Why are some people missing from UK surveys?[1]

Peter Lynn, University of Essex (plynn@essex.ac.uk)

Abstract

This article discusses the issue of population subgroups that may be under-represented in, or missing entirely from, UK population surveys. I do not attempt to quantify the extent of under-representation or to evaluate the impact that this may have on survey estimates. Instead, I attempt to provide a framework for considering the problem: for identifying the groups of interest, for understanding the mechanisms that lead to them being under-represented or excluded, for identifying methods to reduce the under-coverage in future surveys, and for evaluating the merits of implementing such methods on national surveys. The discussion is informed both by a statistical accuracy perspective and an ethical and moral perspective.

Introduction

The UK Government invests heavily in the regular collection of nationally-representative survey data which forms the basis of a range of official statistics as well as providing a research resource. High-profile examples of such surveys include the Labour Force Survey (LFS), Family Resources Survey, National Travel Survey and the Crime Survey for England and Wales. Arguably the highest profile of all is the decennial population Census, which attempts to collect data relating to every person in the country. Aside from Government surveys, national surveys are also carried out by academics, private organisations and third sector organisations, all with the aim of shedding light on various aspects of society.

---

[1] This article is an expanded version of a presentation made at the Radical Statistics conference in London on 24 February 2024 with the title "Who is missing from UK surveys?"

However, to correctly reflect society as a whole, these surveys must correctly represent all sub-sets of society. But surveys – like any other source of statistics – are not perfect tools and do not do this perfectly. Occasionally, concerns about representation become a matter of public concern. Census 1991 in the UK was notable for suffering from a much larger undercount than previous Censuses and for a consequent prolonged media debate about the causes of the undercount and the implications for the utility of the Census (Simpson & Dorling, 1994). Initial reports suggested that around 1 million people were missing from the Census, a figure which made headline news in national newspapers (Waterhouse, 1992), though this was later revised to an estimate of 572,000, or around 1% of the population. But these were not a random 1%: the undercount was greatest amongst young men aged 20-34 and in urban, particularly inner-city, areas. The recent introduction of the Community Charge, a local tax based on people rather than property, was blamed for much of the undercount. While deliberate deregistration in order to avoid the Community Charge may or may not have explained a large proportion of the undercount, it served to highlight the idea that undercount, or non-response, should not be assumed either to be negligible or to be stable over time, and to bring the debate about the importance of countering this into the public realm, at least temporarily.

Other social surveys have not been immune from the impacts of nonresponse. In October 2023, the Office for National Statistics (ONS) was criticised for failing to deal adequately with plummeting response rates to the Labour Force Survey (Cominetti, 2023; Francis-Devine, 2023). This criticism had come about because ONS's Labour Market Statistics, based on the LFS, no longer rang true (Economist, 2023). As a result, release of Labour Market Statistics was suspended for four months and replaced by "experimental" headline statistics (Office for National Statistics, 2023; Office for Statistics Regulation, 2023) which drew upon alternative data sources. This caused considerable difficulties for ONS's data users. The survey's response rate had fallen from over 80% in the 1980s to a little under 50% in 2013 and to less than 15% in mid-2023.

Ultimately, all social surveys suffer from some degree of nonresponse. This may or may not affect estimates based upon the data in meaningful ways, depending on the extent and nature of the nonresponse (Groves and Peytcheva, 2008; Lynn, 2008). Most users of the data, or of statistics or conclusions based upon the data, pay little or no attention to this until the nature of nonresponse suddenly changes in a way that has a noticeable impact on estimates, as in the Census and Labour Force Survey cases mentioned above. Yet the effects of nonresponse could be ever-present, affecting statistical estimates that have not received any critical interrogation. And there are reasons other than non-response for under-representation of sub-groups. These include deliberate exclusions from surveys, and deficiencies in the sampling frames used to select survey samples. The problems caused by under-representation of certain subgroups in surveys, particularly in the context of lower- and middle-income countries, is discussed by Chávez Villegas and Samman (2015) and Carr-Hill (2019).

Reasons for Exclusion from Surveys

Population subgroups can be excluded from surveys at any of three stages in the survey implementation process. These are shown in Figure 1. These stages correspond to the design specification, sample design and data collection stages of the survey process and the potential errors arising from these exclusions are referred to within the total survey error framework (Groves & Lyberg, 2010) as design specification errors, coverage errors and nonresponse errors respectively. The Total Survey Error (TSE) framework illustrates how these sources of subpopulation exclusion can contribute directly to the (in)accuracy of survey-based estimates (Figure 2).

Figure 1: Reasons for exclusion from surveys
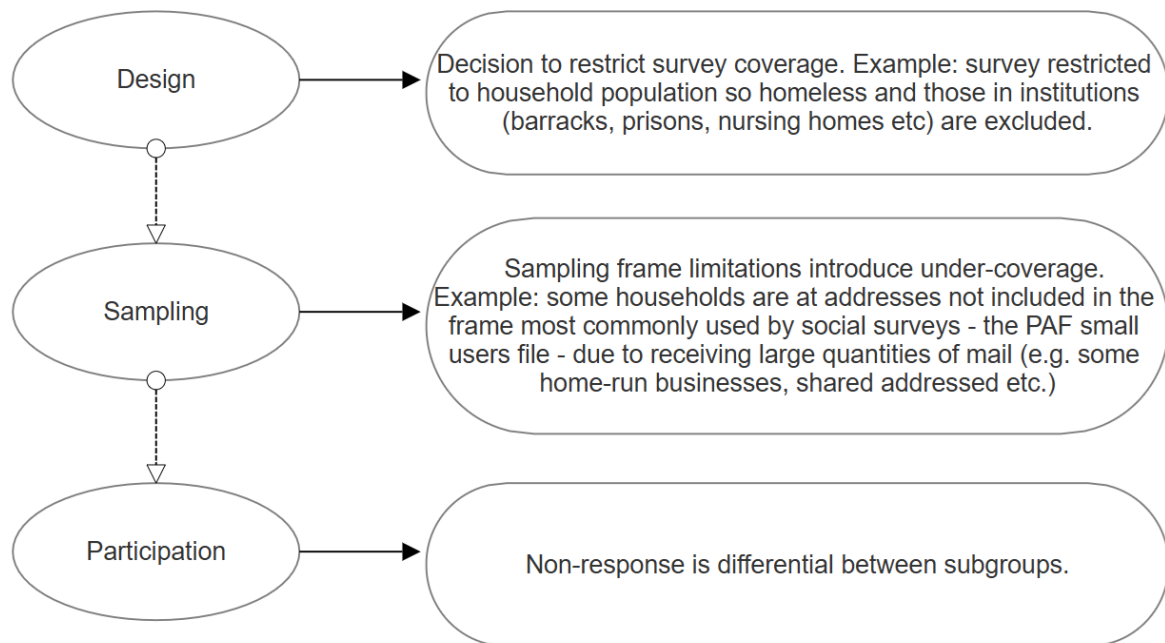
Figure 2: Total Survey Error Components with sources of exclusion highlighted
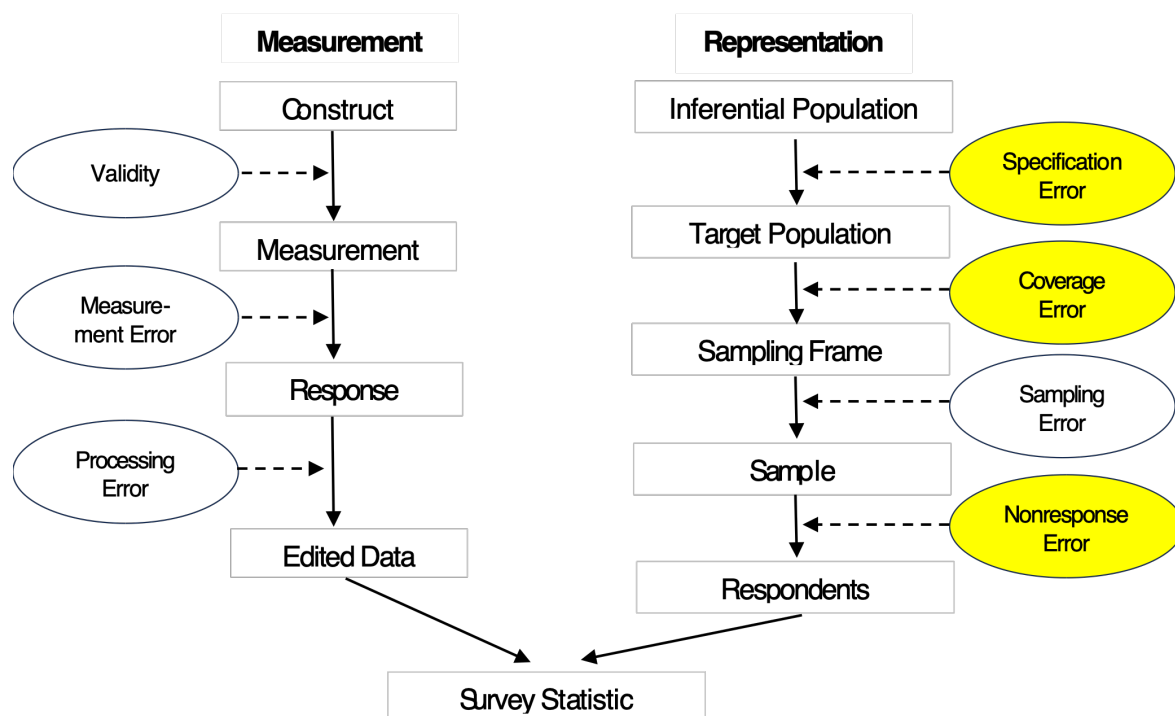
Note: Adapted from Groves & Lyberg (2010)

Design specification error

Exclusion due to a design specification error occurs when the target population for the survey does not include all of the population to which users may make inferences (the inferential population). Survey taking organisations generally get around this by stating that inference should only be made to the target population, so one might question whether it is the target population or the inferential population that is being mis-specified. Either way, data users and users of research based on the data typically have no means of making allowance for the mis-match, so it is inevitable that policy decisions will be informed by research based on data that excludes a subset – typically a non-random subset – of the population affected by the policy.

For example, the Department for Education's School and College Voice Survey is restricted to maintained state schools: private, independent and other non-maintained schools are excluded. The stated aim of the survey is to allow the department "to make more effective policy." The findings are reported with phrases such as "a majority of secondary school teachers thought..." (Department for Education, 2024), which might suggest to the reader that the survey represents all schools. Similarly, most household surveys, in reporting their results, simply refer to "people", implying that they represent all people. For example, the report of the 2020 National Travel Survey (Department for Transport, 2021) states "People made 739 trips on average in 2020..." In reality, people not resident in private households (e.g. those in barracks, prisons, nursing homes and other institutions) are excluded. And yet any transport policy informed by the findings will certainly affect all people, not just those in private households.

Coverage error

Coverage error arises when the sampling frame or sampling method excludes some members of the target population. This can happen as a result of the frame being out-of-date or because some members of the target population do not meet the criteria for inclusion on the frame. For example, if the electoral registers are used as a sampling

frame of residential addresses, it should be realised that addresses at which all residents are either ineligible to vote or have not registered to vote will be excluded. This will lead to systematic under-representation of non-UK nationals (who are less likely than others to register, even when eligible) as well as those aged 18 to 34, ethnic minorities and lower socio-economic groups (groups who are the least likely to be registered according to Electoral Commission (2023)). When using the Postcode Address File (small users) as a frame of residential addresses, as most UK household surveys do, coverage is good, but there are still some households residing at addresses not on the frame, notably those whose address is on the large user file due to receiving a large volume of mail. This could be because a home-based business generates a lot of mail, or because the address is shared with a business or institution.

Non-response error

Non-response error arises when sampled units (households or persons) do not participate in the survey. If the non-participants are systematically different from others, then under-representation of groups with characteristics shared with the non-participants will ensue. There are many reasons for non-participation and each will tend to exclude different groups, but it can be helpful to consider three main categories of reasons for non-participation: the survey may fail to contact the sample member (non-contact), the sample member may be unable to participate, and the sample member may be unwilling to participate.

Non-contact

A failure to contact a sample member means that the sample member does not become aware of the survey. In the case of a face-to-face or telephone survey, this happens when the interviewer does not manage to speak to the sample member or a member of their household. In the case of a self-completion survey, this happens when the sample member does not receive, or see, the invitation to participate. Couper and Ofstedal (2009) sub-divide the reasons for non-contact into two distinct steps in the survey process – location and contact – and it can be helpful to consider these two steps

separately as the causes of exclusion and the implications for under-representation can be quite distinct in the two cases.

On face-to-face interview surveys, successful location involves the interviewer reaching the correct front door, while successful contact involves someone opening the door and talking to the interviewer. It will be obvious that the factors associated with success at each of these steps are rather different (Groves and Couper, 1998). Successful location requires the address details to have been correct on the sampling frame and sufficiently detailed to enable location on the ground. When the frame is an administrative list of some sort, such as a list of benefit recipients or library users, anyone who has changed address since they last had an interaction with the administration system is likely to have the wrong address on the system. In that situation, an interviewer will sometimes, but not often, be able to track the sample member to their new address. Successful contact, conditional on location, requires that the interviewer calls at a time when someone is home, and that someone is willing to open the door. This will depend mainly on interviewer persistence and their ability to be flexible in the times and days when they make call attempts. Households in which there is often no-one at home are at greatest risk, e.g. single-person households with long working hours and/or a long commute.

On self-completion surveys it is hard to distinguish empirically between location and contact, but the two steps remain conceptually distinct. Whether the invitation to participate reaches the sample member depends on correct address details (either postal address or email address depending on the means of delivery employed); whether contact is achieved depends on the sample member noticing, and looking at, the invitation. The latter can be influenced by survey design features that aim to make the mailing stand out, look important, or look interesting (Albaum and Smith, 2012).

Unable to participate

The reasons why some people are unable to participate in surveys differ greatly between self-completion and interviewer-administered modes and are generally greater in the former case. Those most at risk of being unable to participate in an interviewer-administered

interview are those who cannot communicate verbally with the interviewer, either because they suffer from muteness or communication disorder of some other kind, or deafness, or because they and the interviewer do not share a common language. Some surveys make special efforts to include such people, but most do not have the resources to do so. There is a wider set of reasons for being at risk of being unable to participate in a self-completion survey. These include being insufficiently literate, having poor or no sight, and having poor motor control, in addition to not understanding the language of the questionnaire. In the case of web surveys, additional barriers include technical skills, internet access and comfort and familiarity with the internet.

Unwilling to participate

There are many reasons why people may be unwilling to participate in surveys. These include concerns about privacy, data usage and legitimacy and less specific unwillingness to give time and information. The latter appears particularly prevalent amongst those not very socially included and the highest income earners. Other reasons may be situational and related to being particularly busy, stressed or upset. But unwillingness is not independent of survey design. Sample members are less likely to be willing to participate in a survey if it seems irrelevant, difficult, excessively time-consuming or intrusive, or if it is poorly explained, motivated or introduced.

Which groups are under-represented?

Figure 3 summarises the steps in the survey process that bring a risk of exclusion and the barriers that arise at each step (as outlined in the sections above). The third column attempts to illustrate the subgroups consequently at risk of exclusion. It can be seen that a variety of subgroups are at risk. Some groups are likely rather small (those suffering from muteness) and others form a much larger proportion of the population (the socially excluded). And some of the characteristics associated with risk of exclusion may also tend to co-exist, leading to enhanced risk (e.g. socially excluded, poor English, not resident in a private household). The net effect on the representativeness of any particular survey sample will depend on a

range of factors including, crucially, the survey protocols adopted, but also the nature of the sampling frame, the mode(s) of data collection, and the content and complexity of the survey.

Figure 3: Risk of exclusion: survey steps, barriers, and subgroups affected

Step in survey process	Barriers to inclusion	Subgroups at risk
Specification (population surveys)	Not resident in a private household	Resident in an institution; homeless
Specification (general)	Survey-specific	
Coverage (population surveys)	Resident at an address not on PAF	Sharing an address with a business
Coverage (general)	Survey-specific	
Location	Incomplete or incorrect address / incorrect email address	Remote locations (PAF); Recent movers (named sample frames) Without a stable email address (email invites)
Contact (in-home survey)	Interviewer fails to visit at appropriate times/days	Rarely at home
Contact (self-completion survey)	Invitation not noticed / opened	
Unable to participate (interviewer-administered survey)	Communication disorder; Language	Communication disorder; Deafness Poor English
Unable to participate (self-completion survey)	Language; Illiteracy; Visual disability; Motor impairment	Poor English; Poor literacy; Visual disability; Motor impairment
Unwilling to participate	Concerns about privacy / legitimacy; Lack of motivation; Situational: busy, stressed etc	Socially excluded; Busy lifestyles; Highest income earners

What should be done?

A large literature addresses methods for persuading unwilling sample members to participate in surveys. Effective techniques include persuasive and motivational initial communications (Dillman et al, 2014; Wenemark et al, 2011), emphasising salient aspects of the survey (Groves et al, 2000), use of reminders with varied messaging (Muñoz-Leiva et al, 2009), appropriate use of respondent incentives (Singer & Ye 2013; Toepoel 2012), and targeting each of those to sample subgroups where possible (Lynn, 2014, 2017). Additionally, for interviewer-administered surveys the ability of an interviewer to tailor their approach to the circumstances and to recognise when it is best to back-off and re-engage later, rather than prompt a refusal, has been shown to influence co-operation rates (Morton-Williams, 1992; Groves & Couper, 1998). An important distinction should be recognised between surveys that are able to target messaging to particular subgroups and those that cannot. The former include surveys which are a follow-up to a previous data collection exercise, including longitudinal surveys, and those drawing upon an informative sampling frame. The latter include the majority of general population cross-sectional surveys in the UK as these use the PAF as a sampling frame, severely limiting the possibilities for a survey to address particular concerns of subgroups.

Regarding risk of exclusion at any of the other steps in the survey process (design, sampling, contact, ability), for many of the risks identified above, some obvious remedies are available. These are generally resource-intensive, however, and would require justification. For example, survey materials can be translated into multiple languages and interviewers can be recruited and trained to administer interviews in those languages, but this is costly to do and is logistically particularly difficult for a face-to-face survey, as it is hard to predict where in the country an interviewer with particular language skills may be needed. For a web survey, it is less challenging to offer a choice of languages to all participants, but it is still necessary to communicate to the sample member at the invitation stage that it is worth them accessing the survey as they will find a questionnaire in their language. Understanding Society,

the UK Household Longitudinal Study, provides for interviewing in nine languages other than English (Understanding Society, undated), though this is justified largely because the study includes a sizeable ethnic minority boost sample. The NHS England GP Patient Survey provides an online questionnaire in 14 languages, justified by the large sample size: of 759,149 respondents in 2023, 6,199 took part in a language other than English.

As some of the barriers to participation are particular to a survey mode, offering alternative mode options for participation can help to lower the barriers. Mixed-mode approaches are therefore generally favoured when inclusion is of particular importance (Lyberg et al, 2022). Arguably, the barriers to being able to participate are particularly important as they speak to moral and ethical issues around inclusion; the idea that everyone has an equal right to participate in society and to have their voice heard. Individuals have a right to choose whether or not to participate in a voluntary survey, no matter how much the survey organisation would like to persuade them to do so, but they are unable to make that choice if there are insurmountable barriers to participation. It can therefore be argued that it is the duty of the survey organisation to make surveys accessible to all – to the extent possible – over and above any arguments about inclusion being a necessary prerequisite of representativeness, and hence statistical accuracy.

For organisations who regularly carry out multiple surveys, organisation-wide initiatives to reduce exclusion from surveys can be effective and can defray the resource costs. In 2020 the National Statistician set up the Inclusive Data Taskforce with a remit to ensure "that data and evidence across the UK is reflective and inclusive of all" (ONS, 2021). Within this broad remit, one strand of work looked at approaches to data collection. Two of the report's 46 recommendations are closely related to inclusion in surveys:

> 1.4 Data producers should undertake appropriate research to identify the practical barriers to participation and implement best practice in data collection, including ethical considerations, to enhance the inclusiveness of the approaches taken. This might

entail providing internet access to address the barriers for digitally excluded groups, and translators for those not fluent in English.

1.5 Data producers should ensure that data collection instruments are accessible to all, recognising differences in language, literacy, and the relative accessibility of different modes and formats. For example, using multi-mode surveys as standard practice and implementing additional adjustments to enable the participation of adults and children with a range of disabilities, and those who experience other forms of exclusion, including digital exclusion.

An implementation plan (ONS, 2022) set out how ONS would respond to the recommendations and included promises to "carry out research into the value of, and how best to collect survey data from populations not resident in private households, including those living in Communal Establishments, when necessary" and "review operational processes involved in collecting data in our current portfolio of social surveys to determine the barriers to taking part, including among those underrepresented in the statistics. This will enable the design of operational processes where barriers are prevented, removed, or mitigated against." ONS later announced their own survey strategy, one of the pillars of which was to be "fully inclusive by design."

Subsequent work led by the survey and behavioural insights hub at ONS has investigated the experiences of particular groups with ONS surveys in different modes. The groups considered were those with visual impairments, deaf or hard of hearing, neurodivergence, and mental health issues. As of October 2024, findings from this work had not yet been published.

Figure 4 sets out some of the measures that can be considered in order to address the risks of exclusion for each of the groups identified in figure 3 as being at risk at the specification, sampling, contact and ability stages of the survey process. Survey managers and survey agencies are invited to consider whether it might be appropriate to adopt some of these measures either for specific surveys or as an agency standard.

Figure 4: Mitigation measures for subgroups at risk of not receiving a survey invitation or not being able to participate

Subgroups at risk	Mitigation measures
Non-household population	Additional sampling frames of institutions; sampling of homeless persons; sampling non-household population through defined connections to households
Sharing an address with a business	Screen an additional sample of (targeted) business addresses
Remote locations (PAF)	Extra tracing efforts
Recent movers (named sample frames)	Extra tracing efforts; ability to send interviewers to new locations (face-to-face)
Without a stable/correct email address (email invites)	Offer alternative modes of contact/participation
Rarely at home	Incentivising and rewarding interviewer persistence (face-to-face)
Communication disorder	Offer self-completion / alternative modes
Deaf or hard of hearing	Offer visual materials / self-completion, or British Sign Language (BSL)
Poor English	Survey materials and interviews in multiple languages; offer an interpreter, e.g. by video call
Poor literacy	Greatly simplify language and grammar; assisted interviews or ACASI (self-completion surveys)
Visual impairments	Offer interviewer-administered option (if otherwise self-completion)
Motor impairment	Allow for completion using either mouse or keyboard (web surveys); offer video interviewing (if otherwise self-completion)

References

Albaum G & Smith S M (2012) Why people agree to participate in surveys, chapter 11, 179-193, in *Handbook of Survey Methodology for the Social Sciences*, ed. L Gideon, SpringerLink, New York. https://doi.org/10.1007/978-1-4614-3876-2_11

Carr-Hill R (2019) Using Survey Data: Towards Valid Estimates of Poverty in the South. In *Data in Society* (ed.s J. Evans, S. Ruane, H. Southall), Policy Press.

Chávez Villegas C & Samman E (2015) Exclusion in household surveys: Causes, impacts and ways forward. ODI report. https://odi.org/en/publications/exclusion-in-household-surveys-causes-impacts-and-ways-forward/

Cominetti N (2023) Flying blind? The case of the missing employment data. Resolution Foundation. https://www.resolutionfoundation.org/comment/flying-blind/

Couper M P and Ofstedal M B (2009) Keeping in contact with mobile sample members, chapter 11 in *Methodology of Longitudinal Surveys*, ed. P Lynn, Wiley, Hoboken NJ.

Department for Education (2024) School and college voice: February 2024. https://www.gov.uk/government/publications/school-and-college-voice-omnibus-surveys-for-2023-to-2024/

Department for Transport (2021) National Travel Survey: 2020. https://www.gov.uk/government/statistics/national-travel-survey-2020/

Dillman D A, Smyth J D, Christian L M (2014) *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4th ed. Hoboken, NJ: Wiley.

Economist, The (2023) Britain's labour-market figures are less reliable than they were, 5 October. https://www.economist.com/britain/2023/10/05/britains-labour-market-figures-are-less-reliable-than-they-were

Electoral Commission (2023) 2023 Report: electoral registers in the UK. https://www.electoralcommission.org.uk/research-reports-and-data/electoral-registration-research/accuracy-and-completeness-electoral-registers/

Francis-Devine B (2023) Has labour market data become less reliable? House of Commons Library. https://commonslibrary.parliament.uk/has-labour-market-data-become-less-reliable/

Groves R M & Couper M P (1998) *Nonresponse in Household Interview Surveys*. New York: Wiley.

Groves R M & Lyberg L (2010) Total survey error: past, present, and future, *Public Opinion Quarterly* 74(5), 849-879.

Groves R M & Peytcheva E (2008) The impact of nonresponse rates on nonresponse bias: a meta-analysis, *Public Opinion Quarterly* 72(2), 167-189.

Groves R M, Singer E & Corning A (2000) Leverage-saliency theory of survey participation: description and an illustration, *Public Opinion Quarterly* 64(3), 299-308.

Lyberg L, Lynn P & Schouten B (2022) Mode issues in comparative surveys, chapter 23, 321-336, in *Improving the Measurement of Poverty and Social Exclusion in Europe: Reducing Nonsampling Errors* (ed.s P Lynn & L Lyberg), EU Publications.

Lynn P (2008) The problem of non-response, chapter 3, 35-55, in *International Handbook of Survey Research Methods* (ed.s E D de Leeuw, J J Hox & D A Dillman), Lawrence Erlbaum Associates.

Lynn P (2014) Targeted response inducement strategies on longitudinal surveys, chapter 27, 322-338, in *Improving Survey Methods: Lessons from Recent Research* (ed.s U Engel, B Jann, P Lynn, A Scherpenzeel & P Sturgis), Routledge/ Psychology Press.

Lynn P (2017) From standardised to targeted survey procedures for tackling non-response and attrition. *Survey Research Methods* 11(1): 93-103. http://dx.doi.org/10.18148/srm/2017.v11i1.6734.

Morton-Williams J (1992) *Interviewer Approaches*. Aldershot: Dartmouth.

Muñoz-Leiva F, Sánchez-Fernández J, Montoro-Ríos F & Ibáñez-Zapata J A (2009) Improving the response rate and quality in web-based surveys through the personalization and frequency of reminder mailings, *Quality and Quantity* 44, 1037-1052.

Office for National Statistics (2021) Inclusive Data Taskforce: Recommendations Report. Leaving no-one behind. How can we be more inclusive in our data? https://uksa.statisticsauthority.gov.uk/publication/inclusive-data-taskforce-recommendations-report-leaving-no-one-behind-how-can-we-be-more-inclusive-in-our-data/

Office for National Statistics (2022) Inclusive Data Taskforce: Implementation plan https://uksa.statisticsauthority.gov.uk/publication/inclusive-data-taskforce-implementation-plan/

Office for National Statistics (2023) *UK Labour Market: October 2023* (release). https://www.ons.gov.uk/releases/uklabourmarketoctober2023

Office for Statistics Regulation (2023) Review of experimental Labour Market Statistics (produced by the Office for National Statistics (ONS)) https://osr.statisticsauthority.gov.uk/news/osr-publishes-its-findings-on-response-rate-challenges-in-onss-labour-market-statistics/

Simpson S & Dorling D (1994) Those missing millions: implications for social statistics of non-response to the 1991 Census, *Journal of Social Policy* 23(4), 543-567.

Singer E & Ye C (2013) The use and effects of incentives in surveys, *Annals of the American Academy of Political and Social Science* 645(1), 112-141. https://doi.org/10.1177/0002716212458082

Toepoel V (2012) Effects of incentives in surveys, chapter 13, 209-223, in *Handbook of Survey Methodology for the Social Sciences*, ed. L Gideon, SpringerLink, New York. https://doi.org/10.1007/978-1-4614-3876-2_13

Understanding Society (undated) Main survey user guide, section "Questionnaires, questionnaire modules and index terms". https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide

Waterhouse, Rosie (1992) 'Missing million' indicates poll tax factor in Census, *The Independent*, 16 October 1992.

Wenemark M, Persson A, Noorlind Brage H & Svensson T (2011) Applying Motivation Theory to Achieve Increased Respondent Satisfaction, Response Rate and Data Quality in a Self-administered Survey, *Journal of Official Statistics* 27(2), 393-414.