SURVEY FUTURES

SURVEY DATA COLLECTION METHODS COLLABORATION

Working Paper 26:

Within-household clustering of attitudes and implications for survey sample design when only address-based frames are available

Nhlanhla Ndebele[1], Peter Lynn[2], Rory Fitzgerald[1], Ruxandra Comanaru[1]

[1] City St George's, University of London, [2] University of Essex

July 2026

www.surveyfutures.net

Acknowledgements

Survey Futures is an Economic and Social Research Council (ESRC)-funded initiative (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.surveyfutures.net.

The research reported here forms part of the within-household selection subproject in Survey Futures Research Strand 4: Methods for surveys without field interviewer, led by Professor Olga Maslovskaya.

Citation

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:

Ndebele N., Lynn P., Fitzgerald R. & Comanaru R. (2026) 'Within-household clustering of attitudes and implications for survey sample design when only address-based frames are available', *Survey Futures Working Paper* no. 26. Colchester, UK: University of Essex. Available at https://surveyfutures.net/working-papers/.

Contents

Within-household clustering of attitudes and implications for survey sample design when only address-based frames are available

Nhlanhla Ndebele, Peter Lynn, Rory Fitzgerald, Ruxandra Comanaru

Abstract

As many social surveys transition from interviewer-administered to self-completion data collection, there is increasing debate about the most appropriate methods for selecting individuals within households when only address-based sampling frames are available. A pertinent consideration is the extent to which selecting multiple respondents per household influences the precision of survey estimates through its effects on within-household response clustering and sampling variance. For opinion, values, and attitudinal surveys, there is little empirical evidence on the extent of intra-household correlations and their implications for sampling variance and statistical efficiency, limiting evidence-based sample design choices.

This paper addresses this gap using data from the 2021 British Social Attitudes survey. It quantifies the magnitude of within-household response clustering across attitudinal measures, examines its contribution to sampling variance, and compares the statistical implications of selecting one versus two respondents per household. Findings show that although selecting two respondents introduces additional sampling variance through household-level clustering, this is offset by variation in the contribution of unequal selection probabilities to the overall design effect, resulting in modest differences in overall statistical efficiency between the two approaches. These results provide evidence to inform sample design decisions about within-household respondent selection in self-completion household surveys using address-based sampling frames.

Keywords: address-based sampling frames; design effect; intra-class correlation; self-completion surveys; within-household selection methods.

1 Introduction

In countries where individual-level sampling frames are not available for general population surveys, address-based sampling frames are used instead (Maslovskaya et al. 2022). But for surveys of individuals, this introduces the need for an extra stage of sampling, to select individuals at sampled addresses. In face-to-face surveys, selection of respondents within household is carried out by field interviewers. Usually, one individual per household is selected using a random procedure (Koch 2019).

When using address-based samples, household size distribution plays a key role in determining the most appropriate within-household selection method. This is because in one-person households no selection is required, whereas for larger households selection can become more complex and burdensome. In recent years, general declines in survey response rates and rising costs of interviewer fieldwork, together with technological advances in survey data collection (Charman, Mesplie-Cowan, and Collins 2025; Maslovskaya et al. 2025), among other factors, have led many social surveys to transition, or consider transitioning, from interviewer-administered to self-completion methods. The pace of this transition was accelerated by the COVID-19 pandemic, which reduced the interviewer field force (Charman et al. 2025). In the absence of field interviewers, the responsibility for within-household selection rests with sampled households themselves, giving survey organisations limited control over the process of selecting individual respondents (Olson et al. 2019). A number of studies have shown that when households are asked to make a random selection of one person, the person completing the survey is often not the person who should have been selected (Olson, Stange, and Smyth 2014; Williams 2015). This has led to consideration of designs that allow more than one respondent per household, as such designs have potential to greatly reduce the number of 'incorrect' respondents. There are many such possible designs, differing in how individuals are selected within households, and in how many individuals should be selected per household (Clark and Steel 2007; Gaziano 2005; Maslovskaya et al. 2022). These choices have important implications for survey costs and for precision of estimation.

However, there is little literature on how many people should be selected (Clark and Steel 2007). Methods that select multiple respondents per household improve fieldwork efficiency by substantially reducing the number of households that need to be sampled to achieve a target number of responses within a fixed budget (Purdon and Wood 2022). However, one of the drawbacks of these approaches is that they can introduce design effects due to homogeneous survey responses among household members, which reduce the precision of estimates (Clark and Steel 2007; Purdon and Wood 2022). By contrast, methods that select a single respondent per household avoid within-household clustering but incur higher survey costs (Purdon and Wood 2022). Furthermore, selecting a single respondent per household can lead to greater variation in design weights (which are inversely proportional to within-household selection probabilities) and hence a precision loss that may partially counter-balance the gain from not introducing within-household clustering.

In this study, we provide evidence on within-household homogeneity in subjective survey measures and outline the implications for choice of within-household selection method. Section 2 provides an overview of these methods, including probability, quasi-probability, and non-probability approaches, and their practical trade-offs. Section 3 outlines statistical issues

arising when multiple respondents are included per household. Section 4 introduces the data and analytic methods, while Section 5 presents empirical findings on within-household similarity. Section 6 concludes with a discussion of the results and their broader implications for designing cost-effective high quality self-completion surveys.

2 Methods for within-household respondent selection

There are several methodological, operational, and analytical factors to consider when choosing a within-household selection method in surveys that use address-based sampling frames, and these factors affect sample representativeness, respondent burden, and data quality. In general, within-household selection methods can contribute to total survey error in multiple ways (Smyth, Olson, and Stange 2019). In particular, selection methods may introduce error through undercoverage, selection error, and nonresponse.

First, all eligible members of the household need to be considered during the within-household selection process (Smyth et al. 2019). If some eligible members are excluded, undercoverage may occur, while if specific groups are systematically excluded and their characteristics relate to the concepts being measured, this will increase coverage bias in survey estimates (Groves et al. 2009; Smyth et al. 2019). Second, assuming all eligible household members are considered, the selection instructions must be followed accurately. Selecting the wrong household member, whether mistakenly or intentionally, can increase sampling error (Smyth et al. 2019). Third, nonresponse error may arise if the selection procedure discourages certain types of households or selected individuals from participating. This would increase the risk of nonresponse bias if non-respondents differ systematically from respondents (Groves et al. 2009).

Within-household selection methods can be classified into three groups based on the probability characteristics of the resulting sample: (1) probability methods, (2) quasi-probability methods, and (3) non-probability methods (Gaziano 2005; Marlar et al. 2018).

Probability methods randomly select one (or more) eligible respondents from all members of a sampled household, such that each eligible individual has a known, non-zero probability of selection. In face-to-face surveys, a commonly used probability method is the Kish grid (Díaz de Rada 2021). The interviewer enumerates all eligible household members and selects one respondent according to a predefined randomisation rule, providing a clear framework for estimating sampling variance and limiting selection bias (Díaz de Rada 2021; Gaziano 2005). Evidence from face-to-face surveys indicates that the Kish grid generally reduces selection bias, but the enumeration process can be perceived as intrusive and burdensome, potentially increasing nonresponse bias (Díaz de Rada 2021; Gaziano 2005; Marlar et al. 2018; Olson et al. 2019; Smyth et al. 2019; Yan 2009), although various simplifications of Kish's original method that are less intrusive have been developed (Jabkowski 2017; Lynn and Lievesley 1991).

Quasi-probability methods aim to approximate random selection but do not fully ensure known, non-zero, and equal probabilities of selection for all eligible members. Birthday-based selection methods are quasi-probability methods that have also been used in face-to-face surveys (Jabkowski 2017; Jabkowski, Cichocki, and Kołczyńska 2024). They offer a simpler, less intrusive alternative as the interviewer asks to speak to the household member with the next, last, or closest birthday (Gaziano 2005; Marlar et al. 2018). While birthday-based selection methods reduce burden and improve cooperation, this comes at the expense of true random sampling (Díaz de Rada 2021; Olson et al. 2019; Smyth et al. 2019) as births may not be randomly distributed across months of the year and selection is anchored to the fieldwork

period (Gaziano 2005; Olson et al. 2019). This may introduce selection bias in households where more than one person is eligible and variables of interest are related to the dates of birth or seasonality (Gaziano 2005; Smyth et al. 2019).

In self-completion surveys, additional challenges arise because there is no interviewer to motivate participation, clarify instructions, or verify that the selection procedure has been implemented correctly. Probability-based methods are rarely feasible in this context, as enumeration relies on a household informant who may misreport or omit eligible members, and the selection procedure may be misunderstood or ignored (Olson et al. 2019; Smyth et al. 2019). Quasi-probability methods, such as birthday-based selection, are also used in self-completion surveys because they reduce burden (Battaglia et al. 2008), but they are prone to selection inaccuracy if the household informant does not know the birthdates of all eligible members. In addition, recruitment of the selected respondent may be challenging, with evidence suggesting accuracy selection rates from the mid-50% to the mid-70% (Olson and Smyth 2017; Smyth et al. 2019; Stange, Smyth, and Olson 2016; Williams 2015).

Finally, non-probability approaches that allow any multiple adults (e.g., any two, any three, any four) in a household to respond are commonly used in self-completion surveys (Ndebele et al. forthcoming) because of their simple instructions, low respondent burden, and cost-efficiency. In non-probability methods, selection of respondents within households is not governed by a random mechanism, and individual inclusion probabilities are unknown, with some eligible members having a low chance of selection (Gaziano 2005; Marlar et al. 2018; Olson et al. 2019; Smyth et al. 2019). Although these methods are non-probability approaches, this limitation only affects a part of the sample, depending on household size and the number of respondents selected per household. Only when the number of respondents requested to participate per household is smaller than the number of eligible adults in the sampled household does selection become non-random. For example, if a survey requests up to two adults to participate in each household, it is only amongst households with at least three eligible people that selection bias can arise.

Overall, the choice of within-household selection method in self-completion surveys requires careful consideration of trade-offs between respondent burden, selection accuracy, coverage, and nonresponse. Probability-based methods provide clear, measurable selection probabilities but are generally infeasible in self-completion modes, whereas quasi-probability and non-probability approaches offer practical alternatives. Quasi-probability methods, including birthday-based selection, balance reduced burden with partial randomisation but require attention to potential inaccuracies in self-reported household information. Non-probability approaches that allow multiple adults to respond provide flexibility and efficiency, with bias limited to households with more eligible members than requested, making these approaches an alternative for large-scale self-completion surveys.

This study investigates the implications of selecting multiple respondents per household for survey sample design in self-completion surveys that use address-based sampling frames. The study addresses the following research questions:

1) What is the extent of correlation in responses to attitudinal survey questions among respondents from the same household?

2) What is the impact of intra-household correlation on sampling variance when the any-two respondent selection method is used?

3) What are the overall implications for sampling variance and statistical efficiency of selecting two people per household rather than one?

3 Sample clustering effects with multiple-respondent designs

While simple random sampling designs yield samples that most closely approximate the assumptions that observations are independent and identically distributed, assumptions that underpin estimation and inference in most statistical software, such designs are rarely used in survey data collection for both theoretical and practical reasons (Heeringa, West, and Berglund 2017; Kish, Groves, and Krotki 1976). In practice, large-scale scientific surveys employ complex sample designs, incorporating features such as stratification to improve statistical and administrative efficiency, cluster sampling to enhance fieldwork efficiency and reduce costs, and weighting to account for unequal probabilities of selection (Heeringa et al. 2017). While these design features offer clear advantages, they also affect the accuracy and precision of survey estimates. In particular, relative to simple random sampling, complex design features influence the magnitude of standard errors (Heeringa et al. 2017; Kish et al. 1976). At a given sample size, stratification typically reduces standard errors, whereas clustering of sample elements and the use of weights generally increase standard errors compared to those obtained under simple random sampling (Heeringa et al. 2017).

The combined effect of a complex sample design on the standard error of an estimate, relative to that from a simple random sample of equal size, is captured by the *design effect* or *design efficiency factor (DEFF)* (Heeringa et al. 2017; Kish et al. 1976). This is defined as the ratio of the actual sampling variance, accounting for the complexity of the sampling design, to the variance that would be expected under a simple random sample of equal size (Kish 1965):

$$DEFF\ for\ sample\ estimate = \frac{Complex\ sample\ design\ variance\ of\ estimate}{Simple\ random\ sample\ design\ variance\ of\ estimate} \quad (1)$$

The value of the design effect for a particular sample design reflects the combined influences of stratification, cluster sampling, and weighting (Heeringa et al. 2017). A design effect equal to one indicates that the complex design is as efficient as a simple random sample. When the design effect is greater than one, the complex design is less efficient, resulting in larger standard errors than would be expected under a simple random sample. Conversely, a design effect less than one indicates that the complex design is more efficient, producing smaller standard errors compared with a simple random sample (Heeringa et al. 2017; Kish 1965).

We are particularly concerned here with the effect of clustering. When responses from individuals within the same cluster are meaningfully correlated, the total information obtained from a sample of size *n* is less than *n* times the information from a single individual (Kish et al. 1976). For example, when multiple respondents are selected within the same household, their responses may be positively correlated because they share similar environments or characteristics, which reduces the overall statistical efficiency of the sampling design. The key

concern, therefore, is the extent of this loss in efficiency, particularly as a trade-off against the cost savings achieved through clustering (Kish et al. 1976).

A useful way to quantify the extent of this within-cluster similarity is through the *intra-class correlation* (*ICC*, or ρ [rho]), which measures the degree of homogeneity among sample units within the same cluster (Heeringa et al. 2017; Kish 1965; Kish et al. 1976). In surveys that select multiple respondents within the same household, the intra-class correlation represents the average correlation between those individuals and indicates how strongly their responses are related. For the whole survey design, a simple model links the design effect to the intra-class correlation as follows:

$$\rho = \frac{DEFF - 1}{\bar{b} - 1} \tag{2}$$

and

$$DEFF = 1 + \left(\bar{b} - 1\right) \rho \tag{3}$$

where $\bar{b}$ is the average cluster size (Heeringa et al. 2017; Kish et al. 1976). The synthetic estimate of the intra-class correlation is constrained to range between –1 / ($\bar{b}$ – 1) and 1.0 (Heeringa et al. 2017). According to Kish et al. (1976), ρ values close to zero indicate minimal within-cluster correlation, consistent with variables that are approximately randomly distributed. In contrast, ρ values in the range of 0.10 to 0.20 are considered relatively high, as this level of clustering can substantially inflate the design effect and reduce sampling efficiency. In the extreme case, where all individuals within a cluster have identical responses for the variable, ρ approaches 1.0, indicating perfect within-cluster similarity.

4 Data and methods

4.1 Data and sample

This study uses data from the 2021 British Social Attitudes (BSA) survey (NatCen Social Research 2024c), which is designed to yield a representative sample of adults aged 18 or over living in private households in Great Britain. The BSA is a repeated cross-sectional survey that collects data on respondents' attitudes across a range of topics and serves as an important barometer of Britain's social, moral, and political attitudes (NatCen Social Research 2024b). The sampling design differed from pre-2020 waves, which used interviewer-administered face-to-face methods, reflecting the transition to primarily online data collection following the COVID-19 pandemic (Clery et al. 2021; Curtice, Hudson, and Montagu 2020). This resulted in changes to the sampling and selection methods, notably with up to two adults now selected per household (NatCen Social Research 2024a). This design makes the BSA survey particularly suited to investigating the extent of homogeneity in responses to attitudinal questions among respondents from the same household and its implications for survey sample design.

For the 2021 survey, a random sample of 41,205 unclustered addresses was systematically selected from a Postcode Address File (PAF) after sorting by region, population density, and tenure characteristics. Each sampled address was randomly assigned to receive one of eleven questionnaire versions (NatCen Social Research 2024a, 2024b). The survey also included a

small clustered sample assigned to version 12 of the questionnaire (NatCen Social Research 2024a), but this was not included in our analysis. While some core question modules for the unclustered sample were included in all eleven versions, other modules were included only in versions 1 to 6 or only in versions 7 to 11. However, differences between versions 1 to 6 and between versions 7 to 11 are immaterial to our analysis.

The survey employed a mixed-mode push-to-web design, with an opt-in computer-assisted telephone interview for respondents unable or unwilling to complete the survey online. As selected addresses were contacted by post, it was not possible to randomly select a single household at the small number of addresses that would have contained multiple households; instead, the selected household was the one that first opened the invitation letter (NatCen Social Research 2024a). Sampled addresses were sent an advance letter, followed by an invitation letter asking adults aged 18 or over to participate; where more than one eligible adult was present, up to two individuals could participate (NatCen Social Research 2024b). The invitation letter included two sets of unique login details for each address. Non-responding households received two reminder letters, which also contained survey access details. A £10 incentive was provided to respondents upon completion. Fieldwork was conducted between September and October 2021. Adjusting for an estimated proportion of ineligible addresses, based on prior waves, the survey achieved an estimated response rate of approximately 14% (NatCen Social Research 2024a, 2024b).

4.2 Variables

The study was primarily concerned with attitudinal variables, although demographic and socioeconomic characteristics were also included. In total, the analyses included 78 variables. Due to the modular design of the BSA survey, not all respondents were administered the same set of attitudinal items. To maximise the available sample size for each analysis, only questions administered in at least five versions of the questionnaire (i.e., versions 1 – 11, 1 – 6 or 7 – 11) were selected. Demographic and socioeconomic variables included sex, age group, ethnic group, occupational classification and educational attainment. Attitudinal variables included political identification and interest, the left to right scale, the libertarian to authoritarian scale and the welfarism scale (administered in versions 1 – 11), modules on housing and planning, family and work life, disability and living standards (versions 1 – 6), and British identity, the European Union referendum, migration and equal opportunities (versions 7 – 11).

4.3 Methods

4.3.1 Estimation of within-household correlation

Within-household similarity was quantified using intra-class correlation coefficients. Estimation was restricted to households in which two eligible individuals responded (2,492 cases). In this context, the intra-class correlation can be interpreted simply as the correlation between the two respondents within a household, providing a direct measure of within-household similarity. While base samples varied across variables due to the modular design of the BSA survey, respondents within the same household were administered the same questionnaire version.

Analyses were conducted in *Stata 19* (StataCorp 2025b) using the *loneway* command, which fits one-way analysis-of-variance (ANOVA) models to estimate intra-class correlation coefficients

(ICCs) (StataCorp 2025a). ICCs were estimated for each demographic, socioeconomic and attitudinal variable at the household-level primary sampling unit. For each categorical variable, a set of binary indicator variables was created, ICCs were estimated for each category separately, and the mean ICC across categories was reported. Because the *loneway* command truncates negative ICC estimates to zero in its reported output, which can occur when the between-cluster mean square is smaller than the within-cluster mean square, untruncated ICC values were computed from the reported ANOVA mean squares using the following formula (Shrout and Fleiss 1979):

$$\rho = \frac{MS_B - MS_W}{MS_B + (\bar{b} - 1)\, MS_W} \tag{4}$$

where MS_B and MS_W are the between-cluster and within-cluster ANOVA mean squares, respectively, and $\bar{b}$ is the average cluster size.

4.3.2 Design effects and effective sample size

Design effects were estimated using *Stata 19* (StataCorp 2025b). Following estimation of means or proportions for each demographic, socioeconomic and attitudinal variables using the *svyset* estimation framework, design effects were obtained using the *estat effects* post-estimation command. The overall design effect ($DEFF$) was calculated to assess the impact of the complex sampling design, accounting for sample stratification, design weights, and household-level clustering on sampling variance relative to simple random sampling.

To examine the contribution of the individual components of the overall design effects, design effects were estimated under alternative survey design specifications by specifying (i) design weights and stratification only ($DEFF_{ps}$), and (ii) household-level clustering and stratification only ($DEFF_{cs}$). The design effect attributable to household-level clustering ($DEFF_c$) was estimated for each survey variable as:

$$DEFF_c = \frac{DEFF}{DEFF_{ps}} \tag{5}$$

Similarly, the design effect attributable to unequal probabilities of selection ($DEFF_p$) was estimated as:

$$DEFF_p = \frac{DEFF}{DEFF_{cs}} \tag{6}$$

This decomposition enabled the relative contribution of household-level clustering and unequal probabilities of selection to the overall design effect to be examined. As with intra-class correlation coefficients, design effects for categorical variables were analysed for each category, with the mean design effect reported for each variable.

Design effects were estimated separately for two samples. First, they were estimated for the full sample using the realised complex sample design based on the any-two-respondent selection method (5,744 cases) to evaluate the effect for the survey as a whole. For this estimation, to compute design weights, conditional on selecting an address, the probability of selecting any

two respondents for $n_i \geq 2$ is approximated by $2/n_i$, where $n_i$is the number of eligible adults (aged 18 or over) at address i; for $n_i = 1$, the probability is 1. The design weight is the inverse of this selection probability (Heeringa et al. 2017; Lynn 2022). The variable measuring the number of eligible adults contained 33 missing values (0.6% of 5,744 cases), for which the modal value was imputed (Lynn 2022). Second, design effects were estimated for a one-respondent sample (4,370 cases) made up of single-respondent households and those obtained when one respondent was randomly retained from households with two respondents. Design weights were recalculated and conditional on selecting an address, the probability of selecting one respondent was $1/n_i$, giving a design weight of n_i.

Design effects were translated into effective sample size (N_{eff}) using:

$$N_{eff} = \frac{n}{DEFF} \tag{7}$$

The associated loss in efficiency was calculated as:

$$1 - \frac{1}{DEFF} \tag{8}$$

5 Results

5.1 Within-household correlation in survey responses

Table 1 reports descriptive statistics for the intra-class correlation coefficients for the sample of two-respondent households across the 78 variables analysed. The values ranged from approximately – 0.77 to 0.70. The mean was 0.27 (*SD* = 0.20), while the median was 0.23 with 50% of values between 0.18 and 0.32.

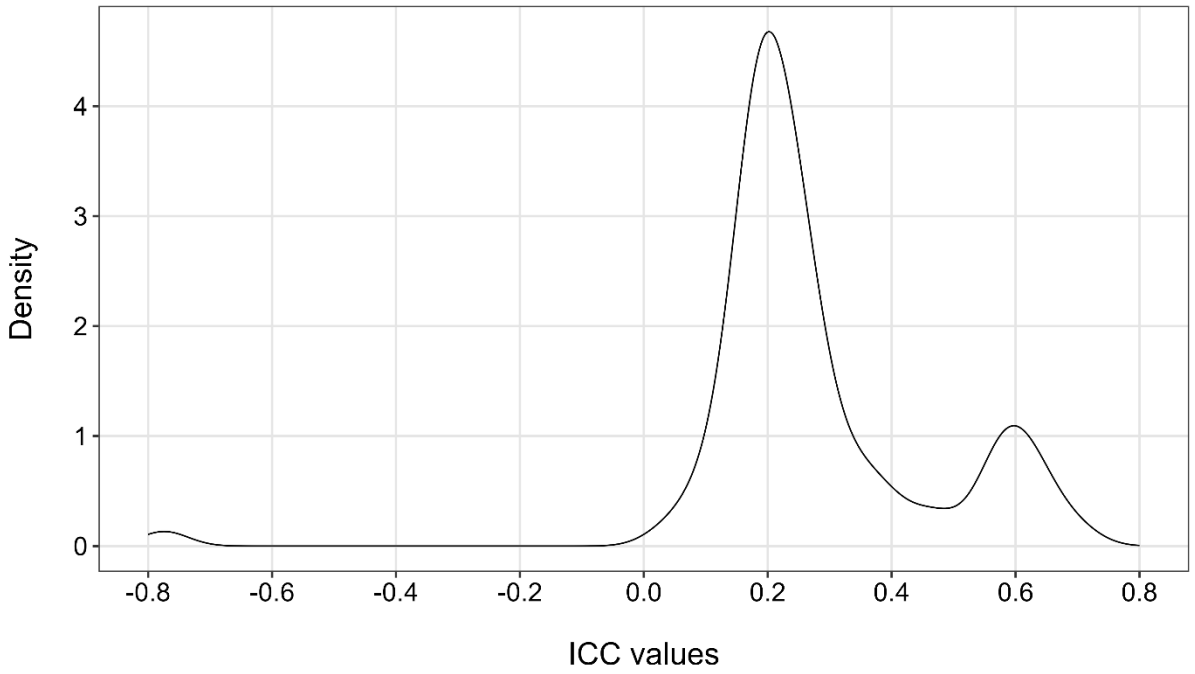
Table 1: Descriptive statistics for intra-class correlation coefficients

Variable	Number of variables	Mean	Std. Dev.	Median	Quartiles		Min.	Max.
					Q1	*Q3*		
ICC values	78	0.271	0.199	0.228	0.178	0.320	– 0.774	0.702

Notes: Based on sample for two-respondent households, *n* = 2,492.

The intra-class correlation coefficients exhibited a bimodal distribution, with a primary mode around 0.2 and a secondary mode around 0.6. The variable ‘sex’ had an extreme negative intra-class correlation coefficient, while all other values were positive (Figure 1).

Figure 1: Distribution of intra-class correlation coefficients



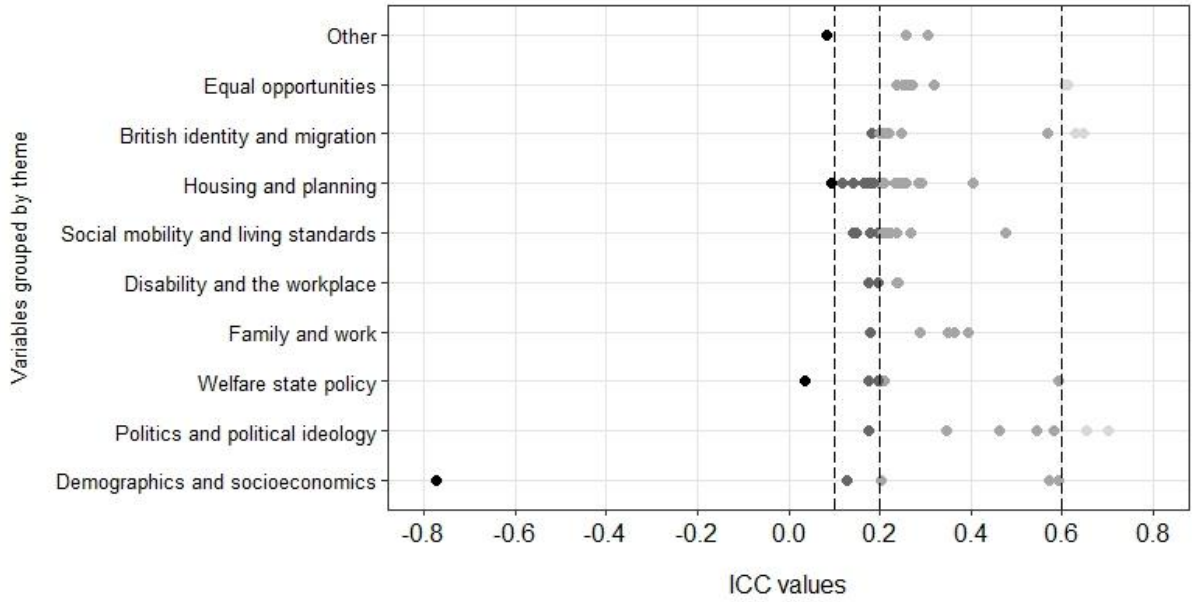
Notes: Distribution of ICC values for 78 variables based on a sample for two-respondent households.

Figure 2 presents estimated intra-class correlation coefficients for the two-respondent households, grouped by theme. Because the estimates were based on the 'any-two' selection method, they may or may not provide an unbiased estimate of the true within-household ICC, depending on the extent to which, if at all, respondent selection in households with more than two eligible adults diverges from random selection. The results indicate substantial heterogeneity in the magnitude of these coefficients across variables and themes.

Four variables exhibited ρ values below 0.10, indicating minimal within-household correlation (Kish et al. 1976). Aside from sex, these variables included three attitudinal measures (views on child maintenance payments when the mother has a new partner, email consultation about new housing developments in the local area, and effective ways to protect personal information). For these variables, responses were largely independent within households, and the households contributed little to the total variance.

The ρ values for a substantial proportion of variables fell within the range 0.10 to 0.20, which are considered relatively high as this level of within-household correlation can substantially inflate the design effects and reduce sampling efficiency (Kish et al. 1976). These variables included attitudinal measures relating to housing and planning, social mobility and living standards, disability and the workplace, and welfare. For these measures, the household constituted a meaningful source of similarity, suggesting that shared environments, resources, and experiences contributed to aligned responses.

Figure 2: Intra-class correlation coefficients by theme



Notes: ICC values for survey variables grouped by theme among households with two respondents and each point represents a variable.

Notably, more than half of the variables exhibited ρ values above 0.20, with some approaching 0.60 or higher. Higher values were most evident among measures of political attitudes, British identity and migration, equal opportunities, and selected socioeconomic perceptions. Such values indicate strong within-household homogeneity, suggesting that a substantial proportion of total variance is attributable to between-household differences rather than within-household variation. Under these conditions, clustering effects are pronounced, leading to substantial design effects and corresponding losses in sampling efficiency.

Overall, the distribution of intraclass correlation coefficients indicates that within-household clustering is a non-negligible feature of the data. For a substantial proportion of variables, sampling clustering within households generates meaningful variance inflation, highlighting the need to account for clustering in both survey design and variance estimation.

5.2 Design effects and relationship with intra-class correlation

Table 2 reports descriptive statistics for the overall and component design effects estimated across the 78 survey variables for the full sample based on the any-two respondent selection method and the simulated one-respondent-per-household sample using the methods described in Section 4.3.2.

Table 2: Descriptive statistics for overall and component design effects

Respondent selection approach	n	$DEFF_p$			$DEFF_c$			$DEFF$			N_{eff}	$Efficiency\ loss$
		Mean (s.d.)	Min.	Max.	Mean (s.d.)	Min.	Max.	Mean (s.d.)	Min.	Max.	$\frac{n}{DEFF}$	$1-\frac{1}{DEFF}$
Any-two	5744	1.077 (0.03)	1.018	1.230	1.131 (0.10)	0.662	1.376	1.208 (0.10)	0.705	1.481	4755	0.172
One-respondent sample	4370	1.188 (0.03)	1.094	1.355	1.000 (0.00)	1.000	1.000	1.182 (0.03)	1.093	1.293	3697	0.154

Notes: Descriptive statistics summarise overall and component design effects estimated for 78 survey variables. Estimates for the any-two approach are based on the realised complex sample design for the full sample, while those for the one-respondent approach are based on the complex sample design for the simulated one-respondent-per-household sample.

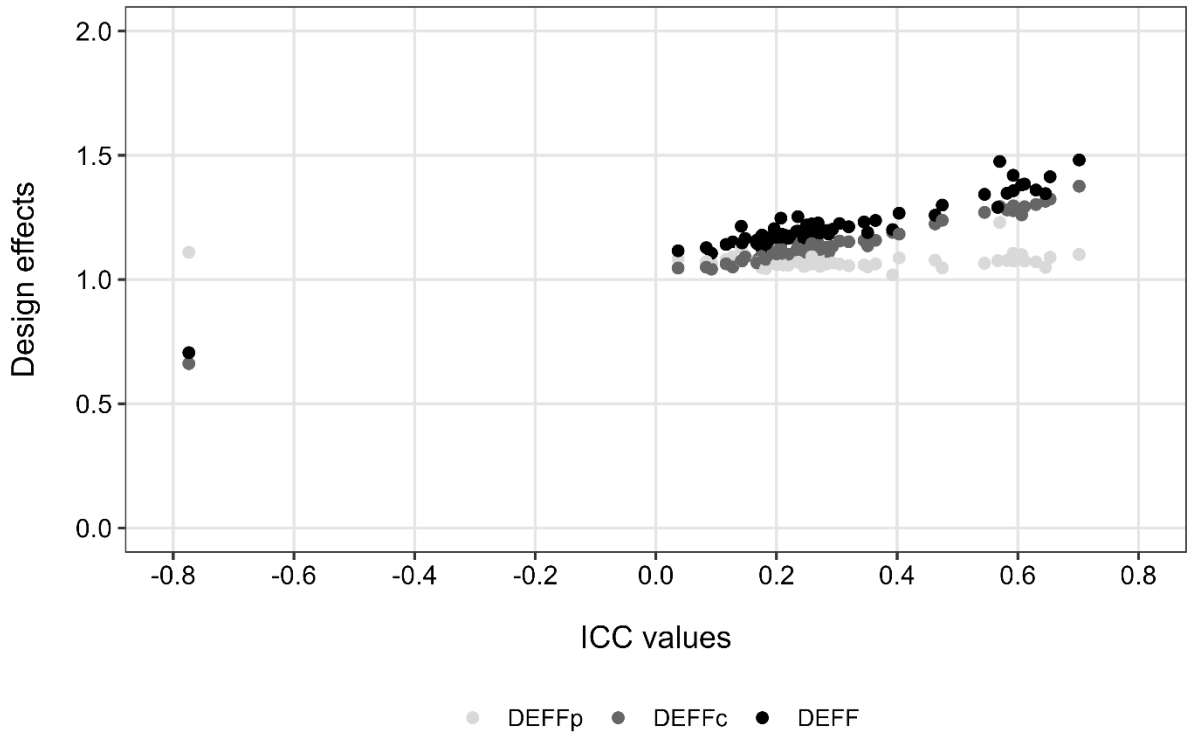
For the full sample based on the any-two respondent selection method, unequal probabilities of selection made a relatively modest contribution to sampling variance. Across the 78 survey variables, the design effect attributable to selection probabilities, $DEFF_p$, averaged 1.08 (*SD* = 0.03), with values ranging from 1.02 to 1.23. These estimates indicate that unequal probabilities of selection increased sampling variances by approximately 8% relative to simple random sampling. The relatively narrow spread of values suggests that this contribution was broadly consistent across survey variables.

In contrast, household-level clustering contributed more to sampling variance and exhibited greater variation across survey variables. Estimates of $DEFF_c$ ranged from 0.66 to 1.38, with a mean of 1.13 (*SD* = 0.10). On average, household-level clustering increased sampling variances by approximately 13%, exceeding the contribution attributable to unequal probabilities of selection. The greater variability in the estimates indicates that the contribution of household-level clustering to sampling variance differed across survey variables.

When all design features were considered simultaneously, the mean overall $DEFF$ was 1.21 (*SD* = 0.10), with values ranging from 0.71 to 1.48. Relative to simple random sampling, the realised complex survey design increased sampling variances by approximately 21%. This corresponds to an average effective sample size of approximately 4,755 respondents from a nominal sample of 5,744, representing an average efficiency loss of approximately 17%. For the variable with the largest overall design effect ($DEFF$ = 1.48), the effective sample size was approximately 3,878 respondents, equivalent to an efficiency loss of approximately 32%.

Figure 3 shows the relationship between the intra-class correlation coefficients reported in Section 5.1 and the corresponding component and overall design effects. The variation in overall design effects ($DEFF$) is driven by variation in intra-class correlation coefficients, whereas the weighting design effect ($DEFF_p$) remained comparatively stable across the observed range of ICC values. Variables exhibiting the highest intra-class correlation coefficients generally produced the largest overall design effects, although no variable exhibited an overall design effect exceeding 1.5. Only one variable (sex), which exhibited the only negative intra-class correlation coefficient (ρ = – 0.77), produced an overall design effect below 1.0.

Figure 3: Relationship between intra-class correlation coefficients and design effects



Notes: Each point represents one of the 78 survey variables. Intra-class correlation coefficients were estimated from households with two respondents, whereas design effects were for the full sample based on the realised complex sample design under the any-two respondent selection design.

Overall, the results indicate that household-level clustering contributed more to the observed design effects than unequal probabilities of selection and that variation in $DEFF$ between variables was driven by variation in ICC. With the any-two respondent selection method, overall design effects were relatively modest.

5.3 Comparison of respondent selection approaches

For the simulated one-respondent-per-household sample, there is no household-level clustering so the mean $DEFF_c$ was 1.00 (*SD* = 0.00) (Table 2). Consequently, the overall design effect was entirely attributable to unequal probabilities of selection. Across the 78 survey variables, $DEFF_p$ averaged 1.19 (*SD* = 0.03), with estimates ranging from 1.09 to 1.36. The corresponding mean overall $DEFF$ was 1.18 (*SD* = 0.03), with estimated values between 1.09 and 1.29. The corresponding average effective sample size was approximately 3,697 respondents from a nominal sample of 4,370, representing an average efficiency loss of approximately 15%.

Table 3 compares the mean design effects and measures of statistical efficiency for the realised complex sample design based on the any-two respondent selection method with those for the simulated one-respondent-per-household sample. Compared to the one-respondent sample, the mean $DEFF_p$ was lower for the any-two respondent selection method (mean 1.08, compared to 1.19), while the mean $DEFF_c$ was higher (mean 1.13, compared to 1.00).

The mean $DEFF$ differed only modestly between the two approaches, averaging 1.21 for the any-two respondent selection method and 1.18 for the one-respondent sample. Correspondingly, relative efficiency increased from 0.83 under the any-two respondent selection method to 0.85 for the one-respondent sample. Overall, despite the contrasting contributions of the component design effects, the any-two respondent selection method and the simulated one-respondent-per-household sample produced similar levels of overall sampling variance and statistical efficiency.

Table 3: Comparison of sampling variance and statistical efficiency for the any-two respondent selection method and the simulated one-respondent-per-household sample

Measure	**Respondent selection approach**		**Difference** (One-respondent – Any-two)
	Any-two	One-respondent sample	
$Mean\ DEFF_p$	1.077	1.188	+ 0.111
$Mean\ DEFF_c$	1.131	1.000	– 0.131
$Mean\ DEFF$	1.208	1.182	– 0.026
$Relative\ efficiency$	0.828	0.846	+ 0.018
$Efficiency\ loss$	0.172	0.154	– 0.018

Notes: Summary statistics based on 78 survey variables. The any-two results are based on the realised complex sample design for the full sample, whereas results for the one-respondent approach are based on the complex sample design for the simulated one-respondent-per-household sample. *Relative efficiency* was calculated as $1/DEFF$ and the difference as one-respondent sample minus any-two.

6 Discussion and conclusion

6.1 Discussion

This study set out to quantify the extent of within-household clustering in survey responses and to evaluate its implications for sampling variance and statistical efficiency in self-completion surveys using address-based sampling frames. By comparing the realised any-two respondent selection method with a simulated one-respondent-per-household sample, the study provides empirical evidence that interviewing multiple respondents per household introduces within-household correlation that is both substantively meaningful and statistically consequential. At the same time, the resulting reduction in overall statistical efficiency is relatively modest, due to a compensating effect of lower variation in selection probabilities when sampling two respondents per household. This highlights the importance of considering both household-level clustering and unequal probabilities of selection when evaluating within-household respondent selection strategies.

The findings first demonstrate that within-household correlation is a common feature of attitudinal survey data, although its magnitude varies considerably across survey variables. Stronger within-household homogeneity was observed for measures relating to political attitudes, national identity, migration, equal opportunities and socioeconomic perceptions,

suggesting that household members frequently share attitudes and experiences in these domains. This pattern is consistent with theoretical expectations that individuals living in the same household are exposed to similar social, economic and informational environments and therefore cannot be regarded as statistically independent. The theoretical background for these findings goes back to the social identity theory (Tajfel and Turner 1979), a long-established framework in social psychology, which explains how people derive their sense of self and identity from the social groups to which they belong, including a household; this sense of belonging to a group influences their social attitudes as well as behaviours. More broadly, the findings from our study are also consistent with the survey methodology literature, which recognises within-household correlation as an inherent characteristic of multi-respondent household surveys (Clark and Steel 2007; Purdon and Wood 2022). This highlights the importance of explicitly considering within-household correlation when evaluating alternative respondent selection strategies.

Secondly, the study demonstrated how household-level clustering contributes to sampling variance under the realised complex sample design based on the any-two respondent selection method. Household-level clustering contributed more to the overall design effect than unequal probabilities of selection. Nevertheless, the overall impact on sampling variance remained moderate, suggesting that selecting two respondents per household does not substantially compromise statistical efficiency.

Thirdly, the comparison of the any-two respondent selection method with the simulated one-respondent-per-household sample highlighted the trade-off between household-level clustering and unequal probabilities of selection in determining sampling variance and statistical efficiency. Selecting two respondents per household introduces a design effect due to intra-household correlation, whereas selecting only one respondent per household increases the contribution of unequal probabilities of selection to sampling variance, as reflected in the weighting design effect. The findings demonstrated that these opposing influences largely offset one another, resulting in only modest differences in overall sampling variance and statistical efficiency between the two respondent selection approaches.

The findings have important implications for the design of self-completion surveys using address-based sampling frames. The comparison of the any-two respondent selection method and the simulated one-respondent-per-household design demonstrated that reducing household-level clustering will simultaneously change the contribution of unequal probabilities of selection to sampling variance. In practice, reducing one source of variance inflation may increase another, meaning that apparent gains in statistical efficiency may be smaller than anticipated when the complete sample design is taken into account. The statistical implications of alternative respondent selection strategies should be assessed by considering the overall performance of the realised complex sample design rather than the contribution of individual design features in isolation.

From a practical perspective, the similarity in overall sampling variance and statistical efficiency between the any-two respondent selection method and the simulated one-respondent-per-household design suggests that the choice between selecting one or two respondents per household should not be based solely on considerations of statistical efficiency. Furthermore,

ICCs were found to be quite high for the political and attitudinal variables studied here. They may be lower for other survey topics, a consideration which could tip the balance of benefits in favour of designs that include more than one respondent per household. Additionally, selecting multiple respondents per household is likely to reduce unit costs of data collection for self-completion surveys (as a single invitation mailing can result in two completed questionnaires) and may also reduce self-selection bias, for which there is greater scope when only one person per household can participate. All these factors should be considered holistically when choosing between within-household respondent selection approaches.

6.2 Limitations and further research

Several limitations of this study should be acknowledged. First, the comparison between within-household selection methods relied on simulated one-respondent-per-household sample derived from the realised any-two respondent selection method rather than an independently implemented random-one respondent selection method. Although this approach provides a practical basis for evaluating the statistical consequences of alternative respondent selection strategies, it cannot fully reproduce all aspects of a survey designed to select a single respondent per household. Secondly, the analyses were restricted to survey variables collected within a single self-completion household survey, most of which were attitudinal measures. The extent of within-household clustering and its contribution to sampling variance may differ for behavioural measures, factual household characteristics, or surveys conducted using alternative modes of data collection.

Further research could extend this study in several ways. First, the effects of intra-household correlation could be examined in surveys selecting larger numbers of respondents per household to evaluate how increasing cluster size influences design effects and sampling variance. Second, effects for populations with different distributions of household size could usefully be estimated in order to broaden the generalisability of the findings of the current study. Third, comparisons could be undertaken using surveys specifically designed to implement a single-respondent selection method rather than relying on a simulated sample derived from a multi-respondent design. Finally, analyses could be extended beyond attitudinal measures to examine household-level clustering and its contribution to sampling variance for behavioural measures and factual household characteristics.

6.3 Conclusion

Overall, this study demonstrates that within-household clustering is a non-negligible feature of attitudinal household surveys and contributes to increased sampling variance when multiple respondents are selected within households. At the same time, the comparison of the any-two respondent selection method and the simulated one-respondent-per-household sample shows that the statistical consequences of alternative respondent selection strategies depend on the combined effects of household-level clustering and unequal probabilities of selection. Although retaining a single respondent eliminates the contribution of household-level clustering to variance inflation, the resulting gains in overall statistical efficiency are modest once changes in the weighting design effect are also taken into account. By examining within-household clustering through intra-class correlation, decomposing overall design effects into their main components, and comparing alternative respondent selection strategies, this study provides a

more comprehensive evaluation of the statistical implications of selecting one or two respondents per household and offers an empirical basis for informing decisions about within-household respondent selection in self-completion household surveys.

Acknowledgements and funding: This research was supported by the UKRI-ESRC strategic research grant ES/X014150/1 for 'Survey data collection methods collaboration: securing the future of social surveys', also known as *Survey Futures*, principal investigator Professor Peter Lynn (University of Essex).

Conflict of interest: None to declare.

Data availability: The data are available from the UK Data Service: SN – 9072, DOI: http://doi.org/10.5255/UKDA-SN-9072-1. Access requires users to register and agree to the relevant terms and conditions of access and use.

7 References

Battaglia, Michael P., Michael W. Link, Martin R. Frankel, Larry Osborn, and Ali H. Mokdad. 2008. 'An Evaluation of Respondent Selection Methods for Household Mail Surveys'. *Public Opinion Quarterly* 72(3):459–69.

Charman, Chris, Sierra Mesplie-Cowan, and Debbie Collins. 2025. *The Post-Pandemic Role of Face-to-Face Fieldworkers: Second Report*. Survey Futures Report No. 6. Colchester, UK: University of Essex. https://surveyfutures.net/reports/.

Clark, Robert G., and David G. Steel. 2007. 'Sampling Within Households in Household Surveys'. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 170(1):63–82.

Clery, E., J. Curtice, S. Frankenburg, H. Morgan, and S. Reid. 2021. *British Social Attitudes: The 38th Report*. *Technical Report*. London: The National Centre for Social Research.

Curtice, J., N. Hudson, and I. Montagu. 2020. *British Social Attitudes: The 37th Report*. *Technical Report*. London: The National Centre for Social Research.

Díaz de Rada, Vidal. 2021. 'Is the Kish Household Sampling Method Better than the Birthday Method?' *Italian Sociological Review* 11(2):485–508.

Gaziano, Cecilie. 2005. 'Comparative Analysis of Within-Household Respondent Selection Techniques'. *Public Opinion Quarterly* 69(1):124–57.

Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. 2nd Edition. Hoboken, NJ: Wiley.

Heeringa, Steven G., Brady T. West, and Patricia A. Berglund. 2017. *Applied Survey Data Analysis*. Second Edition. Boca Raton, FL: CRC Press.

Jabkowski, Piotr. 2017. 'A Meta-Analysis of Within-Household Selection Impact on Survey Outcome Rates, Demographic Representation and Sample Quality in the European Social Survey'. *Ask: Research and Methods* 26(1):31–60.

Jabkowski, Piotr, Piotr Cichocki, and Marta Kołczyńska. 2024. 'Kish Grid vs Birthday Procedures: Survey Refusal Rates by the Type of Within-Household Selection Procedure in the European Social Survey'. *Bulletin de Méthodologie Sociologique* 163:77–87. doi: 10.1177/07591063241258087.

Kish, L., R. M. Groves, and K. P. Krotki. 1976. *Sampling Errors for Fertility Surveys*. *Occasional Papers*. No. 17. World Fertility Survey. The Hague, Netherlands: International Statistical Institute.

Kish, Leslie. 1965. *Survey Sampling*. New York, NY: John Wiley & Sons.

Koch, Achim. 2019. 'Within-Household Selection of Respondents'. Pp. 93–111 in *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, edited by T. P. Johnson, B.-E. Pennell, I. A. L. Stoop, and B. Dorer. Hoboken, NJ: John Wiley & Sons, Inc.

Lynn, Peter. 2022. *Technical Note on Weighting for the ESS Self-Completion Trial in GB*. *Technical Note*. London: European Social Survey ERIC, City, University of London.

Lynn, Peter, and Denise Lievesley. 1991. *Drawing General Population Samples in Great Britain*. *Report*. London: Social and Community Planning Research.

Marlar, Jennifer, Manas Chattopadhyay, Jeff Jones, Stephanie Marken, and Frauke Kreuter. 2018. 'Within-Household Selection and Dual-Frame Telephone Surveys: A Comparative Experiment of Eleven Different Selection Methods'. *Survey Practice* 11(2):1–15.

Maslovskaya, Olga, Lisa Calderwood, Gerry Nicolaas, and Laura Wilson. 2022. *GenPopWeb2: Transitioning from Interviewer-Administered Surveys to Online Data Collection: Experiences, Challenges and Opportunities*. *Final Report*. Southampton: University of Southampton.

Maslovskaya, Olga, Peter Lynn, Lisa Calderwood, Gabriele B. Durrant, Rory Fitzgerald, Gerry Nicolaas, and Joel Williams. 2025. *Survey Futures Position Statement on Response Rates: Supporting Material*. https://surveyfutures.net/wp-content/uploads/2025/06/Response-Rates-Postition-Statement_SupportingMaterial.pdf.

NatCen Social Research. 2024a. *British Social Attitudes 39 - Technical Details*. *Technical Report*. London: National Centre for Social Research.

NatCen Social Research. 2024b. *British Social Attitudes 2021 - User Guide*. *User Guide v4*. London: National Centre for Social Research.

NatCen Social Research. 2024c. 'British Social Attitudes Survey, 2021' [Data collection]. UK Data Service: SN - 9072. http://doi.org/10.5255/UKDA-SN-9072-1.

Ndebele, Nhlanhla, Peter Lynn, Rory Fitzgerald, and Ruxandra Comanaru. forthcoming. *Within-Household Selection Methods for Self-Completion Surveys in the UK: Evidence Review*. Survey Futures Report No. XX. Colchester, UK: University of Essex. https://surveyfutures.net/reports/.

Olson, Kristen, and Jolene D. Smyth. 2017. 'Within-Household Selection in Mail Surveys: Explicit Questions Are Better than Cover Letter Instructions'. *Public Opinion Quarterly* 81(3):688–713.

Olson, Kristen, Jolene D. Smyth, Rachel Horwitz, Scott Keeter, Virginia Lesser, Stephanie Marken, Nancy A. Mathiowetz, Jaki S. McCarthy, Eileen O'Brien, Jean D. Opsomer, Darby Steiger, David Sterrett, Jennifer Su, Z. Tuba Suzer-Gurtekin, Chintan Turakhia, and James Wagner. 2019. *Report of the AAPOR Task Force on Transitions from Telephone Surveys to Self-Administered and Mixed-Mode Surveys*. *Task Force Report*. Alexandria, VA: AAPOR. https://aapor.org/wp-content/uploads/2022/11/Report-of-the-Task-Force-on-Transitions-from-Telephone-Surveys-FULL-REPORT-FINAL.pdf.

Olson, Kristen, Mathew Stange, and Jolene Smyth. 2014. 'Assessing Within-Household Selection Methods in Household Mail Surveys'. *Public Opinion Quarterly* 78(3):656–78.

Purdon, Susan, and Martin Wood. 2022. *The Effects of Selecting Multiple Respondents per Household for a Survey of People in Paid Work: A Statistical and Cost Assessment*. *Report*. Cardiff: WISERD, Cardiff University.

Shrout, Patrick E., and Joseph L. Fleiss. 1979. 'Intraclass Correlations: Uses in Assessing Rater Reliability'. *Psychological Bulletin* 86(2):420–28.

Smyth, Jolene D., Kristen Olson, and Mathew Stange. 2019. 'Within-Household Selection Methods: A Critical Review and Experimental Examination'. Pp. 23–45 in *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment*, edited by P. J. Lavrakas, M. W. Traugott, C. Kennedy, A. L. Holbrook, E. D. de Leeuw, and B. T. West. Hoboken, New Jersey: John Wiley & Sons, Inc.

Stange, Mathew, Jolene D. Smyth, and Kristen Olson. 2016. 'Using a Calendar and Explanatory Instructions to Aid Within-Household Selection in Mail Surveys'. *Field Methods* 28(1):64–78. doi:10.1177/1525822X15604825.

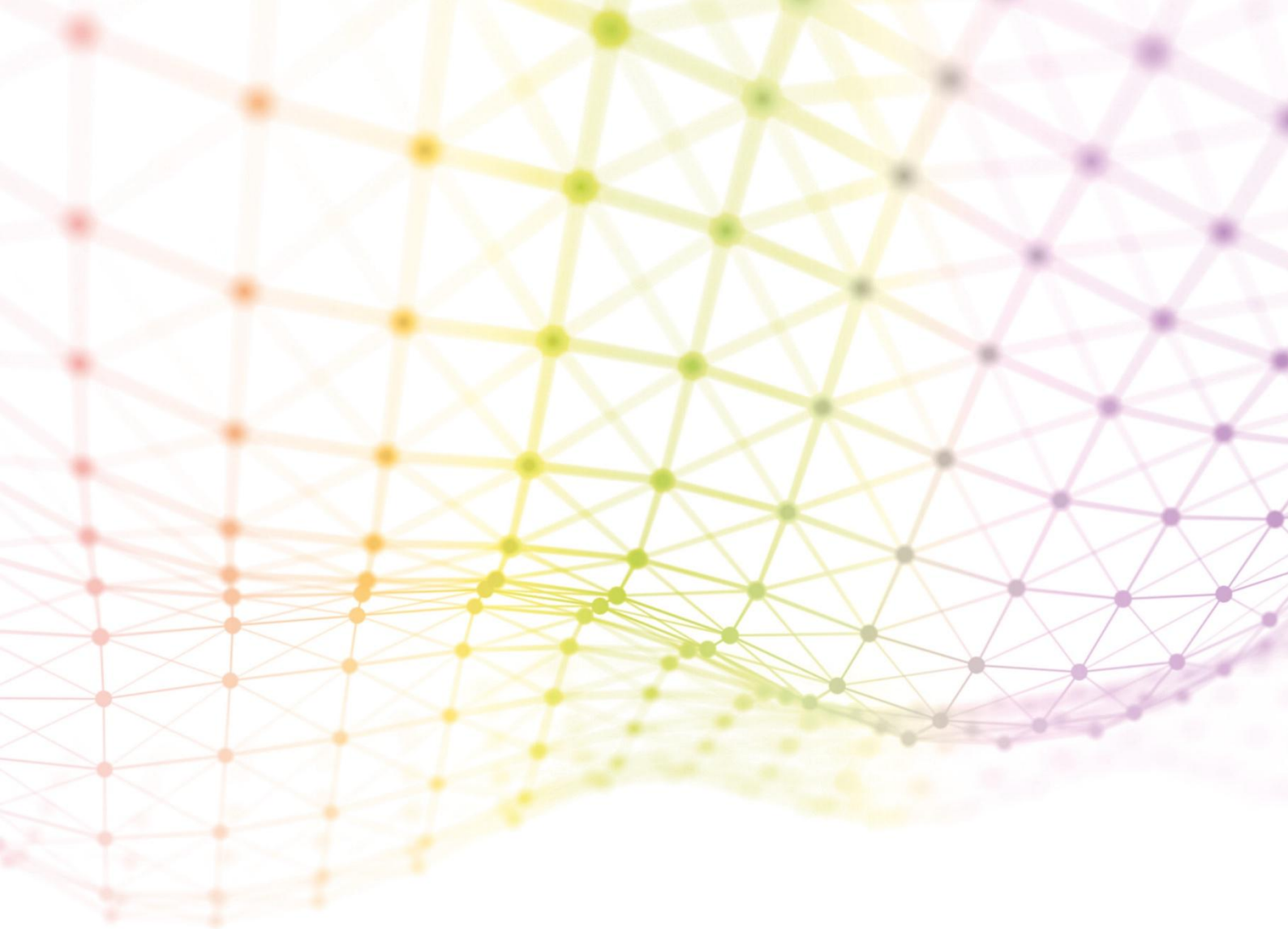
StataCorp. 2025a. *Stata 19 Base Reference Manual*. College Station, TX: Stata Press.

StataCorp. 2025b. 'Stata 19 Statistical Software'.

Tajfel, H., and J. C. Turner. 1979. 'An Integrative Theory of Inter-Group Conflict'. Pp. 33–47 in *The social psychology of inter-group relations*, edited by W. G. Austin and S. Worchel. Monterey, CA: Brooks / Cole.

Williams, Joel. 2015. *Community Life Survey: Investigating the Feasibility of Sampling All Adults in the Household*. *TNS BMRB Report*. London: Cabinet Office. https://assets.publishing.service.gov.uk/media/5a806a11ed915d74e622e50e/The_Community_Life_Survey_Investigating_the_Feasibility_of_Sampling_All_Adults_in_the_Household_FINAL.pdf.

Yan, Ting. 2009. *A Meta-Analysis of Within-Household Respondent Selection Methods*. *AAPOR Report*. Alexandria, VA: AAPOR.



University of Essex

University of Southampton

UKRI Economic and Social Research Council

NCRM National Centre for Research Methods

Office for National Statistics

UCL

Warwick The University of Warwick

Manchester 1824 The University of Manchester

City University of London Est 1894

National Centre for Social Research

LSE The London School of Economics and Political Science

Ulster University

Unil UNIL | Université de Lausanne

Ipsos

verian Formerly Kantar Public

www.surveyfutures.net