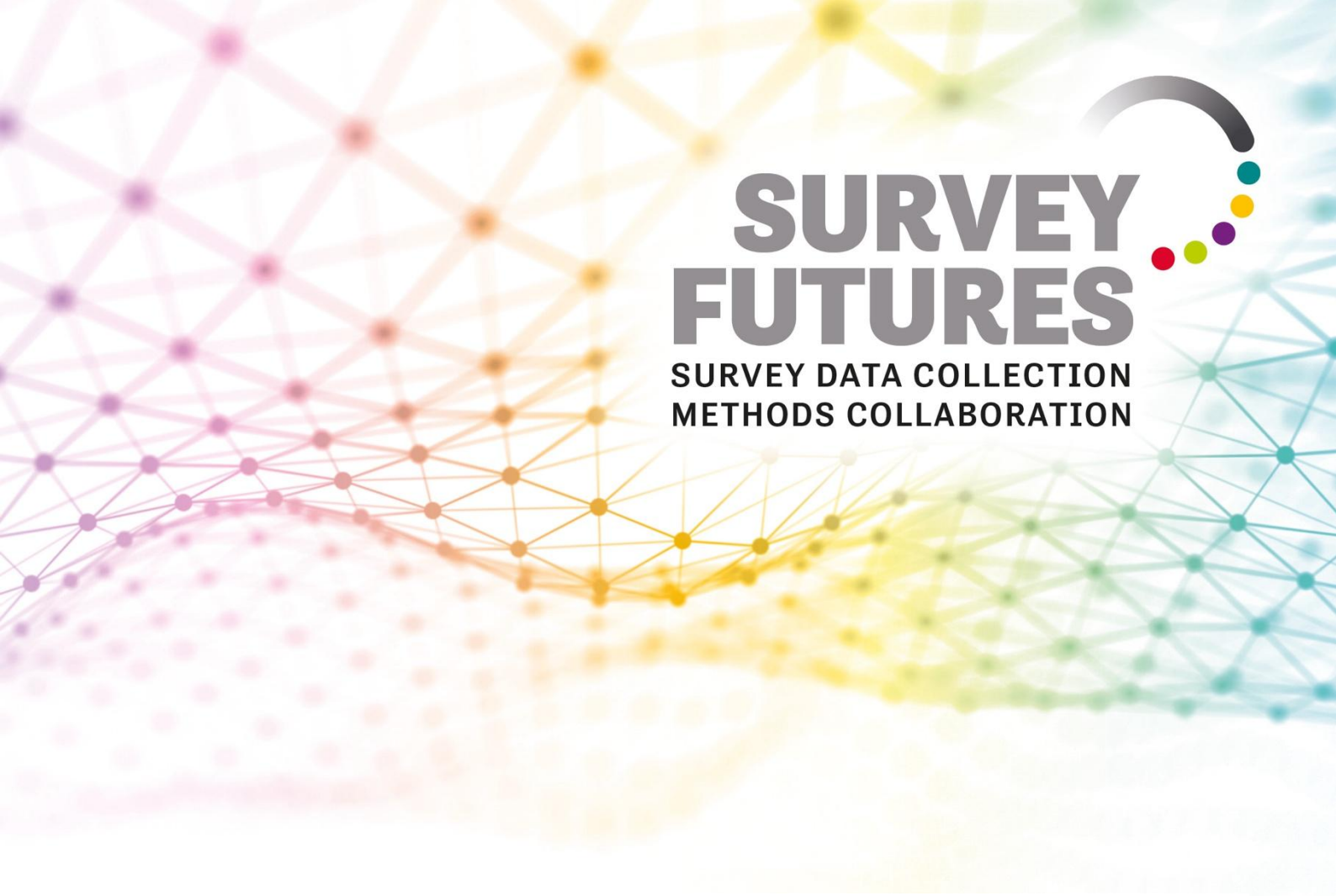


Survey Practice Guide 10: Enhanced Sampling Frames for General Population Surveys

Anna Keyes[1] and Paul A. Smith[2]

[1] National Centre for Social Research; [2] University of Southampton

September 2026

www.surveyfutures.net

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 suggested form of citation for this guide is:

Keyes, A. & Smith, P.A. (*2026*) 'Survey Practice Guide: Enhanced Sampling Frames', *Survey Futures Survey Practice Guide* no. 10. Colchester, UK: University of Essex. Available at https://surveyfutures.net/practice-guides/.

Contents

Executive Summary 1

Introduction 2

Evaluating the quality of UK general population sampling frames 3

Postcode Address File as a sampling frame 5

Addressbase as a sampling frame 7

Administrative data linkage to address sampling frames 9

Administrative data as a general population sampling frame 12

Conclusions 14

Reference list 16

Executive Summary

- This practitioner guide provides an accessible overview of the options for a general population sampling frame in the UK.

- The Postcode Address File (PAF) is the most used sampling frame for general population surveys. It has good coverage and its properties are well known, but it provides only address information and leads to relatively high ineligible rates.

- AddressBase is potentially a step forward but has not been evaluated as a social survey frame, though there is some evidence from the 2021 population censuses on its quality.

- Administrative data could provide additional information to improve sampling, but there are challenges in accessing this data, and in linking it to existing address level information.

- Sampling directly from administrative data sources is possible, but access and ethical permissions can be very challenging to navigate, and there are rarely any sources against which to evaluate the quality of the frame. Administrative data may provide a practical alternative for sampling from specific subpopulations.

Introduction

What this guide covers

This Survey Futures practitioner guide summarises recent published evidence regarding sampling frames that cover UK residents of all ages. The evidence includes relevant academic research, public sector publications, and technical reports. Sampling frames specific to subgroups of the population are not covered and neither are sampling frames that include residents outside the UK. The guide applies to sampling frames for face-to-face, push-to-web, and push-to-telephone modes of survey data collection.

Who this guide is for

This guide is intended for survey statisticians and similar practitioners who are drawing samples for large-scale surveys that need to be representative of the general population within the UK, either adults or of all ages. Its purpose is to summarise the key lessons from available evidence and best practise for UK general population sampling frames, to inform the selection of an appropriate sampling frame.

How to use this guide

The survey practitioner should read this guide to help inform their selection of a sampling frame for a new general population survey within the UK, or when reviewing the existing sampling frame for an established general population survey.

Evaluating the quality of UK general population sampling frames

There is limited twenty-first century evidence from within the UK for sampling frame quality evaluation, due to the dominance of the Postcode Address File (PAF) as a general population sampling frame in recent decades. The characteristics of PAF have changed only in the treatment of addresses with multiple residences since evaluations conducted in the 1980s and 1990s. These include Wilson and Elliot (1987), Lynn and Taylor (1995), and Butcher (1988). Butcher (1988) examines characteristics of PAF including coverage, sampling costs, control of sampling procedures, and choice of clustering unit for clustered samples. Lynn and Taylor (1995) report the results of a split sample experiment in the 1991 British Social Attitudes Survey (BSA), in which half of the sample was drawn from PAF and the other half from the Electoral Register. This experiment enabled a detailed comparison of sample efficiency, weighting efficiency, survey costs, response rates, and estimates of key classificatory variables. Such a split sample experiment could be employed in the future on surveys considering an alternative frame to PAF. The results for BSA 1991 showed PAF to be preferable, primarily due to lower bias in key estimates.

More recent evidence from elsewhere in Europe also provides useful transferable lessons for comparison and evaluation of the different sampling frame options now emerging. These are driven by the desired properties of a general population sampling frame: minimal under-coverage or over-coverage, unique entries, inclusion of identifying information, straightforward access, low cost, and ease of manipulation. It is worth noting the existence of trade-offs between these, notably the greater difficulty of accessing sampling frames containing more identifying information.

Metrics for evaluating the quality of general population sampling frames

Key metrics for assessing and comparing sampling frame quality are as follows (Junes in Lynn and Lyberg 2022):

- Whether the frame has multiple sources of information, for example a population register matched with additional administrative data.
- How frequently the frame is updated.
- Estimated under-coverage rate.
- Which subgroups are disproportionately under-represented.

An additional metric identified in evaluation of the England and Wales Census 2021 frame (ONS 2023) is over-coverage, in the form of duplicate or ineligible entries. Estimates of ineligible addresses in major general population survey samples within the UK since 2022 vary, but can exceed 10%. The 2022/23 National Survey for Wales reported 10.6% of the issued addresses were not eligible (Jenkins and Forbes 2024).

Overall accuracy of the entries within the sampling frame is another crucial metric. In a survey context, this means that the available information has to be sufficient to allow

contact with the sampled case. For a face-to-face survey, an interviewer should be able to locate the address (and named individual, if that information is included). For a web or telephone survey, a survey invitation should be deliverable to the sampled address or individual. When using PAF and other sampling frames based on PAF, maintenance by Royal Mail provides a high level of confidence that addresses will be complete and able to receive mailings. When additional data is matched onto PAF, or administrative data used as a sampling frame, further assessment is required to determine whether addresses remain accurate enough for contact.

Challenges when applying an evaluation framework

Calculating some of the quality metrics listed above is not straightforward and requires additional data sources. Census data and regularly updated ONS mid-year population estimates are vital reference points for the evaluation of general population sampling frames, as well as historical fieldwork outcomes.

Postcode Address File as a sampling frame

Properties of PAF

Royal Mail's Postcode Address File (PAF) has been the most popular UK general population sampling frame for around forty years and is therefore tried and tested across many surveys. It contains more than 30 million addresses and is continuously updated via information from Royal Mail workers making deliveries. Additional expansions of PAF are available that identify addresses with multiple residences and addresses planned or under construction but not yet built. Users of PAF can have high confidence that addresses within it exist and that mail can be delivered to them, as Royal Mail states it is, '*the UK's most up-to-date and accurate address database*' (Royal Mail 2024).

PAF provides trusted and clean address data. It is updated quarterly, is straightforward to access, and raises no confidentiality issues. Although Royal Mail charges for PAF access, this is via licence rather than per use and licences are free for public sector organisations (Royal Mail 2026). Charges vary by type of organisation, the most expensive full PAF licence being for a corporate group at £180,000. However, small businesses and charities can apply for free access (Royal Mail 2025a). For analysis and weighting of the resulting survey data, Census measures, indices of multiple deprivation, and other area-level variables can be easily linked to the data, using postcodes or grid references as identifiers.

Limitations of PAF

PAF was not created for use as a general population sampling frame and is therefore not ideally suited for this purpose. It is a database of postal delivery points, rather than households or persons. Entries in PAF are distinguished by small user and large user classification. A large user is an organisation that receives a minimum of a thousand items of mail a day. Such addresses can be dropped from general population sampling frames, as they are evidently not residential. Small users include organisation names, which enable the identification of some, but not all, commercial addresses within PAF. Royal Mail also keeps records of instances where multiple households share a letter box, known as multiple residency (Royal Mail 2025b). An example of this would be a house with a single front door that has been subdivided into several flats. However, PAF does not include information that would allow student accommodation, holiday homes, nursing homes, and other types of temporary or institutional accommodation to be dropped from a survey sampling frame. While residents of institutions may form part of the eligible population of interest for some UK surveys, temporary visitors in holiday accommodation would not be considered part of the general population.

Nor does PAF provide an indication as to whether the address is occupied or vacant. As a result, it is not possible to systematically drop ineligible addresses from PAF

samples prior to issuing them. Communal residential establishments would require 'time-consuming and costly' screening to identify within PAF (Pickering 2008). As delivery points are not congruent with households, some form of selection within addresses is frequently required, which is a possible source of sampling bias if non-random and a source of additional variance due to differing selection probabilities. Face-to-face surveys can instruct interviewers to randomly select a household within an address if more than one is found, however this is not possible on push-to-web or push-to-telephone surveys.

Case studies

Many large-scale, long-established, and important general population surveys within the UK have used PAF as their sampling frame for decades. These include:

- The British Social Attitudes Survey (NatCen 2025)
- England's National Travel Survey (DfT 2025)
- The UK Labour Force Survey (ONS 2026)
- The Crime Survey for England & Wales (Verian 2026) and Scottish Crime and Justice Survey (Scottish Government et al. 2026)

Addressbase as a sampling frame

Properties of Addressbase

AddressBase Premium combines the PAF for Great Britain with two additional datasets: the local government National Land and Property Gazetteers and the Ordnance Survey Master Map (Fone et al. 2016), using unique property identifiers. It was developed from a similar one-off product used for the 2011 census and covers Great Britain only. A separate product, Addressbase Islands, covers Northern Ireland (Ordnance Survey 2025a). In addition to the properties of PAF, Addressbase provides basic classification of property types into residential or commercial. It is updated every 6 weeks. Detailed technical specifications are available on the Ordnance Survey website (Ordnance Survey 2025b), including the definitions and attributes of each variable included in an Addressbase dataset. Like PAF, Addressbase products are available via licence and free access options can be provided to public sector and charitable bodies on application. However, the cost for other users depends upon how and why the data is accessed via Ordnance Survey's partner network. The pricing structure of Addressbase as a sampling frame is therefore more difficult to predict than that of PAF.

Limitations of Addressbase

Published evidence using Addressbase to date acknowledges the usefulness of its property type classification, while questioning whether they are fully accurate. Lansley et al (2019) attempted to link Addressbase data with the electoral register and land registry in order to create a consumer register, concluding that, 'Address-based frames are imperfect and also create data issues where they do not perfectly correspond to household definitions or property use categories'. There is no specific evidence to validate the property type classification and a single address may include multiple uses. While mixed uses are also an issue for PAF, there is a risk when using Addressbase that addresses with residents could be accidentally excluded from a sampling frame as they have been flagged as commercial.

Addressbase and PAF also share common limitations; no information on whether a residential address is occupied or vacant, or how many households in a residential address, or whether it is a communal establishment.

Case study

Evaluations comparing PAF with Addressbase as a sampling frame are either unpublished or not yet undertaken. Addressbase has however been used to identify residential addresses for England and Wales' largest population survey, the 2021 Census and this constitutes the most systematic evaluation of Addressbase yet completed. Quality assurance of the census frame used multiple administrative datasets included the NHS Patient Demographic Service (PDS), Council Tax

information, Valuation Office Agency data, and the English School Census, which were combined into a Census Intelligence Datasource (CID) to confirm private home and residential institution locations. This exercise comparing a copy of Addressbase taken in summer 2020 to the CID estimated less than 0.5% under-coverage of households in England and Wales in AddressBase (ONS 2023). At the Local Authority (LA) level, only one small LA in England did not meet the quality target of no more than 2% under-coverage.

The primary priority when preparing the census frame was understandably avoiding under-coverage. Nonetheless, over-coverage in Addressbase was also estimated. The initial scale of discrepancy between Addressbase residential records and council tax records was checked in June 2020. This exercise revealed that 25,776,300 addresses were classified as residential by Addressbase and had a council tax record, while a further 1,122,800 were classified as residential by Addressbase but did not have a council tax record (ONS 2023). In addition, 154,200 addresses did not have a residential classification in Addressbase but did have a council tax record. These addresses were treated as eligible. The inclusion of addresses classified as residential by Addressbase which did not have council tax records was the main source of over-coverage, which was estimated at 2.1% in England and 1.7% in Wales. This exceeded the relevant quality target and again varied by LA. 0.42% of addresses in the frame were also identified as duplicates, largely due to the same address appearing multiple times in Addressbase with different UPRNs. In a sampling frame context, these addresses would have two probabilities of selection.

The key conclusion to draw from this evaluation for sampling purposes is that over-coverage of non-residential addresses in England and Wales is much more significant in Addressbase than under-coverage of residential addresses. Of the total 27,053,000 residential addresses identified in the Census quality evaluation as residential by either Addressbase or council tax records, 95% were classified as residential by both, a strong indication that they were indeed occupied residences. 4.2% were classified by Addressbase as residential but did not have a council tax record, requiring further investigation by ONS to determine their status. 0.6% of addresses had a council tax record but were not flagged as residential by Addressbase, indicating inaccurate classification.

Scotland and Northern Ireland's most recent censuses used slightly different methods to prepare their frames, but no reports on the quality evaluations of these approaches have been published to date and so the findings from England and Wales cannot be generalised across the UK. The results nonetheless suggest that, in England and Wales, Addressbase may offer advantages over PAF when seeking to differentiate residential addresses from commercial as the classification of residential addresses is broadly supported by council tax records. However, there is some under-coverage and a greater level of over-coverage, so use of Addressbase rather than PAF for sampling in England and Wales would not result in near-zero ineligible addresses being selected. No explicit comparison of the AddressBase classification against the small/large-user PAF classification is available.

Administrative data linkage to address sampling frames

No research specifically matching administrative data onto PAF or Addressbase to create a general population sampling frame was found in the literature search. Instead, there is considerable evidence on matching administrative data for other reasons. Evidence in this area can be divided into three broad purposes:

- Linking administrative data to survey responses for enhancement or validation,
- Linking administrative data to sampling frames for subgroup identification,
- Linking administrative data to address frames for population estimation.

This evidence provides lessons that are relevant to the potential use of administrative data in general population sampling frames.

How administrative data is currently used with general population sampling frames

Administrative data varies considerably in coverage, content, and quality, as the UK has no single population register. The majority of evidence reviewed involved health record data matching, reflecting the importance of survey research in the health sector as well as more established data-sharing protocols. Screening PAF for residents in particular age groups using health records has been successful for multiple surveys, such as the Northern Ireland Cohort for the Longitudinal Study of Ageing (NICOLA; Neville et al. 2023) and the Scottish Health Survey child boost (ScotCen 2024). For each of these surveys, addresses and demographic data in the relevant health database were used to flag sampled addresses where an individual in the age group of interest lived. While this process cannot be completely accurate, face-to-face survey fieldwork enables monitoring of how well it works. In the Scottish Health Survey 2023, approximately 84% of PAF addresses flagged by health records as inhabited by a child did indeed have at least one child resident (ScotCen 2024a). This approach enables cost reductions as interviewers need to screen fewer addresses to achieve the required level of response from particular age groups. When broad age categories are used, exact matching to named records is not essential and confidentiality implications are minimised. However, it is not possible to measure the level of under-coverage that occurs as a result of sampled addresses failing to be identified in health records. There is a risk of this under-coverage being geographically or demographically non-random and thus introducing systematic bias.

The available evidence consistently reports that matching administrative data onto PAF or another address frame can be a difficult, protracted, and inexact process. Unique identifiers shared between databases are lacking, requiring fuzzy matching methods using text strings. Even with clean survey responses (including name and date of birth) in addition to PAF addresses, more than 95% but not all cases could be matched to health records where 100% matching is desirable (Whiffen et al. 2020; Tingay et al. 2019). At the initial sampling stage, the match rate is likely to be more difficult to measure due to limited data. For example, if a large number of health records matched a sampled address, contact with the household would be required

in order to confirm which were current residents. In the context of an age group boost sample, the match with health records does not need to be exact as long as someone of the required age is resident at the address. It would be a much greater challenge to take a general population sample and attempt to identify the residents of the sampled addresses from health records.

How administrative data could be used to enhance a general population sample frame

In future, being able to screen PAF for broad age categories using health records could, in theory, enable differential response rates to be reflected in the sample stratification, improve the precise of estimates, and facilitate subgroup analysis. At present, it is unknown to what extent such use of health data would improve upon use of granular Census data, which captures the age distribution at Output Area (OA) level. Each OA in England includes between 100 and 625 heterogenous individuals, offering limited potential to oversample by age, for instance. However, incomplete matching of health records could worsen precision if addresses were assigned to the wrong sampling stratum. Moreover, for reasons of data confidentiality adding individual health records onto an address frame is highly unlikely to be possible for surveys other than the Census.

A close approximation to a named general population sampling frame was achieved by a project (Lansley et al. 2019) to create a 'linked consumer register' combining public electoral rolls, PAF, Addressbase Premium, Land Registry records, and commercial databases. This project used a two-stage process to combine data sources. First, a common address spine was constructed, then household composition was attributed to each address. The initial building blocks of the address spine were public versions of electoral rolls from 1997 to 2016, although the authors acknowledge that the rate of opt-outs to this public register has risen considerably since 2003. These were combined with 'consumer registers' from two commercial data brokers, as well as PAF, Addressbase Premium, and Land Registry records for England, Wales, and Scotland. It is important to note that Addressbase and Land Registry data for Northern Ireland was not included, therefore the database is inconsistent across nations. The main barrier to combining these data sources is inconsistency in how names and addresses are recorded, in the absence of any common unique identifier. Unique Property Reference Numbers (UPRNs) are intended to provide this; however these are not present in all databases.

Prior to matching with PAF and Addressbase, address data from administrative and commercial data sources required cleaning by reformatting and standardisation, such as expanding common abbreviations. Changes in postcodes were identified using a Royal Mail lookup and possible duplicates filtered out by searching for multiple-unit postcodes with identical reference co-ordinates. Multiple matching methods were required: rules-based, occupier-based (using concatenated names to search for duplicates within the same postcode for each source), and fuzzy.
Within the final address spine, 87.5% of addresses were found in both the electoral and commercial registers as well as Addressbase, while an additional 2.6% were in

both registers and could be matched to PAF but not Addressbase. 9.1% of addresses could not be matched to PAF or Addressbase but occurred in multiple registers, while the remaining 1% could be found only in a single commercial or electoral register (Lansley et al. 2019). Combining an address frame with administrative data was a major challenge, even before adding household composition data. The intended use of the final register was for timely, granular population estimates rather than as a sampling frame, so it is unlikely to be sufficiently accurate to use as a general population sampling frame for selecting individuals. A notable uncertainty is the role of the commercial data, the sources of which are not stated and may be commercially confidential. If this is of lower quality than the administrative data sources, inclusion could reduce rather than improving data quality.

Key lessons for general population sampling frames from the evidence on matching administrative data are as follows:

- Gaining access to administrative data can be a lengthy and complex process, especially for a new purpose, and can vary considerably across the UK nations and between public bodies.
- Once access is agreed, matching inconsistent and messy administrative (or commercial) data onto PAF or Addressbase is a technical challenge. While a high match rate can be achieved, there is no established straightforward means of validation.
- Data linkage is nonetheless receiving increasing attention as a means of using existing administrative data for social research purposes. Methods are advancing and public sector bodies are willing to share data once robust agreements are in place.

At present, the risks around the practical difficulty and uncertain accuracy of incorporating administrative data into general population sampling frames may appear excessive when PAF is reliable and deemed ‘good enough’ for many intended purposes. Further research is needed to reduce this uncertainty, establish replicable processes, and quantify the benefits in terms of cost reduction and improved representativeness in responding samples.

Administrative data as a general population sampling frame

Evidence on the use of administrative data as a general population sampling frame without PAF or Addressbase is very limited, as the UK does not have an administrative database that is intended as a general population register covering all residents. When UK surveys use administrative data as a sampling frame, their population of interest is almost always a particular subgroup such as students or benefit recipients. However, the data linkage literature demonstrates that National Health Service (NHS) patient registers such as Scotland's Community Health Index (CHI) have particularly high coverage of the general population, the vast majority of whom will have contact with the NHS at some point.

A single case study of using health records as a general population sampling frame was found. While this is insufficient to generalise from, it can offer key considerations for future projects. Moreover, new administrative data sources with potential for use as sampling frames are emerging, such as the RAPID database of Department for Work and Pensions (DWP) and His Majesty's Revenue and Customs (HMRC) records (ADR UK 2026).

Case study

The 2023 Lothian Health Survey was a push-to-web general population survey of adults aged 16 or above living in the area of Scotland covered by the Lothian Health Board (NatCen 2024). The sample was drawn from Community Health Index (CHI) records, with no data matching to other sources. For sampling purposes, the frame was anonymised and only included a unique identifier for each case and the agreed sampling stratification variables: age, sex, Local Authority, and deprivation deciles. During preparation of the sampling frame, cases with incomplete addresses and those whose death had been registered were removed. However, there was no check on how recently each health record had been updated, therefore the likelihood of information being outdated could not be estimated.

The sample stratification variables provided in the frame were used to evaluate the coverage and quality of CHI records. The total number of CHI records in the frame was considerably higher than both 2021 MYEs and 2022 Census estimates of the 16+ population of the Lothian area, an overall difference of +17% from the Census estimate (NatCen 2024). The scale of divergence between CHI records and MYEs also varied both by Local Authority and age group. It was greatest for 16–24-year-olds in the City of Edinburgh, with in excess of a third more individuals in this subgroup in the sampling frame than ONS population estimates.

There was no means of identifying which CHI records could be ineligible at the sampling stage and, due to the push-to-web mode, the true rate of ineligible cases in the issued sample was also unknown. The percentage of letters returned as undeliverable (2.6%) or as named person not at the address (5.2%) provide an indication. The letter returns could indicate inaccurate data entry or addresses that

no longer exist; the absence of the named person could indicate outdated or potentially duplicate records. Manual searches of sampled addresses identified locations that did not appear to be private residences or residential institutions, such as colleges, offices, and sports clubs. Responses were received from some of these addresses, at a much lower rate than the rest of the sample.

Key lessons from this case study include:

- When using a new sampling frame, the sample design must incorporate outcomes of a quality evaluation. Comparison with population estimates should form the starting point, but future projects should also consider requesting tables of available variables not used for sample stratification but important for analysis of survey outcomes, such as deprivation percentile or ethnicity.
- Direct access to an anonymised health record sampling frame need not be required, if the data controller is willing and able to draw the sample to an agreed specification then share confirmation they have done so (including code used to conduct the sampling). This could reduce the complexity of agreeing data sharing, therefore saving time. However, there would minimal difference in data security risk of the two approaches, as long as the anonymised sampling frame and sampled cases could not be linked.
- Future use of health records as a proxy for the general population would need processes to identify and screen out duplicate or ineligible cases. Without this, there is a risk of introducing bias that cannot be controlled for with weighting, as rates of ineligibility can vary by subgroup and are unknown in push-to-web or push-to-telephone surveys. In a face-to-face survey, interviewers would identify ineligible cases in the issued sample and these could be removed from the base during weighting. Without identification of ineligible cases, using selection and non-response weights that adjust the survey responses to match the sampling frame profile could be unhelpful.
- Health record address data is of lower quality than entries in PAF or Addressbase and therefore requires cleaning to ensure that addresses can be found by an interviewer or survey invitations can be delivered.

If a process were established to screen, clean, and validate names, addresses, age, and sex in health records, they could in theory provide a general population sampling frame with major advantages over PAF and Addressbase. In practise, health records are not comprehensive, up-to-date, or collected for the purposes of sampling. Any project using them as a sampling frame needs a strong justification for doing so in order to comply with the UK Data Protection Regulation (GDPR). Further research incorporating additional sample quality evaluation and validation is needed to determine whether health records, or other administrative data, could provide suitable sampling frames for a larger-scale general population survey. To date, administrative records have been used as sampling frames almost exclusively for surveys of specific subgroups.

Conclusions

The current state of UK general population sampling frames

UK general population surveys have relied upon the Postcode Address File as a sampling frame for decades. Well-established sampling and weighting methodologies are available to minimise the impact of its limitations. While Addressbase potentially offers an incremental advance on PAF, enabling commercial addresses to be dropped from sampling frames, higher cost and absence of robust validation of the address classification are barriers to adoption. Administrative data use in general population surveys remains almost entirely confined to targeting subgroups. A case study using health records as a general population sampling frame suggests there are major data quality drawbacks to doing so.

The future of UK general population sampling frames

At present there is understandable caution about moving away from PAF, due to the potential for significant costs, additional bias, or time series breaks. Further work is needed to enable further innovation in general population sampling frames. This should include:

- Development of a robust quality evaluation framework for comparison of general population sampling frames, building on the summary in the first section of this guide.
- Evidence validating the property type classifications in Addressbase Premium.
- Assessments of the accuracy of demographic information and addresses in administrative databases, which would need to be conducted by the data controllers.

The ability to exclude non-residential addresses from general population sampling frames would be an incremental improvement for survey practitioners, as it would reduce costs for all survey modes. This would also improve the quality of weights for push-to-web and push-to-telephone surveys, as ineligible and non-responding addresses cannot be distinguished without an interviewer. A much more significant advance would be the availability of variables to use for individual-level rather than area-level stratification, such as age and sex. Case studies to date suggest the challenges of using existing administrative data in this way are considerable, but not insurmountable.

Ultimately, answering the question of whether there are better UK general population sampling frames than PAF requires parallel run experiments, using face-to-face interviewing so ineligible addresses can be identified. Such an experiment informed the transition from the use of the electoral register to PAF (Lynn and Taylor 1995) by comparing the performance of samples from two different frames for the same survey and fieldwork period in terms of cost, response rate, sampling efficiency, weighting

efficiency, and key variable estimates. The same experimental approach would also be appropriate for the next generation of general population sampling frames.

Reference list

ADR UK. 2026. ‘RAPID: Population Interactions with Employment, Benefits and Pensions - UK - ADR UK’. https://www.adruk.org/data-access/flagship-datasets/rapid-population-interactions-with-employment-benefits-and-pensions-uk/.

Butcher, Robert. 1988. ‘The Use of the Post-Code Address File as a Sampling Frame’. *Journal of the Royal Statistical Society: Series D (The Statistician)* 37 (1): 15–24. https://doi.org/10.2307/2348375.

DfT (GOV.UK). 2025. ‘NTS 2024 Technical Report Chapter 2 Sample Selection’. https://www.gov.uk/government/statistics/national-travel-survey-2024-technical-report/chapter-2-sample-selection.

Fone, David, Jennifer Morgan, Richard Fry, et al. 2016. ‘Change in Alcohol Outlet Density and Alcohol-Related Harm to Population Health (CHALICE): A Comprehensive Record-Linked Database Study in Wales’. *Public Health Research* 4 (3): 1–184. https://doi.org/10.3310/phr04030.

Jenkins, Sarah, and Nick Forbes. 2024. *National Survey for Wales 2022-23: Technical Report*. 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.

Lansley, Guy, Wen Li, and Paul A. Longley. 2019. ‘Creating a Linked Consumer Register for Granular Demographic Analysis’. *Journal of the Royal Statistical Society Series A: Statistics in Society* 182 (4): 1587–605. https://doi.org/10.1111/rssa.12476.

Lynn, Peter, and Lars Lyberg, eds. 2022. ‘Frame Errors in EU-SILC: Under-Coverage’. In *Improving the Measurement of Poverty and Social Exclusion in Europe: Reducing Non-Sampling Errors*. https://op.europa.eu/en/publication-detail/-/publication/798e3ef9-fe65-11ec-b94a-01aa75ed71a1/language-en.

Lynn, Peter, and Bridget Taylor. 1995. ‘On the Bias and Variance of Samples of Individuals: A Comparison of the Electoral Registers and Postcode Address File as Sampling Frames’. *Journal of the Royal Statistical Society. Series D (The Statistician)* 44 (2): 173–94. https://doi.org/10.2307/2348443.

NatCen. 2024. *NHS Lothian Public Health Survey 2023 Technical Report*. https://services.nhslothian.scot/wp-content/uploads/sites/105/2024/06/NHS-Lothian-Public-Health-Survey-2023-Technical-Report-.pdf.

NatCen. 2025. *British Social Attitudes 2024 Technical Details*. https://natcen.ac.uk/sites/default/files/2025-06/BSA%2042%20Technical%20Details.pdf.

Neville, Charlotte, Frances Burns, Sharon Cruise, et al. 2023. 'Cohort Profile: The Northern Ireland Cohort for the Longitudinal Study of Ageing (NICOLA)'. *International Journal of Epidemiology* 52 (4): e211–21. https://doi.org/10.1093/ije/dyad026.

ONS. 2023. 'Evaluation of Addressing Quality: Census 2021 - Office for National Statistics'. https://www.beta.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/evaluationofaddressingqualitycensus2021.

ONS. 2026. 'Labour Force Survey – User Guidance - Office for National Statistics'. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance.

Ordnance Survey (Ordnance Survey). 2025a. 'AddressBase | Data Products | OS'. https://www.ordnancesurvey.co.uk/products/addressbase.

Ordnance Survey. 2025b. 'AddressBase Premium Technical Specification | OS Download Products' Documentation'. September 5. https://docs.os.uk/os-downloads/products/addresses-and-names-portfolio/addressbase-premium/addressbase-premium-technical-specification.

Pickering, Kevin. 2008. *British Crime Survey: Options for Extending the Coverage to Children and People Living in Communal Establishments*.

Royal Mail. 2024. 'PAF'. *Powered By Paf*. https://www.poweredbypaf.com/product/paf/.

Royal Mail. 2025a. *2025 PAF Pricing Factsheet*. https://www.poweredbypaf.com/wp-content/uploads/2025/10/2025-Pricing-Factsheet.pdf.

Royal Mail. 2025b. 'Multiple Residence Data - PAF Complementary Product'. *Powered By PAF*. https://www.poweredbypaf.com/postcode-address-file/multiple-residence/.

Royal Mail (Powered By PAF). 2026. 'Public Sector Licence'. https://www.poweredbypaf.com/public-sector-licence/.

ScotCen. 2024a. *ScotCen Unpublished Internal Report Regarding the Scottish Health Survey 2023*.

ScotCen. 2024b. 'The Scottish Health Survey 2023 - Volume 2: Technical Report'. https://www.gov.scot/publications/scottish-health-survey-2023-volume-2-technical-report/.

Scottish Government, ScotCen, and Ipsos Scotland. 2026. *Scottish Crime and Justice Survey 2024/25 Technical Report*. https://www.gov.scot/binaries/content/documents/govscot/publications/statistics/2026/03/scottish-crime-and-justice-survey-2024-25-supplementary-

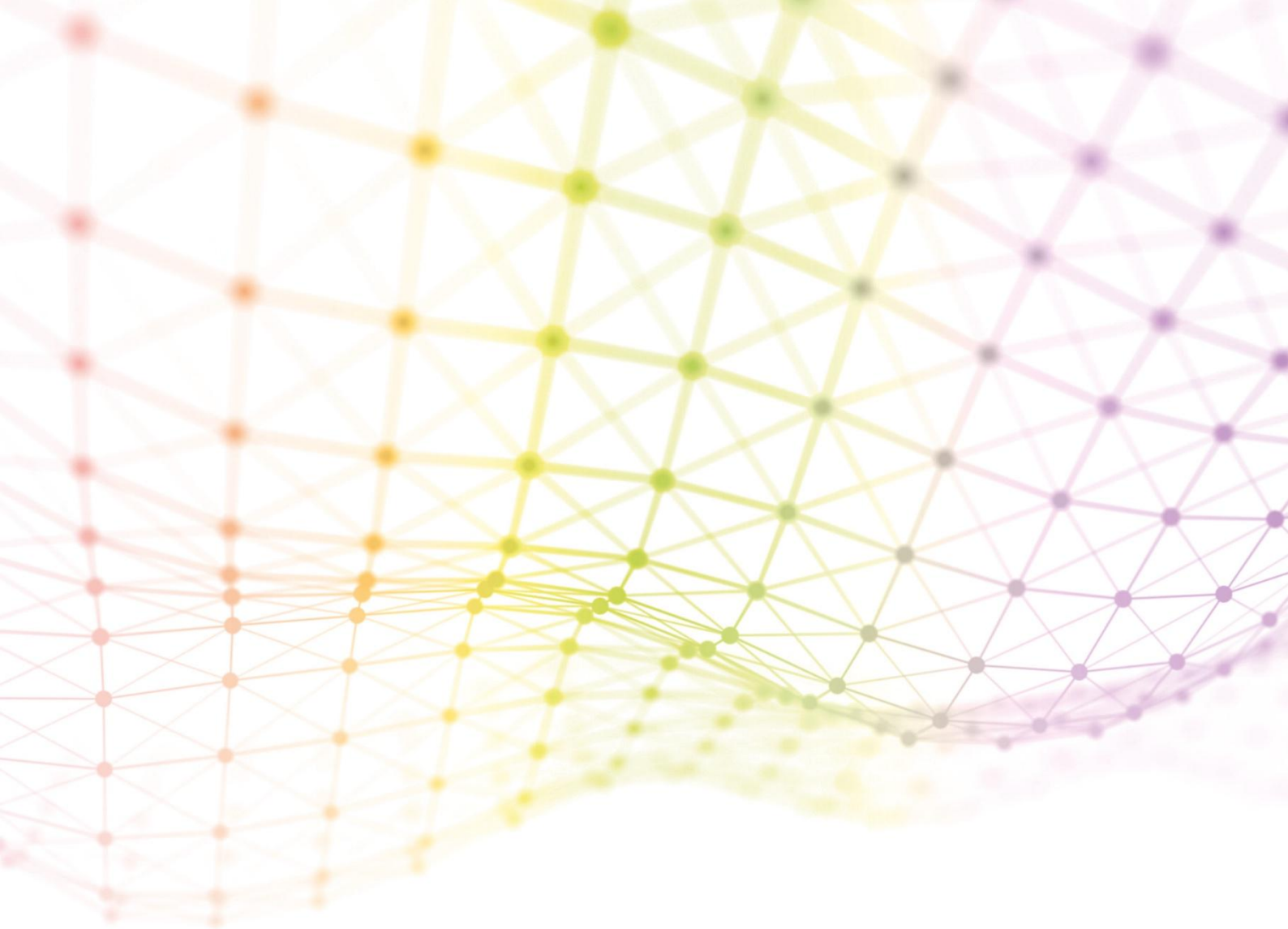
documents/documents/technical-report/technical-report/govscot%3Adocument/scjs-2024-25-technical-report.pdf.

Tingay, Karen Susan, Amrita Bandyopadhyay, Lucy Griffiths, et al. 2019. 'Record Linkage to Enhance Consented Cohort and Routinely Collected Health Data from a UK Birth Cohort'. *International Journal of Population Data Science* 4 (1): 579. https://doi.org/10.23889/ijpds.v4i1.579.

Verian. 2026. *Crime Survey for England and Wales Technical Report 2024/25*. https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/crimeandjustice/methodologies/crimeandjusticemethodology/csew202425technicalreportvol1final.pdf.

Whiffen, Tony, Ashley Akbari, Tony Paget, Sarah Lowe, and Ronan Lyons. 2020. 'How Effective Are Population Health Surveys for Estimating Prevalence of Chronic Conditions Compared to Anonymised Clinical Data?' *International Journal of Population Data Science* 5 (1): 1. https://doi.org/10.23889/ijpds.v5i1.1151.

Wilson, Paul. R., and Elliot, D. J. 1987. 'An Evaluation of the Postcode Address File as a Sampling Frame and Its Use within OPCS'. *Journal of the Royal Statistical Society. Series A (General)* 150 (3): 230–40. https://doi.org/10.2307/2981474.



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