first KtN visit than after subsequent KtN visits.

3.3 RQ3: Do participants recruited via knock-to-nudge differ from those recruited during the initial stage in their substantive responses?

The NSW collects measures across a wide range of substantive topics. As we expect that KtN will recruit higher proportions of individuals from lower educational levels and lower socio-demographic backgrounds, we anticipate differences in responses to substantive topics between nudged and non-nudged respondents, especially when the response is expected to correlate highly with socio-demographic variables. We, therefore, formulate the following hypothesis:

H3.1: Participants recruited during the KtN stage provide different responses to substantive questions compared to those recruited during the initial stage.

3.3 RQ4: Do participants recruited via knock-to-nudge provide data of comparable quality to those recruited at the initial stage?

Using data from the PAMCo survey, Maslovskaya et al. (2025) demonstrate that recruiting more reluctant respondents comes at a cost to data quality as these respondents show substantially higher levels of item non-response than those who responded during the initial

recruitment phase. This aligns with the systematic review by Olson (2013), who identified a positive correlation between recruitment effort and item non-response. Respondents who require multiple follow-up attempts, including converted refusals, tend to have higher item non-response rates than those who are easier to recruit. This relationship is generally attributed to respondent motivation. Specifically, individuals who are less interested in a survey are harder to engage and, once they do participate, often show lower commitment to the response task. Additionally, data quality may suffer due to a negative reaction to persistent recruitment efforts. The association between recruitment effort and item non-response is also mediated by socio-demographic characteristics. As Olson (2013) notes, individuals with lower educational attainment or from disadvantaged socio-economic backgrounds are more likely to provide lower-quality survey responses.

Considering this evidence, we formulate the following hypotheses:

H4.1: Participants recruited during the KtN stage have higher levels of item non-response than participants recruited during the initial stage.

3.4 RQ5: Are participants recruited via knock-to-nudge less likely to participate in an additional task (online modules)?

As the KtN strategy is focusing on recruiting these harder-to-survey groups, we anticipate significant differences in participation in the additional task (online modules) between participants recruited via KtN and those recruited during the initial stage.

In recent years, surveys have increasingly invited respondents to participate in additional data collection tasks beyond the main questionnaire. These tasks can involve activities conducted in different settings or formats when compared to the main survey, including those based on mobile technologies (e.g., location or geo-positioning, scanning and QR/barcode readers, visual data capture through photos or video, Bluetooth-enabled devices, and mobile applications) (Link et al., 2014). They may also require respondents to provide information in formats other than a conventional questionnaire, such as retrospective calendars, time-use diaries, or expenditure diaries, or to take part in further assessments including bio-measures, medical examinations, or cognitive tests (Mitchell et al., 2026). In addition, some surveys invite respondents to complete supplementary questionnaire modules at a later time, often using a different mode or device and at the respondent's convenience (West et al., 2023). A key potential advantage of these additional tasks is that they can reduce respondent burden during the main interview and improve data quality by collecting information that respondents may find difficult to recall or report accurately in a single survey session (Revilla et al., 2019). However, participation in such tasks varies considerably. For example, in the Understanding Society Innovation Panel, participation rates in additional activities involving mobile phone applications (such as completing daily wellbeing questions, providing expenditure data, or submitting photographs and health information) ranged from 13% to 45% (Jäckle et al., 2023; Mitchell & Jäckle, 2026). However, other tasks like additional questions via SMS, can reach participation rates of over 70% (Vine et al., 2025).

This variation in participation rates may reflect differences in task characteristics (including required time, cognitive demands, and the type of data requested), as well as respondent characteristics such as motivation, familiarity with the task, and prior survey experience. As the number of data collection requests increases, respondents may experience survey fatigue and saturation (Groves & Couper, 1998). In line with the response burden framework

proposed by Yan & Williams (2022), the effort required of respondents accumulates across tasks, potentially leading to higher levels of satisficing behaviour and item non-response. Within this framework, respondents who are already reluctant to participate can be particularly susceptible to response burden and can potentially be less likely to engage in additional survey activities. We therefore formulate the following hypotheses:

H9: Participants recruited during the KtN stage are less likely to complete the online modules than those recruited during the initial stage.

4 Methods

4.1 Response rates

We calculate response rates in the 2022-23 NSW in line with the AAPOR Response Rate 6 definition (AAPOR, 2023), which includes both complete and partial interviews in the numerator and excludes ineligible addresses from the denominator. In the NSW, ineligible addresses include those that were not yet built or under construction, demolished or derelict, vacant or empty, non-residential, or not occupied as a main residence (e.g., holiday homes), as well as communal establishments/institutions, addresses the interviewers were instructed not to sample, and those coded by interviewers as inaccessible or not found. Ineligible addresses do not include those that were attempted multiple times by interviewers but where no contact was made, unless there is clear evidence that the address is vacant or not a main residence (e.g., confirmed by neighbours). Addresses with unknown eligibility have been recorded as non-contacts when issued but not attempted, or when it was unclear whether the address was residential or contained eligible residents due to non-contact. Where information was refused about whether the address was residential or contained eligible residents, these cases were recorded as refusals. There are no cases with unknown eligibility in the issued sample. The overall level of ineligible addresses was 10.6% (Jenkins and Forbes, 2024).

To address our RQ1, we calculate response rates across four nested subsets of the administrative dataset (issued sample, including both responding and non-responding households):

- **Subset 1:** Households that were not part of KtN.
- **Subset 2:** Households belonging to Subset 1 *plus* those that received one KtN visit.
- **Subset 3:** Households belonging to Subset 2 *plus* those that received two KtN visits.
- **Full issued sample:** Households belonging to Subset 3 *plus* those that received three KtN visits *plus* those that were part of KtN but for which the number of visits was not recorded.

In addition to the overall survey response rate, we calculate disaggregated response rates for the three administrative variables available for both respondents and non-respondents in the administrative dataset: local authority, OAC, and urban/rural location. Furthermore, to evaluate the effectiveness of the KtN protocol, we calculate the incremental increase in response rates. This was defined as the difference in response rates between each successive subset of participants. For example, the increase attributable to the second KtN visit is derived from the difference between the response rate in Subset 3 and that in Subset 2. This method

allows us to determine which KtN stage was most effective in increasing participation, both overall and for each sub-category of the administrative variables.

4.2 Sample composition

We address representativeness in the 2022-23 NSW from two perspectives: (i) comparing the achieved sample with the issued sample, and (ii) comparing the achieved sample with high-quality population benchmarks.

As the main focus of this paper is on the quality of data collection process and representativeness rather than on quality of inference and survey estimates, we will not be applying weights in this analysis.

4.2.1 Comparison with issued sample

For this analysis, we focus on the differences between the issued and achieved samples, using R-indicators, originally proposed and developed by Schouten et al. (2009). R-indicators are designed to measure the degree to which respondents to a survey resemble the complete sample or population (Maslovskaya et al., 2022; Moore et al., 2018; Rohr et al., 2024; Schouten et al., 2009). With R-indicators, the contrast between respondents and the issued sample is defined with respect to a set of auxiliary variables that must be available for the full issued sample (in our case, urban/rural location, OAC, and local authority).

We start by fitting a logistic regression model using the auxiliary variables, with a binary response indicator y_i equal to 1 when individual i responded to the survey. The response probability for individual i is denoted as $p_i = Pr(y_i = 1)$. The logistic regression model can then be written as:

$$\log \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i} + \dots + \beta_k \cdot x_{ki} = \mathbf{b}^T \mathbf{x}_i, \quad (1)$$

where $\mathbf{b}^T = \{\beta_0, \beta_1, \dots, \beta_k\}$ is a vector of K auxiliary variables and \mathbf{x}_i is a vector of covariates at the individual level.

R-indicators are calculated using the response propensities obtained in the model specified in equation (1). The overall R-indicator R is a transformed response propensity standard deviation, calculated as:

$$R = 1 - 2 \cdot \sqrt{\frac{\sum_{i=1}^n (p_i - \bar{p})^2}{n - 1}}, \quad (2)$$

where n is the sample size, p_i is the estimated response propensity for individual i and \bar{p} is the mean response propensity (e.g., see Moore et al. (2018)). An R-indicator closer to 1 indicates that the achieved sample is representative, while an R-indicator closer to 0 indicates the opposite.

Conditional *partial R-indicators* on the variable level can be calculated for variables used for the response propensity model (Maslovskaya & Lugtig, 2022). They isolate the part of the deviation of representativeness that can be attributed to a variable alone. The partial R-

indicator for variable x_k is obtained by cross-classification of all variables in the model with the exception of the variable itself. This cross-classification results in L cells U_1, U_2, \dots, U_L . Let n_l denote the sample size in cell l , for $l = 1, 2, \dots, L$, and $\sum_{l=1}^L n_l = n$. Let \bar{p}_l denote the mean of the response probabilities in cell l . The conditional partial indicator for categorical variable x_k can be defined as:

$$P_c(x_k) = \sqrt{\frac{1}{n} \sum_{l=1}^L \sum_{i \in U_l} (p_i - \bar{p}_l)^2}, \quad (2)$$

The conditional partial indicator is the remaining within-cell variation of response probabilities if the variable x_k is removed from the cross-classification. The larger the value for $P_c(x_k)$, the larger the role of this variable in the (lack of) representativeness. To address the association of the KtN strategy with representativeness, we calculate R-indicators for the same four subsets of the administrative dataset described in Section 4.1.

To address the contribution of each auxiliary variable available for the issued sample, we calculate four sets of R-indicators for each subset: (i) an *overall* R indicator, based on the logistic regression model in (1) including all three auxiliary variables (OAC, local authority, and urban/rural location); and (ii) three *partial* R-indicators, each based on the logistic regression model in (1) including only one auxiliary variable. We use the `sampcompR` package in R (Rohr & Felderer, 2025) to calculate the R-indicators.

4.2.2 Comparison with population benchmarks

We also compare distributions of socio-economic variables in the NSW sample (achieved sample, considering respondents only) against high-quality population benchmarks for the target population (adults aged 16 years or over residing in households in Wales). Benchmarks are sourced from the 2021 Census for England and Wales (Office for National Statistics, 2021) and include only residents in Wales.

We evaluate the following variables: sex, age, cross-classified sex and age, ethnicity, highest educational qualification, economic activity, marital status, household size, household tenure, Welsh language proficiency. In addition, we assess urban/rural location, OAC, and local authority. We base our comparison on the following nested sub-samples:

- **Sub-sample A:** Responding households that were not part of KtN.
- **Sub-sample B:** Responding households belonging to Sub-sample A *plus* responding households that received one KtN visit.
- **Sub-sample C:** Responding households belonging to Sub-sample B *plus* responding households that received two KtN visits.
- **Full achieved sample:** Responding households belonging to Sub-sample C *plus* responding households that received three KtN visits *plus* those that were part of KtN but for which the number of visits was not recorded.

We compare the unweighted distributions of each variable in Sub-samples A, B, C, and the full sample against the population benchmarks. We calculate Z-tests to assess whether each sample proportion differs significantly from the population benchmark.

To evaluate representativeness more systematically, we calculate a *dissimilarity index* for each variable, based on the method developed by Duncan & Duncan (1955). This index quantifies

the proportion of respondents that would need to be reallocated across categories of a given variable to match the population distribution. The dissimilarity index DS_k for variable k , which has been divided into M categories, is defined as:

$$DS_k = \frac{1}{2} \cdot \sum_{m=1}^M |p_{km} - P_{km}|, \quad (3)$$

where p_{km} is the proportion of respondents in category m of variable k in the sample, and P_{km} is the corresponding proportion in the target population. For example, lower dissimilarity index in Sub-sample B (compared to Sub-sample A) suggests that adding KtN respondents improves representativeness. This approach enables us to assess the relative contribution of each survey stage to improving alignment between the sample and the target population.

4.3 Substantive data

We analyse substantive responses for the 41 “ask-all” questions used later in the item non-response analysis. Twenty-one of these questions are binary, allowing participants to answer “Yes” or “No”, with the additional options to refuse or say “don’t know”. For these questions, we compare the proportions of “Yes” responses between nudged and non-nudged addresses.

The remaining 20 questions are single-choice categorical variables, where participants must select one option from a list of three or more. For consistency and simplicity, we transform responses to these questions into binary variables for analysis and calculate the proportion of participants who selected an option of interest for each question. For example, for the “*Good health*” variable, we create a binary variable equal to 1 if participants indicate their health is “Good” or “Very good”, and 0 if it is “Fair”, “Bad”, or “Very bad”. Similarly, for the “*Concern about climate change*” variable, the binary variable equals 1 if participants say they are “Very concerned” or “Fairly concerned”, and 0 if they select any of the other three options. “Don’t know” answers, refusals, and skipped responses are kept intact in these transformations and not included in this analysis. The full list of transformed variables and groups considered in the analysis is provided in Table A1 in the Appendix. For these 20 single-choice questions, we compare the proportions of responses in category coded as 1 for each of the variables analysed between nudged and non-nudged addresses. We employ a χ^2 test to assess the statistical significance of response differences between nudged and non-nudged addresses. The focus of this analysis is to assess variability between the two sub-samples (nudged and non-nudged) as the variability between these two groups would provide additional evidence for improved representativeness by including KtN stage of recruitment. We also calculate a χ^2 test to compare responses between non-nudged addresses and the full sample to assess whether nudged cases added sufficient value.

4.4 Item non-response

We address data quality by comparing item non-response indicators between nudged and non-nudged participants. Our analysis focuses on the 41 questions that were administered to all respondents and were not derived variables. These include 32 questions from the telephone component of the survey, along with 9 questions from the additional online modules. The full list is provided in Table A1 of the Appendix.

For each respondent j , we compute three item non-response indicators I_{jq} . All the indicators take the following form:

$$I_{jq} = \frac{N_{jq}}{T_j}, \quad (4)$$

where $q = \{1, 2, 3\}$ and T_j denotes the number of questions participant j was eligible to answer. For participants who completed both survey components (telephone and online), $T_j = 41$. For those who only completed the telephone questionnaires $T_j = 32$. N_{jq} denotes the number of questions for which respondent j provided a (non-) response of type q . We consider three types of item non-response: (i) spontaneous “don’t know” responses ($q = 1$), (ii) spontaneous refusals ($q = 2$), and (iii) skipped questions, either due to the interview being terminated early or because the respondent indicated that the question did not apply to them ($q = 3$). We also compute a total item non-response indicator I_j , which considers all three types of item non-response, and is defined as follows:

$$I_j = \frac{N_j}{T_j} = \frac{\sum_{jq} N_{jq}}{T_j} \quad (5)$$

We first plot the distribution of the three types of item non-response and the total item non-response indicator and compare them between nudged and non-nudged addresses. We then compute descriptive statistics (mean and 95% confidence intervals), comparing the item non-response indicators across survey stages and KtN visits.

4.5 Participation in the additional online modules

We measured participation in the online modules (additional task) by calculating the proportion of main survey respondents who completed the required additional online modules. Mean participation rates and 95% confidence intervals were computed separately for nudged and non-nudged addresses.

Additionally, we fit a binary logistic regression (Agresti, 2013) for the participation variable, controlling for the socio-demographic characteristics. A significant coefficient for the Knock-to-nudge status variable in this model would indicate that nudged addresses had a significantly different probability of responding to the online modules compared to non-nudged households.

5 Results

5.1 Response rates

The administrative dataset (issued sample) contains 32,447 addresses, of which 3,276 were ineligible (vacant, empty, demolished, or non-residential). This resulted in 29,171 eligible addresses, of which 11,128 provided a response, yielding an overall response rate of 38.1%. Table 1 presents the number of responding addresses per number of KtN visits, along with the cumulative response rate for each stage of the recruitment process.

Table 1. Survey response by KtN visits

| Number of KtN visits | Eligible addresses | Responding addresses | % of responding addresses | Cumulative response rate (%) |
|---------------------------------|--------------------|----------------------|---------------------------|------------------------------|
| No KtN visits | 6,143 | 4,015 | 36.1 | 13.8 |
| One KtN visit | 11,340 | 4,658 | 41.9 | 29.8 |
| Two KtN visits | 5,817 | 1,458 | 13.1 | 34.8 |
| Three KtN visits | 4,644 | 667 | 6.0 | 37.1 |
| No info on number of KtN visits | 1,227 | 320 | 2.9 | 38.1 |
| Total | 29,171 | 11,128 | 100.0 | - |

Note: AAPOR RR6 was calculated.

The table shows that 63.9% of responses in the 2022-23 NSW were obtained during the KtN stage, which represents a substantial improvement in the overall response rate. The initial stage achieved a 13.8% response rate. The first KtN visit yielded 4,655 additional responses, raising the overall response rate to 29.8%. The first KtN visit is the most effective stage for increasing participation, adding 15.9 percentage points. In contrast, the second visit contributed a gain of only 5.0 percentage points, while the third visit (and households for which number of visits is not available) increased the response rate by just 3.4 percentage points compared to two visits. These results support our H1.1 and H1.3.

Table 2 presents disaggregated response rates by local authority, OAC, and household location, for the nested subsets described in Section 4.1, which, in practice, allow us to calculate cumulative response rates for each KtN stage. Consistent with our H1.2, response varies substantially across local authorities, with the lowest overall response rate (29.4%) observed in Blaenau Gwent (located in the southeast of Wales), and the highest (46.4%) in Carmarthenshire (in the southwest). Similarly, in terms of OAC, response varies between 28.0% (“ethnicity central”, i.e. those likely to belong to non-white ethnic groups) to 43.3% (“suburbanites”, or those residing in the outskirts of urban areas). Observed response is higher for rural addresses (39.1%) compared to urban addresses (37.4%).

Table 2. Survey response per socio-demographic variable and administrative sample subset

| Variable | Category | Issued sample size | Response rate (%) | | | |
|-----------------|-------------------|--------------------|-------------------|----------|----------|-------------|
| | | | Subset 1 | Subset 2 | Subset 3 | Full issued |
| Local authority | Isle of Anglesey | 843 | 14.8 | 33.9 | 39.6 | 43.7 |
| | Gwynedd | 1,213 | 16.0 | 30.3 | 35.3 | 38.3 |
| | Conwy | 990 | 14.4 | 32.0 | 37.1 | 38.3 |
| | Denbighshire | 741 | 12.8 | 34.0 | 39.4 | 43.0 |
| | Flintshire | 1,190 | 12.4 | 29.8 | 35.4 | 39.2 |
| | Wrexham | 1,001 | 13.0 | 30.4 | 36.3 | 39.7 |
| | Powys | 3,333 | 20.3 | 30.3 | 32.5 | 33.0 |
| | Ceredigion | 972 | 16.0 | 28.4 | 31.5 | 33.3 |
| | Pembrokeshire | 887 | 19.3 | 38.2 | 43.3 | 46.1 |
| | Carmarthenshire | 1,234 | 16.5 | 36.3 | 42.2 | 46.4 |
| | Swansea | 1,946 | 12.4 | 28.9 | 35.2 | 38.4 |
| | Neath Port Talbot | 1,383 | 10.0 | 29.8 | 36.1 | 39.6 |
| | Bridgend | 1,120 | 12.3 | 31.8 | 37.0 | 38.8 |
| | Vale of Glamorgan | 1,121 | 13.0 | 27.8 | 33.7 | 37.6 |
| | Cardiff | 2,809 | 13.9 | 28.0 | 31.7 | 37.3 |
| | Rhondda Cynon Taf | 1,988 | 9.1 | 24.9 | 31.3 | 37.0 |
| | Merthyr Tydfil | 902 | 8.8 | 26.8 | 29.9 | 35.4 |
| | Caerphilly | 1,430 | 9.0 | 26.7 | 33.5 | 38.4 |
| | Blaenau Gwent | 1,101 | 8.5 | 21.5 | 26.3 | 29.4 |
| Torfaen | 787 | 13.9 | 35.2 | 41.0 | 42.2 | |

| Variable | Category | Issued sample size | Response rate (%) | | | |
|------------------------------|-----------------------------|--------------------|-------------------|-------------|-------------|-------------|
| | | | Subset 1 | Subset 2 | Subset 3 | Full issued |
| | Monmouthshire | 954 | 18.6 | 31.7 | 36.7 | 40.4 |
| | Newport | 1,226 | 12.2 | 29.7 | 36.1 | 39.5 |
| OAC | Rural residents | 6,702 | 19.7 | 35.3 | 39.5 | 41.6 |
| | Cosmopolitans | 671 | 8.6 | 22.2 | 27.9 | 31.9 |
| | Ethnicity central | 164 | 9.8 | 20.1 | 22.6 | 28.7 |
| | Multicultural cosmopolitans | 758 | 10.0 | 23.9 | 28.8 | 33.9 |
| | Urbanites | 3,402 | 14.5 | 30.5 | 35.1 | 38.9 |
| | Suburbanites | 5,603 | 16.0 | 33.8 | 39.2 | 43.4 |
| | Constrained city dwellers | 2,244 | 9.6 | 23.4 | 28.9 | 31.7 |
| | Hard pressed living | 9,627 | 9.7 | 25.9 | 31.3 | 34.9 |
| Location | Urban | 18,290 | 11.8 | 28.0 | 33.5 | 37.5 |
| | Rural | 10,881 | 17.1 | 32.7 | 36.9 | 39.3 |
| Overall response rate | | 29,173 | 13.8 | 29.8 | 34.8 | 38.1 |

Note: PP is percentage point. AAPOR RR6 is calculated.

5.2 Sample composition

5.2.1 Comparison with issued sample

Figure 1 presents the partial R-indicators calculated for each sample subset, as well as the full issued sample, for each of the three variables available for both respondents and non-respondents (local authority, OAC, and urban/rural), as well as for the overall R-indicator including all three variables.

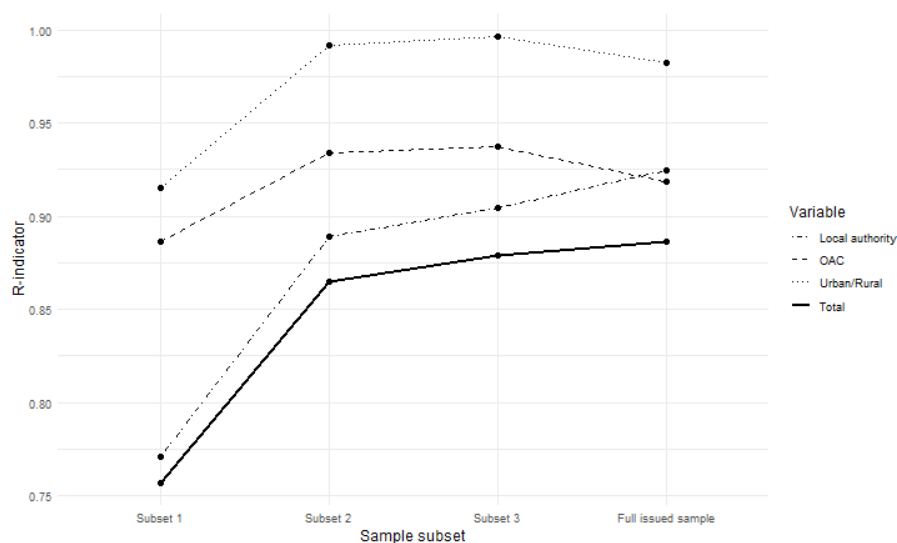


Figure 1. R-indicators per variable and survey stage

Figure 1 demonstrates that the R-indicators for the full sample are higher than those obtained for the initial sample (Subset 1), confirming our H2.1. The improvement is more pronounced from Subset 1 to Subset 2, both across all individual variables, and overall. Specifically, adding the households that were nudged only once to the initial sample results in an overall improvement of the R-indicator calculated across all three variables from 0.760 (Subset 1) to 0.860 (Subset 2). Further gains are observed in Subset 3 (0.880) and the full sample (0.890), although these improvements diminish with each KtN visit, consistent with our H2.2. Moreover, when R-indicators are calculated separately for the urban/rural location or OAC, the full issued sample yields a lower R-indicator than Subset 3, suggesting that a third KtN visit

does not improve sample representativeness with respect to these variables and may even have the opposite effect.

5.2.2 Comparison with population benchmarks

Table 3 compares the unweighted distributions of the socio-demographic variables with the population benchmarks for the three nested sub-samples described in Section 4.2.2.

Table 3. Sample composition and representativeness – Main telephone survey

| Variable | Category | Composition by sub-sample (%) | | | | Benchmark | Dissimilarity Index by sub-sample (%) | | | |
|-----------------------|---------------------|-------------------------------|--------------|---------------|------------------|-----------|---------------------------------------|--------------|---------------|------------------|
| | | A (3,992) | B (8,629) | C (10,078) | Full (11,049) | | A (3,992) | B (8,629) | C (10,078) | Full (11,049) |
| Sex | Female | 56.7*** | 57.1*** | 57.7*** | 57.8*** | 51.5 | 5.2 | 5.6 | 6.1 | 6.2 |
| | Male | 43.3*** | 42.9*** | 42.3*** | 42.2*** | | | | | |
| Age | 16 to 24 | 1.9*** | 2.8*** | 3.1*** | 3.4*** | 12.9 | 25.8 | 20.2 | 18.0 | 16.2 |
| | 25 to 34 | 7.8*** | 9.7*** | 10.5*** | 11.1*** | | | | | |
| | 35 to 49 | 13.7*** | 16.6*** | 17.7*** | 18.5*** | | | | | |
| | 50 to 64 | 28.1*** | 27.8*** | 27.6*** | 27.4*** | | | | | |
| | 65 or over | 48.4*** | 43.2*** | 41.2*** | 39.6*** | | | | | |
| Sex * Age | Female * 16 to 24 | 1.2*** | 1.5*** | 1.7*** | 1.8*** | 6.3 | 26.0 | 20.7 | 18.6 | 17.4 |
| | Female * 25 to 34 | 4.9*** | 6.1*** | 6.7*** | 7.2* | | | | | |
| | Female * 35 to 49 | 8.4*** | 10.3* | 10.9 | 11.3 | | | | | |
| | Female * 50 to 64 | 16.2*** | 16.1*** | 16.1*** | 15.9*** | | | | | |
| | Female * 65 or over | 26.1*** | 23.1*** | 22.3*** | 21.5*** | | | | | |
| | Male * 16 to 24 | 0.8*** | 1.3*** | 1.4*** | 1.5*** | | | | | |
| | Male * 25 to 34 | 2.9*** | 3.5*** | 3.8*** | 3.9*** | | | | | |
| | Male * 35 to 49 | 5.3*** | 6.3*** | 6.7*** | 7.2*** | | | | | |
| | Male * 50 to 64 | 11.9 | 11.7 | 11.5* | 11.5* | | | | | |
| Male * 65 or over | 22.4*** | 20.1*** | 18.9*** | 18.1*** | | | | | | |
| Local authority | Blaenau Gwent | 2.3 | 2.7*** | 2.8*** | 2.9*** | 2.2 | 17.7 | 11.8 | 10.9 | 10.0 |
| | Bridgend | 3.4*** | 4.1** | 4.1** | 3.9*** | | | | | |
| | Caerphilly | 3.2*** | 4.4*** | 4.7*** | 4.9*** | | | | | |
| | Cardiff | 9.7*** | 9.1*** | 8.8*** | 9.4*** | | | | | |
| | Carmarthenshire | 5.1** | 5.2*** | 5.1*** | 5.1*** | | | | | |
| | Ceredigion | 3.9*** | 3.2*** | 3*** | 2.9*** | | | | | |
| | Conwy | 3.6 | 3.7 | 3.6 | 3.4* | | | | | |
| | Denbighshire | 2.4** | 2.9 | 2.9 | 2.9 | | | | | |
| | Flintshire | 3.7*** | 4.1*** | 4.2*** | 4.2*** | | | | | |
| | Gwynedd | 4.8** | 4.2* | 4.2* | 4.2* | | | | | |
| | Isle of Anglesey | 3.1*** | 3.3*** | 3.3*** | 3.3*** | | | | | |
| | Merthyr Tydfil | 2.0 | 2.8*** | 2.7*** | 2.9*** | | | | | |
| | Monmouthshire | 4.4*** | 3.5* | 3.4* | 3.5* | | | | | |
| | Neath Port Talbot | 3.4*** | 4.8 | 4.9* | 4.9* | | | | | |
| | Newport | 3.7*** | 4.2*** | 4.3*** | 4.3*** | | | | | |
| | Pembrokeshire | 4.3 | 3.9 | 3.8 | 3.7* | | | | | |
| | Powys | 16.9*** | 11.7*** | 10.7*** | 9.9*** | | | | | |
| | Rhondda Cynon Taf | 4.5*** | 5.6*** | 6.1*** | 6.6*** | | | | | |
| | Swansea | 6*** | 6.5*** | 6.8*** | 6.7*** | | | | | |
| Torfaen | 2.7 | 3.2 | 3.2 | 3 | | | | | | |
| Vale of Glamorgan | 3.6* | 3.6** | 3.7** | 3.8* | | | | | | |
| Wrexham | 3.2*** | 3.5*** | 3.6*** | 3.6*** | | | | | | |
| Ethnicity | Other | 2.6*** | 3*** | 3.2*** | 3.4*** | 5.4 | 2.8 | 2.3 | 2.2 | 2.0 |
| | White | 97.4*** | 97*** | 96.8*** | 96.6*** | | | | | |
| Highest qualification | Level 1 | 3*** | 3.9*** | 4*** | 4.1*** | 8.7 | 22.6 | 18.3 | 18.1 | 18.4 |
| | Level 2 | 13.2* | 14.4 | 14.3 | 14.2 | | | | | |
| | Level 3 | 9.2*** | 9.1*** | 9.3*** | 9.3*** | | | | | |
| | Level 4 | 50.7*** | 45.3*** | 44.8*** | 45*** | | | | | |
| | No qualifications | 12.1*** | 14.5*** | 14.5*** | 14.3*** | | | | | |
| | Others | 11.7*** | 12.7*** | 13.1*** | 13.2*** | | | | | |

| Variable | Category | Composition by sub-sample (%) | | | | Benchmark | Dissimilarity Index by sub-sample (%) | | | |
|--|--|-------------------------------|--------------|---------------|------------------|-----------|---------------------------------------|--------------|---------------|------------------|
| | | A (3,992) | B (8,629) | C (10,078) | Full (11,049) | | A (3,992) | B (8,629) | C (10,078) | Full (11,049) |
| Economic activity | Employed | 42.5*** | 45.1*** | 46.9*** | 48.5*** | 53.5 | 23.1 | 17.8 | 15.5 | 13.9 |
| | Unemployed | 1*** | 1.2*** | 1.4*** | 1.4*** | 3.1 | | | | |
| | Inactive: Student | 1.2*** | 1.4*** | 1.6*** | 1.8*** | 5.7 | | | | |
| | Inactive: Retired | 47.8*** | 42.5*** | 40.2*** | 38.6*** | 24.7 | | | | |
| | Inactive: Long-term sick or disabled | 4.8*** | 5.9 | 5.9 | 5.8 | 5.9 | | | | |
| | Inactive: Looking after home/family | 2.1*** | 3.4*** | 3.4*** | 3.4*** | 4.3 | | | | |
| | Inactive: Other | 0.6*** | 0.5*** | 0.5*** | 0.5*** | 2.9 | | | | |
| Marital status | Never married/in a civil partnership | 21.9*** | 24.5*** | 25.8*** | 26.8*** | 37.2 | 15.3 | 12.7 | 11.4 | 10.4 |
| | Married/in a civil partnership | 47.8*** | 47.1*** | 46.2*** | 45.6*** | 43.8 | | | | |
| | Separated still married/in a civil partnership | 2.1 | 2.3 | 2.3 | 2.4* | 2.1 | | | | |
| | Divorced/civil partnership dissolved | 13.9*** | 13.3*** | 13.3*** | 13.2*** | 9.9 | | | | |
| | Widowed/surviving civil partner | 14.3*** | 12.9*** | 12.5*** | 12.1*** | 7.1 | | | | |
| Welsh language | Does not speak Welsh | 84.9 | 84.9 | 85.1 | 84.9 | 84.7 | 0.2 | 0.2 | 0.4 | 0.2 |
| | Speaks Welsh | 15.1 | 15.1 | 14.9 | 15.1 | 15.3 | | | | |
| Household size | 1 person | 36.7*** | 33.9*** | 33.5*** | 33.2** | 31.9 | 13.8 | 8.0 | 6.6 | 5.9 |
| | 2 people | 44.1*** | 41.1*** | 40.1*** | 39.7*** | 35.1 | | | | |
| | 3 people | 9.8*** | 12.6*** | 12.9*** | 13.2*** | 15.8 | | | | |
| | 4 or more people | 9.3*** | 12.4*** | 13.5*** | 14.0*** | 17.2 | | | | |
| Tenure | Owned | 81.2*** | 76.4*** | 74.8*** | 74.2*** | 66.4 | 14.8 | 10.6 | 9.4 | 8.8 |
| | Private rented: Other | 1.2*** | 1.1*** | 1.1*** | 1.2*** | 2.7 | | | | |
| | Private rented: Private landlord | 8.6*** | 9.9*** | 10.7*** | 11.2*** | 14.5 | | | | |
| | Social rented: Other | 6.1*** | 8.5* | 8.9*** | 8.8*** | 7.8 | | | | |
| | Social rented: Council or local authority | 2.9*** | 4.2*** | 4.6*** | 4.6*** | 8.7 | | | | |
| Mean dissimilarity index for the sample | | | | | | | 15.2 | 11.7 | 10.6 | 9.9 |

Note: The table compares, for each variable and sub-sample, the unweighted proportion of respondents in each category with the population benchmark for each category. Z-tests are used for comparing proportions. Significance levels: *p < 0.05, **p < 0.01, ***p < 0.001. The sizes of each sub-sample are in parentheses in the header.

In line with our H2.1, the inclusion of KtN respondents reduces the dissimilarity indices for most variables, bringing sample composition closer to population benchmarks. Specifically, KtN visits improve representativeness across several key characteristics: age (by raising the share of younger participants), ethnicity (by increasing the proportion from non-White British backgrounds), education (by including more participants with no formal qualifications), economic activity (by adding more employed participants), household size (by including more households with three or more members), and tenure (by adding more respondents who rent, either from a private landlord or a local council). Two variables do not improve after including KtN respondents: sex and Welsh language proficiency (language proficiency is well-represented at each stage of recruitment).

Consistent with our H2.2, the most substantial improvements in representativeness occur after the first KtN visit for all variables where dissimilarity is reduced. Specifically, the mean dissimilarity index falls from 15.2% in Sub-sample A to 11.7% in Sub-sample B, which includes households visited once. Subsequent visits yield smaller gains, with the index reaching 9.9% for the full sample. Notably, for the sex and highest educational qualification variables, the dissimilarity index increases slightly after incorporating households visited three times. These findings confirm that the effectiveness of the KtN protocol diminishes after the initial visit, which supports H2.2.

It is important to note that the overall dissimilarity index decreased substantially between the initial stage and the full sample, suggesting that the overall KtN efforts were important in improving sample composition.

5.3 Substantive data

Table 4 summarises the descriptive analysis conducted for the substantive variables. We compare responses obtained from nudged and non-nudged addresses and also from non-nudged addresses and full achieved sample to assess whether KtN stage added sufficient value. The table also provides the sample size (excluding “don’t know” responses, refusals, and skipped responses) and the proportion estimated for the full sample.

Table 4. Percentage of respondents with positive response to substantive questions

| Module | N | Question | Group | Sample size | Percentage of respondents (%) | | | Comparisons (χ^2) | |
|---------------------------|----|--|---|-------------|-------------------------------|------------|--------|--------------------------|--------------------|
| | | | | | Full sample | Non-nudged | Nudged | Nudged vs. Non-nudged | Full vs Non-nudged |
| Sociodemographic | 1 | Religion | No religion | 11,044 | 45.9 | 42.1 | 48.0 | 35.4*** | 16.6*** |
| | 2 | Time living at this address | 5 years or more | 11,050 | 72.3 | 74.9 | 70.8 | 20.7*** | 9.8** |
| | 3 | Satisfaction with accommodation | Fairly satisfied/ Very satisfied | 11,051 | 93.9 | 95.3 | 93.1 | 20.6*** | 10.2** |
| Internet | 4 | Has access to the internet | Yes | 11,052 | 91.6 | 91.9 | 91.5 | 0.5 | 0.2 |
| | 5 | Uses internet | Yes | 11,053 | 87.8 | 88.2 | 87.6 | 0.7 | 0.3 |
| Welsh | 6 | Understands spoken Welsh | Yes | 11,045 | 28.8 | 30.1 | 28.1 | 4.8* | 2.2 |
| | 7 | Speaks Welsh | Yes | 11,054 | 17.1 | 17.2 | 17.1 | 0.0 | 0.0 |
| | 8 | Reads Welsh | Yes | 11,044 | 20.5 | 20.8 | 20.4 | 0.2 | 0.1 |
| | 9 | Writes Welsh | Yes | 11,043 | 17.5 | 17.8 | 17.3 | 0.4 | 0.2 |
| Occupation/ Education | 10 | Currently studying for a degree | Yes | 11,050 | 3.7 | 2.9 | 4.1 | 10.3** | 5.0* |
| Transport | 11 | Car or van available for use | Yes | 11,052 | 85.6 | 87.4 | 84.5 | 17.1*** | 8.2** |
| | 12 | Used bus in the last 12 months | Yes | 11,050 | 39.3 | 39.6 | 39.1 | 0.3 | 0.1 |
| | 13 | Used train in the last 12 months | Yes | 11,048 | 38.8 | 37.5 | 39.5 | 4.3* | 2.0 |
| | 14 | Satisfaction with bus | Fairly satisfied/ Very satisfied | 11,053 | 41.7 | 41.3 | 41.9 | 0.3 | 0.1 |
| | 15 | Satisfaction with train | Fairly satisfied/ Very satisfied | 11,053 | 40.7 | 40.7 | 40.7 | 0.0 | 0.0 |
| Arts/Culture/ Heritage | 16 | Visited an archive or records office | Yes | 11,047 | 7.9 | 9.2 | 7.1 | 14.9*** | 6.7* |
| Climate change | 17 | Is climate changing | Definitely/ Probably | 10,775 | 96.7 | 97.1 | 96.4 | 3.5 | 1.7 |
| | 18 | Cause of climate change | Entirely or Mainly by human activity | 10,671 | 53.1 | 55.0 | 52.0 | 8.3** | 3.9* |
| | 19 | Concern about climate change | Fairly concerned/ Very concerned | 10,877 | 75.8 | 78.5 | 74.3 | 24.8*** | 11.8** |
| | 20 | Impact of climate change in Wales | Already having an impact | 10,395 | 76.7 | 80.1 | 74.7 | 39.4*** | 18.7*** |
| | 21 | Responsibility of the general public | A lot | 10,736 | 59.8 | 61.8 | 58.7 | 10.2** | 4.7* |
| | 22 | Responsibility of business | A lot | 10,674 | 69.0 | 70.3 | 68.2 | 5.3* | 2.5 |
| | 23 | Responsibility of government | A lot | 10,705 | 76.6 | 78.2 | 75.7 | 8.6** | 4.0* |
| | 24 | Government responding | A lot | 8,233 | 77.2 | 77.2 | 77.1 | 0.0 | 0.0 |
| Social care | 25 | Carer for someone else | Yes | 11,030 | 28.5 | 28.1 | 28.7 | 0.4 | 0.2 |
| | 26 | Received help for themselves | Yes | 11,030 | 5.2 | 5.3 | 5.2 | 0.1 | 0.0 |
| | 27 | Received help to care for someone else | Yes | 11,025 | 6.5 | 6.6 | 6.5 | 0.0 | 0.0 |
| Health/GP/ Hospital | 28 | Health in general | Good/Very good | 11,021 | 64.5 | 65.4 | 64.1 | 1.8 | 0.8 |
| | 29 | Seen a GP | Yes | 11,016 | 65.4 | 66.5 | 64.7 | 3.8 | 1.8 |
| | 30 | Seen other health professional at GP surgery | Yes | 11,018 | 48.5 | 52.8 | 46.0 | 46.6*** | 21.8*** |
| | 31 | Had a hospital appointment | Yes | 11,018 | 41.5 | 43.6 | 40.3 | 11.4** | 5.3* |

| Module | N | Question | Group | Sample size | Percentage of respondents (%) | | | Comparisons (χ^2) | |
|---------------------------------|----|--|---|-------------|-------------------------------|------------|--------|--------------------------|--------------------|
| | | | | | Full sample | Non-nudged | Nudged | Nudged vs. Non-nudged | Full vs Non-nudged |
| | 32 | Attended A&E | Yes | 11,025 | 14.1 | 13.5 | 14.5 | 1.8 | 0.8 |
| Recycling | 33 | Satisfaction with recycling services | Fairly satisfied/ Very satisfied | 9,296 | 84.8 | 86.0 | 84.0 | 6.6* | 3.0 |
| | 34 | Informed about recycling services | Tend to agree/ Strongly agree | 9,168 | 68.0 | 69.7 | 67.0 | 6.9** | 3.1 |
| | 35 | Has altered clothing | Yes | 9,275 | 45.8 | 50.7 | 42.8 | 54.8*** | 24.3*** |
| | 36 | Has repaired any household item | Yes | 9,281 | 38.8 | 41.2 | 37.2 | 14.7*** | 6.5* |
| Physical punishment of children | 37 | Sometimes necessary | Agree | 8,817 | 29.8 | 32.5 | 28.1 | 19.3*** | 8.4** |
| | 38 | To stop children harming themselves | Yes, as last resort/ Yes, whenever necessary | 8,624 | 52.3 | 55.6 | 50.2 | 23.1*** | 10.2** |
| | 39 | To stop children harming another child | Yes, as last resort/ Yes, whenever necessary | 8,607 | 49.6 | 52.6 | 47.7 | 19.9*** | 8.8** |
| | 40 | To stop out of control behaviour | Yes, as last resort/ Yes, whenever necessary | 8,589 | 30.5 | 32.7 | 29.1 | 12.1*** | 5.3* |
| | 41 | As a punishment for misbehaving | Yes, as last resort/ Yes, whenever necessary | 8,590 | 31.4 | 33.8 | 29.9 | 14.1*** | 6.2* |

Note: The last two columns in the table present the results of two χ^2 tests. The second-to-last column compares the distribution of each variable among nudged addresses with that among non-nudged addresses. The final column compares the distribution of each variable among non-nudged addresses with that in the full sample. In both cases, significance levels are indicated as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

We found significant differences in responses between nudged and non-nudged addresses for 25 of the 41 variables analysed (61%). Similarly, responses in 20 of these variables also differ when comparing non-nudged addresses with the full sample. Both results are consistent with H3.1 and suggest that the inclusion of KtN stage improved representativeness and added sufficient value. Compared with respondents in non-nudged addresses, those in nudged addresses are less likely to have lived at their current address for five or more years, to be satisfied with their accommodation, to have access to a car or van, and to have visited an archive or records office in the past twelve months. Nudged respondents are more likely than non-nudged respondents to state they have no religion and to have used a train in the last twelve months. They are also less likely to have seen a GP or other health professional at a GP surgery, or to have had a hospital appointment.

Notable differences emerge in attitudes towards climate change. While there is no difference between nudged and non-nudged addresses in the belief that the climate is changing, nudged respondents are less likely than non-nudged ones to believe that climate change is caused entirely or mainly by human activity, or that it is currently having an impact in Wales. They are also less concerned about climate change and less likely to believe that the general public, businesses, or the government have “a lot” of responsibility in tackling it. Additionally, nudged respondents are less likely to feel informed about their local recycling services or to have carried out recycling-related actions (such as altering clothing or repairing household items), compared with non-nudged respondents. Nudged respondents are also generally less likely to believe in physical punishment for children.

Interestingly, we found no significant differences between nudged and non-nudged respondents in questions related to internet use or social care. In addition, nudged participants were just as likely as non-nudged participants to rate their own health as “good” or “very good”.

As noted in section 5.4 below, for certain questions, respondents who say “don’t know” may be expressing genuine lack of awareness about the topic, rather than engaging in satisficing behaviour. Table 5 summarises the proportion of “don’t know” responses to these questions.

Table 5. Percentage of respondents providing “don’t know” responses to substantive variables

| Module | N | Question | Group | Sample size | Percentage of respondents (%) | | | Comparisons (χ^2) | |
|----------------|----|--------------------------------------|---|-------------|-------------------------------|------------|--------|--------------------------|---------------------|
| | | | | | Full sample | Non-nudged | Nudged | Nudged vs. Non-nudged | Full vs. Non-nudged |
| Transport | 14 | Satisfaction with bus | Fairly satisfied/ Very satisfied | 11,053 | 20.0 | 20.3 | 19.8 | 0.4 | 0.2 |
| | 15 | Satisfaction with train | Fairly satisfied/ Very satisfied | 11,053 | 22.5 | 21.8 | 22.9 | 1.8 | 0.8 |
| Climate change | 17 | Is climate changing | Definitely/ Probably | 10,775 | 2.5 | 1.8 | 2.8 | 10.6** | 5.2* |
| | 18 | Cause of climate change | Entirely or Mainly by human activity | 10,671 | 3.4 | 2.7 | 3.8 | 9.2** | 4.5* |
| | 19 | Concern about climate change | Fairly concerned/ Very concerned | 10,877 | 1.6 | 1.0 | 1.9 | 13.1*** | 6.7* |
| | 20 | Impact of climate change in Wales | Already having an impact | 10,395 | 5.9 | 4.5 | 6.7 | 21.2*** | 10.5** |
| | 21 | Responsibility of the general public | A lot | 10,736 | 2.8 | 2.2 | 3.2 | 7.7** | 3.8 |
| | 22 | Responsibility of business | A lot | 10,674 | 3.4 | 2.7 | 3.8 | 8.5** | 4.2* |
| | 23 | Responsibility of government | A lot | 10,705 | 3.1 | 2.5 | 3.5 | 8.5** | 4.1* |
| | 24 | Government responding | A lot | 8,233 | 25.5 | 23.6 | 26.5 | 10.9** | 5.2* |

Note: The last two columns in the table present the results of two χ^2 tests. The second-to-last column compares the distribution of each variable among nudged addresses with that among non-nudged addresses. The final column compares the distribution of each variable among non-nudged addresses with that in the full sample. In both cases, significance levels are indicated as follows: ***p < 0.001, **p<0.01, *p<0.05.

This table demonstrates that there are no differences in the proportion of “don’t know” responses between nudged and non-nudged addresses in the questions about satisfaction with public transport. However, nudged respondents are significantly more likely to provide “don’t know” responses for all climate change-related questions.

5.4 Item non-response

Figure 2 illustrates item non-response for four indicators across the 41 “ask-all” questions analysed, distinguishing between nudged and non-nudged addresses for both the telephone and online components. For this analysis, the total number of respondents to the online component excludes those who did not participate in that component. The figure illustrates that “don't know” responses are the most common form of item non-response, and that all

four item non-response indicators are generally higher among nudged respondents compared to non-nudged respondents.

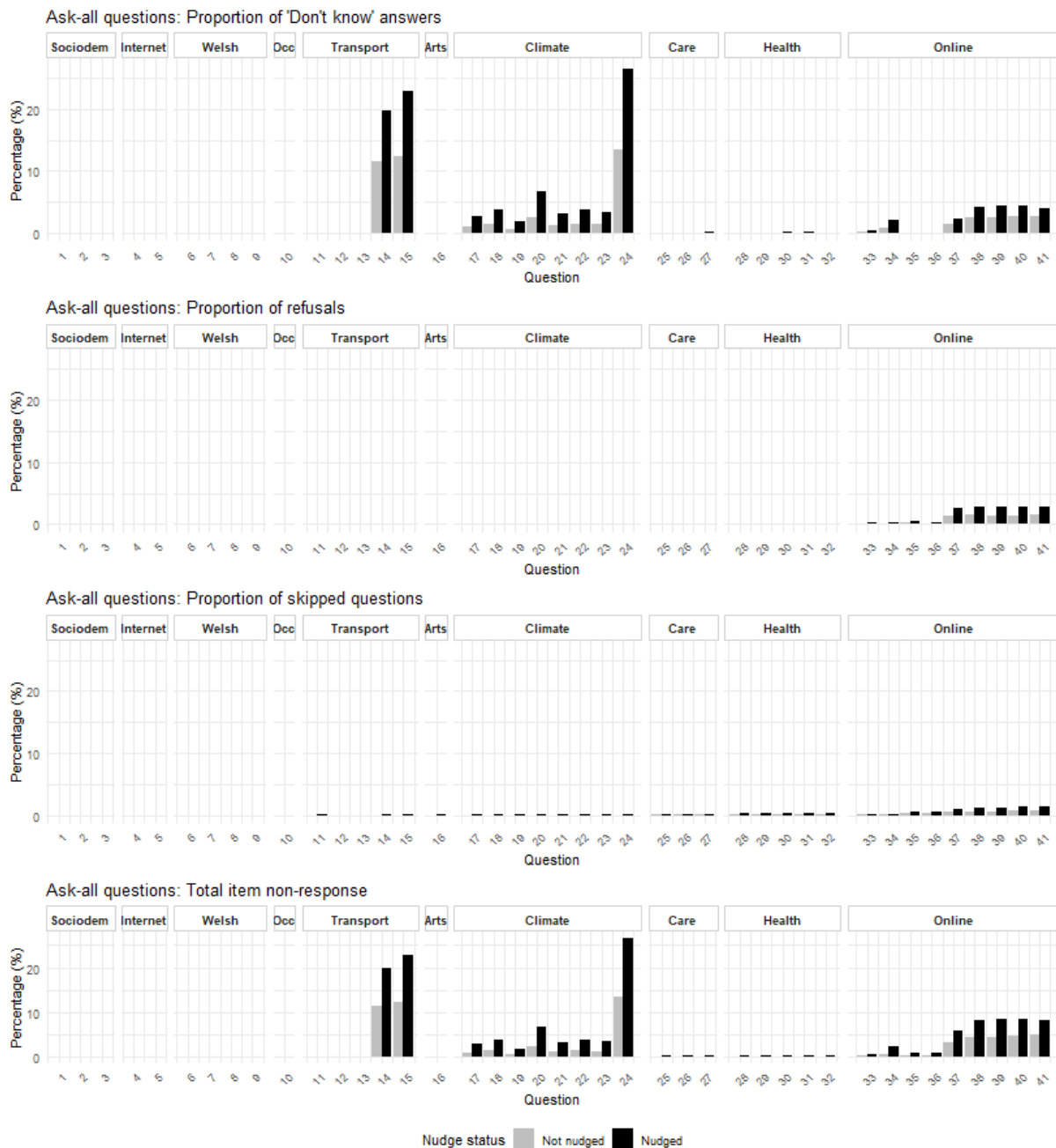


Figure 2. Indicators of item non-response for ask-all questions by survey stage

Three groups of questions demonstrate relatively high levels of “don’t know” responses. The first group comprises two questions (14 and 15) relating to satisfaction with buses and trains, which were asked of all participants regardless of whether they had used public transport in the previous 12 months. Among non-nudged participants, “don't know” responses to these questions range from 11.5% to 12.3%, whereas nudged participants show higher proportions, ranging between 19.8% and 22.9%. It should be noted that over 97% of “don't know”

responses to these questions were provided by participants who stated that they had not used trains or buses in the preceding 12 months. This suggests that “don't know” answers to these questions indicate a genuine lack of knowledge or an opinion on this topic, rather than data quality issues, mentioned in section 5.3.

The second group of questions with high levels of item non-response comprises the eight questions (17 to 25) about climate change. These exhibit the highest levels of item non-response overall. Specifically, the question “Thinking now about how the government is responding to climate change, do you think the government is doing...”, presented on a three-point Likert scale (“too little”, “enough” or “too much”), elicited 13.4% 'don't know' responses from non-nudged respondents and 26.5% from nudged respondents. Other questions within this module have item non-response levels between 0.5% and 1.5% for non-nudged respondents and between 1.9% and 6.7% for nudged respondents. As with the previous module, “don't know” answers to these questions may indicate a genuine lack of knowledge or opinion on this topic, rather than data quality issues.

The third group of questions with higher levels of item non-response comprises the “ask-all” questions in the online modules are questions 33 to 41. For nudged respondents, these questions show proportions of “don't know” responses of up to 4.5%, refusal levels of up to 2.9%, and skipped questions of up to 1.3%. Among non-nudged respondents, the maximum proportion of “don't know” responses was 2.8%, refusals reached up to 1.5%, and skipped questions reached up to 0.7%. In the remaining modules, item non-response levels are substantially lower, never exceeding 2.0%.

Table 6 demonstrates that respondents in the survey have a mean total level of item non-response of 2.20% across the 41 “ask-all” questions analysed. “Don't know” responses are the most frequent form of item non-response (mean value of 1.71% across “ask-all” questions), while the proportion of refusals reached a mean value of 0.57%, and the proportion of skipped questions reached 0.42%.

Table 6. Item non-response by survey stage and number of KtN visits

| Type of item non-response | KtN status | Number of KtN visits | Item non-response (%) | | Sample size |
|---------------------------|--------------|-----------------------|-----------------------|-------------------------|---------------|
| | | | Mean | 95% Confidence Interval | |
| Don't know | Non-nudged | No KtN visits | 1.51 | [1.42, 1.61] | 3,993 |
| | Nudged | All nudged | 1.82 | [1.73, 1.91] | 7,061 |
| | | One KtN visit | 1.76 | [1.66, 1.87] | 4,638 |
| | | Two KtN visits | 1.90 | [1.70, 2.10] | 1,450 |
| | | Three KtN visits | 1.90 | [1.61, 2.19] | 658 |
| | | No info on KtN visits | 2.14 | [1.67, 2.61] | 315 |
| | Total | | 1.71 | [1.64, 1.78] | 11,054 |
| Refusal | Non-nudged | No KtN visits | 0.40 | [0.34, 0.45] | 3,993 |
| | Nudged | All nudged | 0.66 | [0.60, 0.72] | 7,061 |
| | | One KtN visit | 0.61 | [0.54, 0.68] | 4,638 |
| | | Two KtN visits | 0.79 | [0.64, 0.93] | 1,450 |
| | | Three KtN visits | 0.78 | [0.55, 1.00] | 658 |
| | | No info on KtN visits | 0.56 | [0.27, 0.86] | 315 |
| | Total | | 0.57 | [0.52, 0.61] | 11,054 |
| Skipped | Non-nudged | No KtN visits | 0.27 | [0.21, 0.32] | 3,993 |
| | Nudged | All nudged | 0.50 | [0.45, 0.56] | 7,061 |
| | | One KtN visit | 0.47 | [0.41, 0.54] | 4,638 |
| | | Two KtN visits | 0.51 | [0.39, 0.64] | 1,450 |
| | | Three KtN visits | 0.72 | [0.49, 0.95] | 658 |

| Type of item non-response | KtN status | Number of KtN visits | Item non-response (%) | | Sample size |
|---------------------------|--------------|-----------------------|-----------------------|-------------------------|---------------|
| | | | Mean | 95% Confidence Interval | |
| | | | | No info on KtN visits | |
| | | Total | 0.42 | [0.38, 0.46] | 11,054 |
| Total item non-response | Non-nudged | No KtN visits | 1.92 | [1.80, 2.04] | 3,993 |
| | Nudged | All nudged | 2.36 | [2.25, 2.47] | 7,061 |
| | | One KtN visit | 2.29 | [2.15, 2.42] | 4,638 |
| | | Two KtN visits | 2.59 | [2.33, 2.85] | 1,450 |
| | | Three KtN visits | 2.26 | [1.93, 2.58] | 658 |
| | | No info on KtN visits | 2.51 | [1.99, 3.03] | 315 |
| | Total | 2.20 | [2.12, 2.28] | 11,054 | |

Consistent with our H4.1, there is a significant difference in the level of total item non-response between respondents in nudged addresses (2.36%) and those in non-nudged addresses (1.92%) as evidenced by the non-overlapping 95% confidence intervals for the two means.

5.5 Participation in online modules

Most participants responding to the main telephone survey also participate in the online modules ($n = 9,349$, 84.6% of the main survey respondents). The proportion of respondents in the online modules is significantly higher for the non-nudged addresses (89.7%) than for the nudged addresses (81.7%) ($p < 0.001$). Participation rates, along with their confidence intervals, are presented in Table 7, which demonstrates that non-nudged participants are significantly more likely to engage with the online modules than participants recruited via KtN, regardless of the number of visits. This aligns with H5.1. In addition, respondents recruited after three KtN visits are significantly less likely to complete the online modules than participants nudged once or twice. This confirms that participants requiring more recruitment effort are less likely to engage in additional activities in surveys.

Table 7. Item non-response per survey stage and number of KtN visits

| KtN status | Number of KtN visits | Participation in follow-up online survey (%) | | Sample size (Main survey) | Follow-up participants |
|--------------|-----------------------|--|-------------------------|---------------------------|------------------------|
| | | Mean | 95% Confidence Interval | | |
| Non-nudged | No KtN visits | 89.7 | [88.8, 90.7] | 3,992 | 3,582 |
| Nudged | One KtN visit | 82.5 | [81.4, 83.6] | 4,637 | 3,825 |
| | Two KtN visits | 81.0 | [79.0, 83.0] | 1,449 | 1,174 |
| | Three KtN visits | 75.5 | [72.2, 78.8] | 658 | 497 |
| | No info on KtN visits | 86.6 | [82.8, 90.4] | 313 | 271 |
| | All nudged | 81.7 | [80.8, 82.6] | 7,057 | 5,767 |
| Total | | 84.6 | [83.9, 85.3] | 11,049 | 9,349 |

Interestingly, the difference in completion of the online modules persists after controlling for socio-demographic variables, as shown by the logistic regression model (full results are presented in Table A2 in the Appendix). Specifically, the OR for the KtN status indicator is 0.55 ($p < 0.001$), which implies that nudged participants were significantly less likely to complete the online modules than non-nudged participants. This supports H5.1. We found no significant decline in completion of the online modules for households that responded after two KtN visits, compared to those that responded after having one visit only. Male participants

(compared to females), participants from ethnicities other than White British (compared to White British participants), and participants renting their homes from a private landlord or letting agency (compared to those who own their house) are also less likely to engage in the online modules.

6 Discussion

This paper examined the role of a KtN recruitment strategy in a telephone survey. Our findings demonstrate that, compared to initial recruitment alone, a KtN strategy effectively increases sample sizes and response rates and improves sample composition by incorporating participants from groups that are typically under-represented in social surveys, including younger individuals, renters, those without formal qualifications, and individuals from non-White British backgrounds.

The results also indicate that KtN enriches the sample and improves representativeness in a meaningful way, incorporating respondents who provide different answers to substantive questions. However, this comes with a trade-off, as these participants appear to be less engaged, demonstrating higher levels of item non-response and the valid “don’t know” responses and a lower propensity to participate in additional tasks such as completion of the online modules. Furthermore, our analysis suggests diminishing returns in improvements in response and sample composition after the first KtN visit, implying that a single visit may be sufficient to secure most of the benefits of this recruitment effort.

While the KtN approach shows clear promise for enhancing participation and representativeness in surveys that lack a face-to-face fieldwork component, key questions remain for researchers and survey practitioners. A robust experimental design would be valuable for isolating the causal mechanisms behind the improvements in response and representativeness. Ultimately, the decision to implement KtN will depend on a cost-benefit analysis, weighing the gains in sample quality and size against the substantial costs of deploying interviewers (around 30% of additional costs for fieldwork budget) and the potential data quality challenges associated with recruiting less-engaged respondents.

This work has a number of limitations. Although interviewers were instructed to conduct KtN visits at all non-responding addresses, some were never visited. This could have been due to a security assessment deeming the location unsafe, interviewers being refused access, or a lack of available interviewers in the local area. Sometimes interviewers did not follow the protocol and visited households, for which telephone numbers were already obtained, and sometimes the paradata has inaccuracies, which makes it difficult to establish how many visits specific households received among other issues. Additionally, we do not have the information about the reminder mailing dates and therefore it is sometimes not possible to attribute responses specifically to a reminder or to a visit. In future studies, it is important that all details of recruitment are carefully recorded in order to enable robust analysis of associations without introducing additional assumptions. Another limitation of this work was that only one indicator of data quality was analysed due to the specific types of questions included in the survey (e.g., no attitudinal questions are collected), and it would be beneficial to conduct a more extensive assessment of data quality in future studies.

Further research that incorporates detailed cost data is essential to properly establish the cost-effectiveness of this method and to explore whether its benefits in terms of participation and

representativeness outweigh the substantial financial costs and operational complexity and risks of inclusion of in-person fieldwork.

References

- AAPOR. (2023). *Standard Definitions Final Dispositions of Case Codes and Outcome Rates for Surveys*. <https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf>
- Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). Wiley.
- Cavazos, M., Ferrelly, M., & Levine, B. (2023). *Examining nonresponse bias in an address-based sample through an in-person nonresponse follow-up* Annual conference of the American Association of Public Opinion Research AAPOR, Philadelphia, PA, US.
- Cleary, A. (2023). *European Social Survey (ESS) - the impact of using fieldworkers to encourage response to the ESS* (28th ESS ERIC CST Meeting 7e). European Social Survey (ESS). https://www.europeansocialsurvey.org/sites/default/files/2024-09/Report_Fieldworkers_Andrew_Cleary.pdf
- Cornick, P., d'Ardenne, J., Maslovskaya, O., Mesplie-Cowan, S., Nicolaas, G., & Smith, P. (2022). *Review of options for the National Survey for Wales* (GSR report number 27/2022). <https://gov.wales/national-survey-wales-development-work#section-18671>
- Daikeler, J., Bošnjak, M., & Manfreda, K. L. (2020). Web versus other survey modes: An updated and extended meta-analysis comparing response rates. *Journal of Survey Statistics and Methodology*, 8, 513–539. <https://doi.org/10.1093/jssam/smz008>
- Dillman, D. A. (2020). Towards survey response rate theories that no longer pass each other like 'Strangers in the Night'. In P. S. Brenner (Ed.), *Understanding Survey Methodology: Sociological Theory and Applications* (Vol. 4, pp. 15–44). Springer.
- Dodge, N., & Chapman, R. (2018). Investigating recruitment and completion mode biases in online and door to door electronic surveys. *International Journal of Social Research Methodology*, 21(2), 149–163. <https://doi.org/10.1080/13645579.2017.1336596>
- Domarchi, C., Maslovskaya, O., & Smith, P. W. F. (2025). *Knock-to-nudge methods for recruitment in self-administered surveys: Evidence review* (Survey Futures Report, No. 8). Colchester, UK, U. o. Essex. <https://surveyfutures.net/reports/>
- Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2), 210–217. <https://doi.org/10.2307/2088328>
- Elis, J., Goerres, A., Mayer, S. J., & Spies, D. C. (2024). Comparing the mobilising effects of in-person canvassing to postal reminders – experimental evidence from a longitudinal election study. *International Journal of Social Research Methodology*, 27(5), 605–611. <https://doi.org/10.1080/13645579.2023.2189804>

- Eurostat. (2020). *Data collection for the EU-Labour Force Survey in the context of the COVID-19 crisis*. https://ec.europa.eu/eurostat/documents/10186/10693286/LFS_guidance.pdf
- Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5), 646–675. <https://doi.org/10.1093/poq/nfl033>
- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. John Wiley & Sons.
- Jäckle, A., Burton, J., Couper, M. P., & Perelli-Harris, B. (2023). *Participation of household panel members in daily burst measurement using a mobile app: effects of position of the invitation, bonus incentives, and number of daily questions* (Understanding Society Working Paper Series, No. 2023-01). <https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2023-01/>
- Jenkins, S., & Forbes, N. (2024). *National Survey for Wales 2022-23 and 2021-23 Combined Years Data Technical Report*. Welsh Government, Office for National Statistics. <https://www.gov.wales/sites/default/files/statistics-and-research/2024-02/national-survey-for-wales-2022-23-and-2021-23-combined-years-data--technical-report.pdf>
- Kastberg, S., & Siegler, V. (2022). *Impact of COVID-19 on ONS social survey data collection*. Office for National Statistics. Retrieved 16 January 2024 from <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/impactofcovid19ononssocialsurveydatacollection>
- Kunz, T., Daikeler, J., & Ackermann-Piek, D. (2024). Interviewer-observed paradata in mixed-mode and innovative data collection. *International Journal of Market Research*, 66(1), 14–26. <https://doi.org/10.1177/14707853231184742>
- Link, M. W., Murphy, J., Schober, M. F., Buskirk, T. D., Childs, J. H., & Tesfaye, C. L. (2014). Mobile technologies for conducting, augmenting and potentially replacing surveys: Executive summary of the AAPOR task force on emerging technologies in public opinion research. *The Public Opinion Quarterly*, 78(4), 779–787. <https://www.jstor.org/stable/24545969>
- Luijckx, R., Jónsdóttir, G. A., Gummer, T., Stähli, M. E., Frederiksen, M., Ketola, K., Reeskens, T., Brislinger, E., Christmann, P., Gunnarsson, S. Þ., Hjaltason, Á. B., Joye, D., Lomazzi, V., Maineri, A. M., Milbert, P., Ochsner, M., Pollien, A., Sapin, M., Solanes, I., Verhoeven, S., & Wolf, C. (2021). The European Values Study 2017: On the way to the future using mixed-modes. *European Sociological Review*, 37(2), 330–346. <https://doi.org/10.1093/esr/jcaa049>
- Maslovskaya, O., Domarchi, C., & Smith, P. W. F. (2025). *Effectiveness of the knock-to-nudge approach for establishing contact with the respondents: Evidence from the National Readership Survey in the UK* (Survey Futures Working Paper, No. 10). Colchester, UK, U. o. Essex. <https://surveyfutures.net/working-papers/>

- Maslovskaya, O., & Lugtig, P. (2022). Representativeness in six waves of Cross-National Online Survey (CRONOS) Panel *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(3), 851–871. <https://doi.org/10.1111/rssa.12801>
- Maslovskaya, O., Lynn, P., Calderwood, L., Durrant, G., Fitzgerald, R., Nicolaas, G., & Williams, J. (2025). *Survey Futures Position Statement on Response Rates* (Survey Futures Position Statement, No. 1). https://surveyfutures.net/wp-content/uploads/2025/06/Response-Rates-Posititon-Statement_SurveyFutures.pdf.
- Millar, M. M., & Dillman, D. A. (2011). Improving response to Web and mixed-mode surveys. *Public Opinion Quarterly*, 75(2), 249–269. <https://doi.org/10.1093/poq/nfr003>
- Mitchell, J., & Jäckle, A. (2026). *How panel members react to multiple requests for additional data over time* (Understanding Society Working Paper Series, No. 2026/01). <https://www.understandingsociety.ac.uk/wp-content/uploads/working-papers/2026-01.pdf>
- Moore, J. C., Durrant, G. B., & Smith, P. W. F. (2018). Data set representativeness during data collection in three UK social surveys: Generalizability and the effects of auxiliary covariate choice *Journal of the Royal Statistical Society Series A: Statistics in Society*, 181(1), 229–248. <https://doi.org/10.1111/rssa.12256>
- Nicolaas, G. (2002). *Within-household selection for push-to-web surveys*. GenPopWeb2 Report. <https://www.ncrm.ac.uk/documents/Within%20household%20selection%20for%20push%20to%20web%20surveys.pdf>
- Office for National Statistics. (2011). *2011 Residential-based Area Classifications*. Office for National Statistics,. Retrieved 9 October 2025 from <https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications>
- Office for National Statistics. (2015). *Pen Portraits for the 2011 Area Classification for Output Areas*. Retrieved 20 June 2024 from <https://www.ons.gov.uk/file?uri=/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications/penportraitsandradialplots/penportraits.pdf>
- Office for National Statistics. (2021). *2021 Census [Create a custom dataset]*. UK Data Service. <https://www.ons.gov.uk/datasets/create>
- Office for National Statistics, & Welsh Government. (2024). *National Survey for Wales, 2022-2023* (SN: 9144; Version 2) [Data collection]. <https://doi.org/10.5255/UKDA-SN-9144-2>
- Olson, K. (2013). Do non-response follow-ups improve or reduce data quality?: A review of the existing literature. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 176(1), 129–145. <https://doi.org/10.1111/j.1467-985X.2012.01042.x>

- Revilla, M., Couper, M. P., & Ochoa, C. (2019). Willingness of online panelists to perform additional tasks. *Methods Data Analyses*, 13(2), 223–251. <https://doi.org/10.12758/mda.2018.01>
- Rohr, B., & Felderer, B. (2025). *sampcompR: Comparing and visualizing differences between surveys*. In *R Package* (Version 0.3.2) <https://github.com/bjoernrohr/sampcompr>.
- Rohr, B., Silber, H., & Felderer, B. (2024). Comparing the accuracy of univariate, bivariate, and multivariate estimates across probability and nonprobability surveys with population benchmarks. *Sociological Methodology*, 55(1), 121–154. <https://doi.org/10.1177/00811750241280963>
- Salmon, C. T., & Nichols, J. S. (1983). The next-birthday method of respondent selection. *The Public Opinion Quarterly*, 47(2). <http://www.jstor.org/stable/2749026>
- Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the representativeness of survey response *Survey Methodology*, 35(1), 101–113.
- Vine, J., Jäckle, A., Burton, J., & Couper, M. P. (2025). *Getting consent to survey in new ways: evidence from experiments about questions by text message in the Understanding Society Innovation Panel* (Understanding Society Working Paper Series, No. 2025-14). <https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2025-14/>
- West, B. T., Zhang, S., Wagner, J., Gatward, R., Saw, H.-W., & Axinn, W. G. (2023). Methods for improving participation rates in national self-administered web/mail surveys: Evidence from the United States. *PLoS ONE*, 18(8), e0289695. <https://doi.org/10.1371/journal.pone.0289695>
- Wolf, C., Christmann, P., Gummer, T., Schnaudt, C., & Verhoeven, S. (2021). Conducting general social surveys as self-administered mixed-mode surveys. *Public Opinion Quarterly*, 85(2), 623–648. <https://doi.org/10.1093/poq/nfab039>
- Yan, T., & Williams, D. (2022). Response Burden – Review and Conceptual Framework. *Journal of Official Statistics*, 38(4), 939–961. <https://doi.org/10.2478/jos-2022-0041>

Appendix

Table A1. Questions included in the item non-response and substantive analysis

| Mode | Module | Order | Variable name | Question | Variable type | Group selected for analysis |
|-----------|-----------------------|-------|-------------------|---|---------------|---|
| Telephone | Sociodemographic | 1 | Religion | <i>What is your religion?</i> | Single choice | No religion |
| | | 2 | ResTme | <i>How long have you lived at this address?</i> | Single choice | Five years or more |
| | | 3 | HSat | <i>And, how satisfied are you with this accommodation?</i> | Single choice | Fairly satisfied / Very satisfied |
| | Internet | 4 | IntHhHave | <i>Does your household have access to the internet at home?</i> | Binary | Yes |
| | | 5 | IntPersUse | <i>Do you personally use the internet at home, work, or elsewhere?</i> | Binary | Yes |
| | Welsh | 6 | WelUndSpk | <i>Can you understand spoken Welsh?</i> | Binary | Yes |
| | | 7 | WelSpk | <i>Can you speak Welsh?</i> | Binary | Yes |
| | | 8 | WelRead | <i>Can you read Welsh?</i> | Binary | Yes |
| | | 9 | WelWrite | <i>Can you write Welsh?</i> | Binary | Yes |
| | Occupation/Education | 10 | HighEd | <i>Are you currently studying for a university degree or equivalent level qualification?</i> | Binary | Yes |
| | Transport | 11 | CarUse | <i>Is there a car or van normally available for use by you or anyone in your household?</i> | Binary | Yes |
| | | 12 | Bus12M | <i>In the last 12 months, have you used bus services in Wales?</i> | Binary | Yes |
| | | 13 | Train12M | <i>In the last 12 months, have you used train services in Wales?</i> | Binary | Yes |
| | | 14 | BusOverSat | <i>Overall, how satisfied or dissatisfied are you with bus services? We're interested in your views, even if you don't use buses.</i> | Single choice | Fairly satisfied / Very satisfied |
| | | 15 | TrainOverSat | <i>Overall, how satisfied or dissatisfied are you with train services? We're interested in your views, even if you don't use trains.</i> | Single choice | Fairly satisfied / Very satisfied |
| | Arts/Culture/Heritage | 16 | ArchAtt | <i>In the last 12 months, have you used an archive or records office in Wales or elsewhere, either in person or online?</i> | Binary | Yes |
| | Climate change | 17 | CliChanView | <i>You may have heard the idea that the world's climate is changing due to increases in temperature over the past 100 years. What is your personal opinion on this? Is it...</i> | Single choice | Definitely changing / Probably changing |
| | | 18 | CliChanCause2 | <i>Do you think that climate change is caused by natural processes, human activity, or both?</i> | Single choice | Entirely by human activity / Mainly by human activity |
| | | 19 | CliChanCon | <i>How concerned, if at all, are you about climate change?</i> | Single choice | Fairly concerned / Very concerned |
| | | 20 | CliChanWhen | <i>When, if at all, do you think climate change will start to have an impact in Wales?</i> | Single choice | Already having an impact |
| | | 21 | CliChanRespPublic | <i>How much responsibility do you think each of the following has to tackle climate change in Wales?: The general public</i> | Single choice | A lot |
| | | 22 | CliChanRespBus | <i>How much responsibility do you think each of the following has to tackle climate change in Wales?: Business</i> | Single choice | A lot |
| | | 23 | CliChanRespGov | <i>How much responsibility do you think each of the following has to tackle climate change in Wales?: The government</i> | Single choice | A lot |
| | Social care | 24 | CliChanGov | <i>Thinking now about how the government is responding to climate change, do you think the government is doing...</i> | Single choice | A lot |
| | | 25 | Carer | <i>Do you look after or give any help or support to family members, friends, neighbours or others because of long-term physical or mental ill-health or disability, or problems related to old age?</i> | Binary | Yes |
| | | 26 | SCUseYN | <i>In the last 12 months, have you received help for yourself from care and support services in Wales?</i> | Binary | Yes |

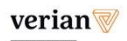
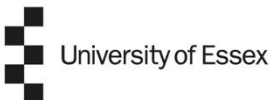
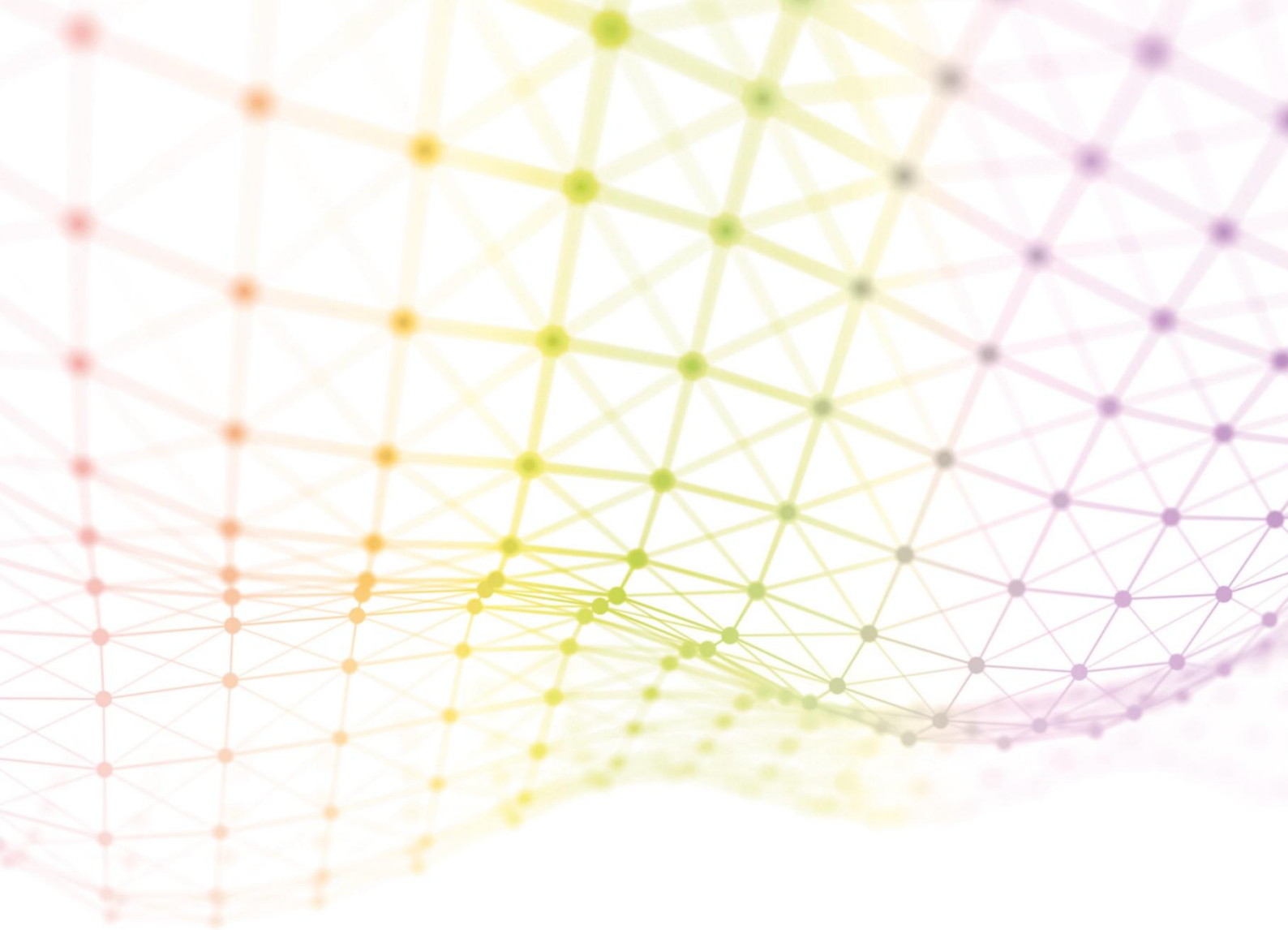
| Mode | Module | Order | Variable name | Question | Variable type | Group selected for analysis |
|--------|---------------------------------|-------|----------------|--|---------------|---|
| | | 27 | SCUse2 | <i>In the last 12 months have you received any help from care and support services in Wales to care for, or arrange care for, someone else?</i> | Binary | Yes |
| | Health/GP/Hospital | 28 | GenHealth | <i>How is your health in general; is it...</i> | Single choice | Good / Very good |
| | | 29 | GpSeenDr | <i>Thinking about the last 12 months, have you had an appointment with a GP or family doctor about your own health?</i> | Binary | Yes |
| | | 30 | GpSeenPract | <i>Have you had an appointment with any other health professional based at your GP surgery in the last 12 months</i> | Binary | Yes |
| | | 31 | HspHadApp | <i>In the last 12 months, have you had an NHS hospital appointment?</i> | Binary | Yes |
| | | 32 | HspAE | <i>have you attended an Accident and Emergency or a Minor Injury Unit for your own health?</i> | Binary | Yes |
| Online | Recycling | 33 | UaRecycQual | <i>How satisfied or dissatisfied are you with the recycling collection service provided by [council]?</i> | Single choice | Fairly satisfied / Very satisfied |
| | | 34 | UaRecycComm | <i>[council] keeps me informed about its recycling collection service</i> | Single choice | Strongly agree / Tend to agree |
| | | 35 | RecycHHClo | <i>In the last 12 months, have you carried out or arranged a repair or alteration of clothing that would otherwise have been unused or thrown away?</i> | Binary | Yes |
| | | 36 | RecycHHFrn | <i>In the last 12 months, have you carried out or arranged a repair of any household items, such as furniture, a fridge or a kettle, that would otherwise have been unused or thrown away?</i> | Binary | Yes |
| | Physical punishment of children | 37 | SmackAgree | <i>To what extent do you agree or disagree that it is sometimes necessary to smack a child?</i> | Single choice | Tend to agree / Strongly agree |
| | | 38 | SmackHarmSelf | <i>Which of these statements comes closest to your opinion on whether it's appropriate for a parent to smack a child in the following circumstances? To stop them doing something which is dangerous or harmful to themselves</i> | Single choice | Yes, whenever necessary / Yes, as a last resort |
| | | 39 | SmackHarmAnoth | <i>Which of these statements comes closest to your opinion on whether it's appropriate for a parent to smack a child in the following circumstances? To stop them doing something which is dangerous or harmful to another child</i> | Single choice | Yes, whenever necessary / Yes, as a last resort |
| | | 40 | SmackOOC | <i>Which of these statements comes closest to your opinion on whether it's appropriate for a parent to smack a child in the following circumstances? To stop behaviour that's out of control – for example, a tantrum or meltdown.</i> | Single choice | Yes, whenever necessary / Yes, as a last resort |
| | | 41 | SmackPun | <i>Which of these statements comes closest to your opinion on whether it's appropriate for a parent to smack a child in the following circumstances? As a punishment for misbehaving.</i> | Single choice | Yes, whenever necessary / Yes, as a last resort |

Note: In this table, "Order" indicates the question number appearing in the item non-response plots.

Table A2. Logistic regression model of participation in online modules – Full results

| Variable | Category | Beta | SE | OR |
|------------------------|--------------------------------|-----------|-------|-------|
| Knock-to-nudge status | Non-nudged (Reference) | - | - | - |
| | Nudged | -0.598*** | 0.063 | 0.550 |
| Sex | Female (Reference) | - | - | - |
| | Male | -0.235*** | 0.055 | 0.791 |
| Age | 16-24 (Reference) | - | - | - |
| | 25-34 | 0.472 | 0.152 | 1.116 |
| | 35-49 | 0.133 | 0.151 | 1.255 |
| | 50-64 | 0.316 | 0.157 | 1.170 |
| | 65+ | 0.402 | 0.163 | 1.147 |
| Highest qualifications | Level 1 (Reference) | - | - | - |
| | Level 2 | 0.333* | 0.140 | 1.395 |
| | Level 3 | 0.330* | 0.150 | 1.390 |
| | Level 4 | 0.317* | 0.128 | 1.373 |
| | No qualifications | 0.112 | 0.139 | 1.247 |
| | Others | 0.283* | 0.140 | 1.327 |
| Ethnicity | White British | - | - | - |
| | Other (Reference) | -0.570*** | 0.131 | 0.565 |
| Marital status | Single (Reference) | - | - | - |
| | Married/Civil partnership | 0.188 | 0.087 | 1.121 |
| | Separated | 0.097 | 0.166 | 0.759 |
| | Divorced | 0.913 | 0.099 | 1.011 |
| | Widowed | 0.630 | 0.115 | 1.057 |
| Household size | 1 | - | - | - |
| | 2 | 0.885 | 0.084 | 1.012 |
| | 3 | 0.958 | 0.102 | 0.995 |
| | 4+ | 0.408 | 0.109 | 0.914 |
| Tenure | Owned (Reference) | - | - | - |
| | Private rented | 0.510 | 0.261 | 1.187 |
| | Other private rented/Rent free | -0.258** | 0.085 | 0.772 |
| | Rented from council/LA | 0.261 | 0.098 | 0.896 |
| | Other social rented | 0.266 | 0.126 | 0.869 |
| Wales region | Blaneau Gwent (Reference) | - | - | - |
| | Bridgend | 0.586 | 0.207 | 0.893 |
| | Caerphilly | 0.497 | 0.204 | 1.149 |
| | Cardiff | 0.775 | 0.184 | 1.054 |
| | Carmarthenshire | 0.273 | 0.197 | 0.805 |
| | Ceredigion | 0.656 | 0.241 | 1.113 |
| | Conwy | 0.099 | 0.209 | 0.709 |
| | Denbighshire | 0.350 | 0.220 | 0.814 |
| | Flintshire | 0.270 | 0.203 | 0.799 |
| | Gwynedd | 0.651 | 0.219 | 1.104 |
| | Isle of Anglesey | 0.206 | 0.218 | 0.759 |
| | Merthyr Tydfil | 0.470 | 0.218 | 0.855 |
| | Monmouthshire | 0.094 | 0.240 | 1.496 |
| | Neath Port Talbot | 0.149 | 0.193 | 0.757 |
| | Newport | 0.314 | 0.212 | 1.238 |
| | Pembrokeshire | 0.129 | 0.212 | 0.725 |
| | Powys | 0.877 | 0.193 | 1.030 |
| | Rhondda Cynon Taf | -0.449* | 0.183 | 0.638 |
| | Swansea | 0.589 | 0.193 | 1.110 |
| | Torfaen | 0.739 | 0.227 | 1.079 |
| Vale of Glamorgan | 0.112 | 0.204 | 0.723 | |
| Wrexham | 0.240 | 0.209 | 0.782 | |
| Urban/Rural | Rural (Reference) | - | - | - |
| | Urban | 0.753 | 0.069 | 0.979 |
| Intercept | | 1.921*** | 0.262 | 6.826 |
| Log-likelihood | -4611.696 (<i>df</i> = 46) | | | |
| AIC | 9315.4 | | | |

Note: OR – Odds Ratio, SE – Standard Error.



www.surveyfutures.net