



SURVEY FUTURES

**SURVEY DATA COLLECTION
METHODS COLLABORATION**

Working Paper 17: **Enhanced sampling frames: A literature review**

Anna Keyes¹, Paul A. Smith²

¹ National Centre for Social Research

² University of Southampton

May 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.

This paper is an output of *Survey Futures* Research Strand 1, “Enhanced sampling frames and procedures,” led by Professor Paul A. Smith (University of Southampton).

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:

Keyes A, Smith P A (2026) ‘Enhanced sampling frames: A literature review’, *Survey Futures Working Paper* no. 17. Colchester, UK: University of Essex. Available at <https://surveyfutures.net/working-papers/>.

Executive Summary

Survey researchers seeking a representative sample of the UK general population do not have the advantage of access to a population register, as there is no such register in the UK.

Databases of addresses are instead widely used as general population sampling frames. Royal Mail's Postcode Address File (PAF) has been the most popular of these for around thirty years. Although it has the advantage of containing reliable and regularly updated address data, it does not enable the identification of inhabited residences.

A more recent address database product, Addressbase Plus, adds administrative data onto PAF with the aim of differentiating between residential and commercial properties. However, this classification is imperfect. Addressbase Plus was used as the initial address frame for the 2021 England and Wales Census, then council tax records and other administrative data were used to confirm private home and residential institution locations. Given the known limitations of using address databases as general population sampling frames, there has been increasing interest in the use of individual-level administrative data in survey literature and practice.

Now that alternatives to using PAF on its own are available, there is a need to set evaluation criteria for UK general population sampling frames. Outside the UK context, there is European research to draw upon, suggesting the following assessment criteria:

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

General population sampling frames cannot be evaluated in isolation and require recent population estimates for comparison.

The addition of administrative data to PAF has been trialled for various purposes, including identification of subgroups, such as children, to be oversampled and to create small area population estimates. Research to date has emphasised the challenges of combining administrative databases with PAF, due to the imprecision of the matching process. The use of health records as a general population sample frame has also been trialled in the 2023 Lothian Health Survey. Analysis of the survey sample estimated the percentage of ineligible records using comparison with Census and Office for National Statistics (ONS) mid-year estimate data. The results suggested that ineligibility levels in the frame varied by Local Authority and age group and could be high. Examination of the address data also showed that it was of lower quality than PAF, containing errors that made some survey invitations undeliverable.

In conclusion, decision-makers cannot yet have confidence that alternatives to PAF would offer benefits without risks. For this to change, a quality evaluation framework for alternative sampling frames would need to be applied, requiring access to any administrative data sources being considered for sampling purposes. In the short term, incremental benefits could be offered by Addressbase but there is limited evidence for these.

1. Introduction

General population surveys conducted in the United Kingdom all contend with the challenge of achieving a representative sample when no single national register of the population exists to use as a sampling frame. Unlike most other European nations, in the UK records of names and addresses are fragmented across many administrative bodies. Access to these records for sampling general population surveys remains limited, although there is increasing recognition of their potential. In recent decades, UK general population surveys have instead relied upon sampling addresses without any information on who lives within. The Postcode Address File (PAF), administered by Royal Mail, is the most widely-used such database. A more recent option for sampling addresses is Addressbase, a product offered by Ordnance Survey that matches PAF records to unique property reference numbers (UPRNs) and classifies addresses by residential or commercial use. Both provide valid addresses that survey invitation letters can be sent to, but neither can offer residents' names or any of their demographic information.

Since the COVID-19 pandemic, response rates to UK general population surveys have generally fallen (Maslovskaya et al. 2025) especially among harder to reach demographic groups, which is increasing the challenge of achieving representative samples. Consequently, there has been increased innovation in sampling and weighting over the past five years. Moreover, demand for high-quality survey data from public bodies has encouraged a willingness to share administrative records for sampling purposes in some nations. Reviewing the available evidence on basic and enhanced sampling frames is therefore timely, so that best practise and future opportunities can be identified. Where administrative data has already been used in sampling frames, it is important to identify the benefits and costs to inform forthcoming surveys.

The aims of this literature review are threefold. Firstly, to investigate the current properties of PAF and Addressbase as sampling frames for UK general population surveys. Secondly, to identify the known advantages and drawbacks of using administrative data sources in combination with or instead of PAF and Addressbase. Finally, to synthesise advice from the academic and grey literature for survey practitioners on best practise for general population sampling frames, as well as highlighting key areas of uncertainty where evidence is currently lacking.

2. Scope and Approach

The purpose of this literature review is to synthesise and inform practise as well as research, therefore materials from the grey literature (such as technical and government reports) were included in scope as well as journal papers and academic books. Many general population surveys conducted in the UK are commissioned and funded by public sector bodies, therefore a considerable proportion of the relevant work can be found in the grey literature. However, open access to datasets from such surveys has enabled a wide range of published academic research, some of which is also relevant to this review. Given the specificity of issues around achieving a representative general population survey sample in the UK, studies were screened

out if they were not written in English, only applicable outside the UK, or focused only on a population subgroup. No time limit was applied to the searches, however very few relevant materials were found to be published more than fifteen years ago.

Relevant materials were identified by searching academic databases such as Google Scholar using a series of keyword strings. Searches proceeded from more general to more specific. Initially broad strings were attempted to estimate their prevalence in the databases ('Postcode address file survey' and 'postcode address file survey data linkage'). These searches both returned more than ten thousand results from Google Scholar, so the strings were narrowed until a more reasonable number of potentially relevant results were achieved. The string "'postcode address file" + survey sampling' returned approximately one thousand results and 'addressbase sampling frame survey' returned less than four hundred, reflecting the more recent adoption of Addressbase. Search results and abstracts were visually screened to weed out research that simply analysed data from a survey without any evaluation of the sampling methodology. Such research is very common and not easy to distinguish from methodological work using search terms, so a high proportion of search results were deemed outside the scope. Relevant materials were saved in a Zotero database.

3. The Postcode Address File (PAF)

The Postcode Address File (PAF) website states that it is 'the UK's most up-to-date address database' (Royal Mail 2024). It is managed by Royal Mail, contains more than 30 million addresses, and is continuously updated via information from Royal Mail workers making deliveries. PAF is licensed for public, charitable, and commercial use for a variety of purposes, for example allowing businesses to verify delivery addresses. A PAF license provides access to a comprehensive and regularly updated database of UK addresses, with obvious utility for survey sampling. 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. However, PAF does not identify whether an address is residential or of another type, nor does it provide an indication of whether the address is occupied or vacant. When a sample is drawn from PAF for a general population survey, a percentage of the addresses selected are therefore likely to be ineligible as no-one is living in them.

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 2025). 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.

PAF became a popular sampling frame for UK general population surveys in the latter decades of the twentieth century. A 1988 paper in the journal of the Royal Statistical Society (Butcher 1988) discussed the impact on major government surveys of shifting to PAF as the sampling frame. The most popular option prior to this was sampling from electoral registers. Lynn and Taylor (1995) compared the two sampling frames and conclude that PAF had better coverage and was 'clearly the superior frame for sampling households'. The rationale for this is evident: PAF has excellent coverage of residential addresses, while electoral registers have only partial coverage. Addresses with no residential electors cannot be sampled from electoral registers. To illustrate the current extent of disparity between the two, the most recent Office for National Statistics (ONS) figures on electoral registrations state there were 48,844,292 local government electoral registrations in the UK as of December 2021 (ONS 2022a). The 2021 mid-year population estimate of people aged 18 or over living in the UK was 53,188,204 (ONS 2022b), suggesting that more than 4 million adults are not on local government electoral registers. It should be noted that entitlement to vote in the UK and residence in the UK are not fully congruent. British citizens, qualifying Commonwealth citizens, and citizens of the Republic of Ireland can vote, even if not resident in the UK. Conversely, residents in the UK who are citizens of nations outside the EU cannot vote (Johnston 2024). Moreover, electoral registers are opt-in and registration needs to be regularly renewed, whereas UK residents cannot remove their address from the PAF.

In addition to concern about the total number of persons on electoral registers, literature comparing them with PAF notes that some subgroups of adults are less likely to be registered. Sampling from electoral registers will therefore under-represent these groups, representing a source of bias. Lynn (1997) conducted an analysis of the British Crime Survey (BCS), which had used PAF as its sampling frame for the first time in 1992 after sampling from electoral registers in previous waves. The study examined what proportion of the PAF sample and responding sample were also present on electoral registers. It was found that 7.7% of achieved BCS interviews were carried out at addresses not on electoral registers. Lynn (1997) concluded that households included in PAF but not on electoral registers are demographically distinctive in multiple ways: more likely to be privately rented, a flat rather than a house, of minority ethnicity, single person, and without children. Evidently this may have changed in subsequent decades and there is no suggestion in recent literature that electoral registers should once again be used as a general population sampling frame. Nonetheless, the comparative evaluations undertaken in the 1990s provide helpful precedents for comparing PAF with administrative data sampling frames. This is particularly valuable given the paucity of recent research evaluating options for UK survey sampling frames.

As of 2025, many large, established, and important general population surveys in the UK with extensive time series use PAF as their sampling frame. These include the British Social Attitudes Survey, England's National Travel Survey (NTS), the UK Labour Force Survey, and Crime Surveys for England & Wales and Scotland. PAF provides trusted and clean address data, is updated

quarterly, is straightforward to access and raises no confidentiality issues. For analysis and weighting of the resulting survey data, Census measures, indices of multiple deprivation, and other area-level variables can be easily matched on using postcodes.

The main disadvantages of PAF as a general population sampling frame all result from it not being designed for that purpose. PAF is a database of postal delivery points, rather than persons, and systematically dropping non-residential addresses isn't practicable. On face-to-face surveys, non-residential addresses are identified by interviewers and the percentage recorded as 'deadwood'. The proportion is not insubstantial, for example 8% in England's NTS 2023 (DfT 2024) and 9% in the Scottish Health Survey 2023 (ScotCen 2024c). In addition to non-residential buildings, ineligible addresses include empty residences and second or holiday homes as these do not have permanent residents. Inviting ineligible addresses of this kind to participate adds to the survey costs, without providing any responses. When a survey is web-first, ineligible non-residential addresses cannot be differentiated from non-responding residential addresses, making weighting to adjust for non-response bias more difficult. Moreover, delivery points are not congruent with households, requiring some form of selection within PAF addresses, another possible source of bias and/or variance.

General population surveys also struggle to represent communal establishment residents when using PAF as a sampling frame. A report on extending the coverage of the British Crime Survey to children and communal establishments (Pickering 2008) came to the conclusion that, 'The PAF doesn't identify communal establishments and identifying a useable sample would require considerable address screening, which would be time-consuming and costly'. It is not straightforward to either target or remove communal establishments from PAF samples. Without interviewer visits, the number of eligible residents at any sampled but non-responding address is unknown. This is an increasingly pressing issue for web surveys, which have become more popular since the COVID-19 pandemic began, yet have consistently lower response rates than face-to-face surveys. For example, the 2019 British Social Attitudes Survey (BSA) was conducted entirely face-to-face. During fieldwork 8.6% of sampled PAF addresses were found to be ineligible and 40.5% provided a survey response (NatCen 2020). For COVID-19 safety reasons, BSA 2021 switched to web and phone interviews and 13% of issued addresses provided at least one response, with an unknown level of ineligible addresses (NatCen 2022).

In summary, PAF became the predominant sampling frame for UK general population surveys in the absence of preferable alternatives. Its greatest strength, comprehensive coverage of UK addresses, can also be a weakness in the context of survey sampling. PAF samples contain non-trivial proportions of non-residential and therefore ineligible addresses, increasing fieldwork costs and requiring adjustments when weighting survey responses. To be better suited to use as a survey sampling frame, PAF would ideally include additional information classifying addresses by type of use. Addressbase, developed by Ordnance Survey using PAF, attempts to provide such a database.

4. Addressbase

AddressBase Premium combines PAF 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 2025). In theory, Addressbase provides classification of property types, enabling residential and commercial addresses to be distinguished. In practise, recent literature acknowledges the usefulness of this classification while questioning whether it is fully accurate. In two recent papers, Fone et al. (2016) used Addressbase to identify alcohol outlet density and residential density in the vicinity and Lansley et al. (2019) attempted to link it with the electoral register and land registry in order to create a consumer register. The latter acknowledged that unique identifiers within Addressbase are not present in other databases, therefore, 'A barrier to our core objective of establishing a reliable linked data product is that individuals, businesses or local authorities may use differing conventions for recording names and addresses' (Lansley et al. 2019). The authors do not discuss the use of Addressbase combined with other data as a survey sampling frame, however their conclusions are also applicable to such a use, notably that, 'Address-based frames are imperfect and also create data issues where they do not perfectly correspond to household definitions or property use categories' (Lansley et al. 2019).

As of July 2025, PAF remains the established sampling frame for major UK general population surveys. Evaluations comparing it with Addressbase for this purpose 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: 'The initial frame is based on a copy of AddressBase taken in summer 2020 to give time to print 26 million initial contact letters. A further copy will be taken later in 2020 to account for any new addresses and to flag addresses that should no longer be included on the frame' (ONS 2020). To supplement and validate the Addressbase list, ONS combined it with specific datasets on student halls, care homes, armed forces bases, prisons, boarding schools, and traveller sites from a range of government and other sources. The work ONS undertook to quality assure Addressbase for the 2021 Census therefore constitutes the most systematic evaluation of Addressbase yet completed. The full Census 2021 quality report has not yet been published. However, an assessment of the address frame's quality is available (ONS 2023a) and states that, 'We chose to use AddressBase Premium as the main source for the household section of address frame because it was and still is the most authoritative, current and accurate dataset on addresses in England and Wales'. As ONS does not support this statement with any specific evidence, it rests upon the assumption that the additional information within Addressbase is sufficiently accurate to offer significant enhancement to PAF in England and Wales.

Quality assurance of the Addressbase frame was undertaken using multiple administrative datasets, combined into a Census Intelligence Datasource (CID). These included the NHS Patient Demographic Service (PDS), Council Tax information, Valuation Office Agency data, and the English School Census (ONS 2020). The scale of discrepancy between Addressbase

residential records and council tax records was checked in June 2020 while preparing the initial frame. 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. Further investigation was undertaken by ONS to determine whether the latter group comprised genuine residential establishments. The final census address frame comprised 27,053,200 households and 49,900 communal establishments (ONS 2023a). As noted above, communal establishments could not be reliably identified using Addressbase therefore administrative data for specific establishment types was required.

The primary priority when preparing the census frame was understandably avoiding under-coverage. Over-coverage was of less concern. The CID and census frame comparison estimated that under-coverage within the frame was 0.35% for England and 0.49% in Wales, well within quality targets (ONS 2023a). 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. Over-coverage was also assessed using CID and an average of 2.1% for England and 1.7% for Wales was identified. This exceeded the relevant quality target and again varied by LA. The main cause of over-coverage was stated as the inclusion of addresses classified as residential by Addressbase which did not have council tax records (ONS 2023a). 0.42% of addresses in the frame were also identified as duplicates, largely due the same address appearing multiple times in Addressbase with different UPRNs. In a sampling frame context, these addresses would have two probabilities of selection.

A comparable evaluation of Scotland's census address frame has yet to be published. Scotland's 2022 census also used Addressbase Premium as the basis of the frame. A 'tailored enumeration approach' was then used to classify communal establishments such as holiday parks (Scotland's Census 2025). In Northern Ireland, the 2021 census used the Land & Property Services (LPS) POINTER database as the address frame, rather than Addressbase Islands (NISRA 2024). This was the same approach as NI took for the 2011 census. POINTER is an NI-specific address database maintained by the public sector organisation LPS, using information from Local Authorities and Royal Mail. It is similar to PAF, in that property uses are not identifiable. To prepare the census frame, desk-based checking and liaison with LAs were used to confirm which addresses from POINTER should be included in the census address frame. Addresses for communal establishments were sourced separately, using administrative and public data depending on the type of establishment. The Northern Ireland Statistics and Research Agency Census 2021 General Report stated that, 'coverage measured at the end of the operational phase at around 99% and accuracy measured at around 96%' (NISRA 2024).

The results of England and Wales' census frame quality assurance exercise provide useful lessons for the sampling of general population surveys, although they may not apply to Scotland and Northern Ireland. The results suggest that Addressbase offers advantages over PAF when seeking to differentiate residential addresses from other types, despite the known limitations of its classifications. The levels of under- and over-coverage identified in the census frame cannot however be generalised to using Addressbase as a sampling frame. ONS were able to draw upon a wide range of administrative datasets that other organisations could not readily access in

order to replicate the lengthy and complex census frame preparation process. Rather, the quality levels that ONS was able to achieve suggest an upper limit for what is currently possible in terms of sampling frame coverage using Addressbase. For most general population surveys, the scope of administrative data use would be much more limited. Nonetheless, the results demonstrate that administrative datasets can be used to improve Addressbase coverage and that deadwood can be reduced to much lower levels than are commonly observed in general population surveys sampled from PAF.

5. Evaluation of General Population Sampling Frames

The evaluation of the England and Wales Census 2021 frame also provides an indication of how UK general population sampling frame quality can be assessed. Nonetheless, it is important to note that the census evaluation did not compare Addressbase and administrative data as alternative frames. Instead, it compared Addressbase enhanced with additional administrative data with administrative data alone, without the intention of using the latter as a frame. In the context of general population surveys, there is a need to identify the best available sampling frame by comparing the quality of the available alternatives. The census evaluation suggests some suitable measures of quality: percentages of under- and over-coverage across geographical areas and address types. Without access to a range of administrative data sources, the most recent aggregate population and household estimates published by ONS would be appropriate comparators.

Looking beyond Census 2021, there is no single widely used framework to assess and compare the quality of general population sampling frames. This is in part due to the variation in available general population sampling frames between countries and lack thereof over time in the UK. Prior to the introduction of Addressbase, PAF was firmly established over decades as the best available general population sampling frame and therefore assessment work focused on identification of specific subgroups rather than wider quality assessment, for example Bramley et al. (2018). Such work is a valuable element of the wider quality assessment picture, as systematic under- or over-representation of particular groups would undermine the overall quality of a sampling frame. Indeed, this was the reason for the switch away from using the UK electoral register as a sampling frame (Lynn and Taylor 1995).

A chapter by Junes on sampling frame errors in European Union Statistics on Income and Living Conditions (Junes 2022) compares the quality of sampling frames across EU countries. Although the UK is no longer a member of the EU, the framework for comparison is deliberately designed for international application and therefore remains relevant. Junes comments that there were no harmonised indicators of sample frame quality across EU states at the time of writing. The indicators subsequently identified for comparing sampling frame quality are:

- 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.

Junes concludes that information about the quality of sampling frames used is ‘rather limited’ and, ‘Without more detailed information about sampling frame coverage, we cannot really conclude whether or not frame under-coverage makes a significant contribution to estimation bias. The possibility certainly cannot be ruled out’ (p69, Lynn and Lyberg 2022). Respondent-driven sampling is suggested as a means of quantifying the under-representation of hard-to-reach groups.

In the UK context, PAF and Addressbase are both continuously updated, although Addressbase updates are released every six weeks (ONS 2020) and updates to PAF small users quarterly. When sampling general population surveys from PAF, additional data is not available to identify residential addresses. By contrast, Addressbase includes information from the National Land and Property Gazetteers and Ordnance Survey maps. The material already discussed suggests that over-coverage is more prevalent than under-coverage in PAF and Addressbase due to their nature as databases of addresses rather than of residence. This is an important fifth indicator to add to Junes’ list, as over-coverage is unlikely to be randomly distributed, just as under-coverage is not. It is therefore another potential source of bias within samples, in addition to increasing the sample sizes required in order to meet a particular achieved sample target. In web surveys where ineligible addresses are not distinguishable from non-respondents, over-coverage is challenging to control for when weighting responses. Using non-response modelling to weight the responding sample back to the profile of the sampling frame would not control for this, however calibration to population estimates of individual characteristics could mitigate it. That said, bias due to under-coverage is of greater concern as this cannot be controlled for at all. Consequently, the difficulty of sampling residential institutions and resulting under-coverage of populations living in schools, army barracks, prisons, and similar locations in PAF and Addressbase has led to the ‘general population’ being defined as persons living in private households.

It is worth noting that precisely measuring the scale of both under- and over-coverage presents significant challenges. Subgroups that are disproportionately under-represented can readily be hypothesised based on the characteristics that would render them less likely to be included in the frame. For example, both PAF and Addressbase do not immediately include every new residential property constructed, and also lag in removing demolished addresses. A frame based on administrative data collected for the purposes of a particular public sector function, for example administering tax, will not include residents who do not engage with that function, for example those too young to directly pay tax. In both situations, another source of data is needed to compare with the frame and estimate the scale of under-coverage. By comparison, identifying the information sources and update frequency of a sampling frame from its documentation is generally much more straightforward.

To estimate over-coverage levels, collating fieldwork outcomes across major general population surveys that currently sample from PAF or Addressbase and use face-to-face interviewing would also provide valuable indicators. Such surveys customarily report overall levels of ineligible and deadwood addresses with their response rate. As well as comparison between surveys, it would

be very helpful if these metrics could identify variations in sampling frame over-coverage between UK nations, regions, and urban-rural categories, as well as any changes over time. However this would require accurate and disaggregated fieldwork outcome information to be available for a wide range of general population surveys, which raises co-ordination challenges. At present, deadwood proportions are generally published in technical reports at an aggregate level, as shown in table 1.

Table 1: Recent deadwood rates for UK face-to-face general population surveys

| Survey Name | Nations included | Year of fieldwork | Proportion of deadwood addresses | Source |
|------------------------------------|------------------|-------------------|----------------------------------|--|
| Crime Survey for England and Wales | England & Wales | 2022 | 8.2% | ONS (2024) |
| Crime Survey for England and Wales | England & Wales | 2024 | 8.7% | ONS (2025) |
| English Housing Survey | England | 2023 | 5.6% | Ministry of Housing, Communities, & Local Gov (2025) |
| English Housing Survey | England | 2022 | 5.8% | Ministry of Housing, Communities, & Local Gov (2024) |
| Health Survey for England | England | 2022 | 6.50 | NHS England (2024) |
| Living Costs and Food Survey | UK | 2022 | 6.1% | ONS (2025) |
| Living Costs and Food Survey | UK | 2023 | 7.7% | ONS (2025) |
| Living Costs and Food Survey | UK | 2024 | 10.0% | ONS (2025) |
| National Survey for Wales | Wales | 2023 | 10.6% | Jenkins & Forbes (2024) |
| National Travel Survey | England | 2022 | 7.2% | DfT (2023) |
| National Travel Survey | England | 2023 | 7.7% | DfT (2024) |
| Scottish Crime and Justice Survey | Scotland | 2023/4 | 7.8% | Bell et al. (2025) |
| Scottish Health Survey | Scotland | 2022 | 10.0% | Scot Gov (2023) |
| Scottish Health Survey | Scotland | 2023 | 9.3% | ScotCen (2024c) |

Deadwood rates published for recent surveys vary from 5.6 to 10.6% of sampled addresses. The limited data shown may suggest that deadwood rates in Scotland and Wales could be higher than in England. However, more evidence would be needed to confirm this.

The key point regarding sample frame evaluation is that it cannot occur in isolation – the England and Wales Census 2021 frame evaluation demonstrates that additional data sources are essential. Census data and regularly updated mid-year population estimates are vital reference points for the evaluation of UK general population sampling frames, as well as historical fieldwork outcomes.

6. Administrative Data Linkage

The England and Wales Census 2021 frame provides the largest example of administrative data linkage with Addressbase, but there are also examples of smaller scale surveys using data linkage as well as work with wider methodological relevance to the use of administrative data as an enhancement to PAF or Addressbase sampling frames. While the purpose of matching administrative data to PAF or Addressbase may be outside the scope of this literature review, the methods used offer lessons relevant to general population sampling frames. No research specifically matching administrative data onto PAF or Addressbase in order to create a general population frame for survey sampling was found in the literature search. Early work investigating the potential for replacing regular censuses with administrative data is relevant here; a summary of the research to date is linked from ONS (2023b). This research includes feasibility studies on applying a dynamic population model (DPM) to administrative data to generate population estimates.

Three broad categories of relevant data-matching research were identified: matching administrative data onto survey responses for validation or enhancement, matching onto PAF to boost subgroups of interest, and matching to produce a database for population estimation.

6.1. Matching administrative data onto survey responses

The first category is of least direct relevance to sampling frames. An exemplar is Whiffen et al (2020), who matched clinical data with Welsh Health Survey (WHS) responses in order to compare estimated rates of several chronic health conditions between the two data sources. The paper notes that validation of health survey estimates in this manner is unusual. The WHS sample was drawn from PAF and clinical data came from multiple sources, including GP records, hospital admissions, and the Welsh Demographics Service Dataset (WDSD). The authors report that of the 11,694 2013 and 2014 respondents who consented to data linkage, 96.8% were matched with medical records including both good and ‘fuzzy’ matches. Similar work by Tingay et al. (2019) produced linkage rates of 97.6% for Scotland and 99.9% for Wales, but rates varied considerably between health data sources from 65.3% to 99.6%. By contrast Ernsten et al. (2018) matched postcodes from Scottish Longitudinal Study (SLS) participants aged 16-64 sampled from Census 2011 with NHS data, producing a postcode match rate of just under 90%. A further 9% of SLS participants could be traced within NHS records, but their Census postcode and NHS record postcode did not match; the remaining 1% could not be found within NHS records. Two key points are worth noting. First, that even with clean survey response data (name, date of birth, etc) in addition to PAF addresses, not all cases could be matched with health records. Second, that such linkage of admin data to survey data requires

respondent consent that is not always forthcoming. The challenges of data linkage will be correspondingly greater at the sampling stage.

The ONS has similarly matched census data with survey responses in the Nonresponse link studies. Here the census acts like an administrative dataset, providing additional variables which are then used in an assessment of nonrespondent characteristics. These can be used in a nonresponse adjustment. There were regular publications of these analyses from census up to 2001. The most accessible analysis for the 2011 study is Moore et al. (2018); in 2021 data were collected, but the analysis was not undertaken due to other COVID-related priorities.

6.2. Matching administrative data onto PAF for subgroup identification

In the second category, using administrative data for subgroup identification, a good example is the Scottish Health Survey (SHeS), an annual general population survey which uses health data linkage at the sampling stage to screen addresses for a child boost sample. Community Health Index (CHI) records are matched onto an address sample drawn from PAF in order to identify addresses with a child resident. SHeS 2023 sample sizes were calculated on the basis of 64% accuracy in the CHI record matching (ScotCen 2024c), in other words 64% of addresses flagged in CHI records as having children resident were assumed to actually have them. Subsequent fieldwork monitoring demonstrated that the actual accuracy was notably higher, around 84% for the 2023 child boost (ScotCen 2024b). This screening approach enables the target number of child interviews to be achieved much more efficiently than the prior approach, of drawing a larger PAF sample and conducting interviewer screening for resident children on the doorstep. While the accuracy rate found during 2023 fieldwork is cautiously encouraging, it cannot be used as a proxy for actual data matching accuracy. It is probable that some PAF addresses were flagged as having children resident then this was confirmed in fieldwork despite the matched CHI record not being that of the actual resident child. This is because some addresses are more likely to have children resident than others, varying depending on the building and location characteristics. In the context of a child boost, it doesn't matter if the child flagged as resident in CHI records is not resident at the address any longer as long as another child is living there instead (equivalently, the linked administrative data only needs to be a good *predictor* of the characteristic(s) of interest to be useful, it doesn't need to be *correct*).

Similarly, during sampling for the fourth National Survey of Sexual Attitudes and Lifestyles (Natsal-4), the potential for using NHS health records to screen for a young person (age 16-29) boost sample was trialled in Scotland and Wales. The match rate for PAF addresses and health records was found to vary by property type. Addresses containing 'flat' had a 52% match rate with health records, while addresses with a number then street had an 84% match rate (NatCen 2023a). Note that this doesn't necessarily mean the correct person's record was matched, just that the health database contained at least one match for the address. Where interviewers made contact with someone at the address, this could be checked. Non-trivial levels of both false positives and false negatives were found during the trial, suggesting that using health records to pre-screen the sample could drop potentially eligible cases as well as retaining some ineligible cases. This approach was therefore not adopted for Natsal-4 fieldwork overall.

The Northern Ireland Cohort for the Longitudinal Study of Ageing (NICOLA) project used health records in a similar way. A stratified random sample of private households with residents aged 50 or over was drawn from the Business Services Organisation (BSO) General Practitioner Register Database, then merged with POINTER address records, a similar database to PAF (Neville et al. 2023). A comparable exercise in Scotland was proposed by Douglas, Rutherford, and Bell (2018) to screen PAF addresses for residents aged 50+ using health records, in order to provide the pilot sample for a proposed longitudinal study of ageing in Scotland (Healthy Ageing in Scotland, or HAGIS). The pilot report (Douglas et al. 2018) described the linkage process as ‘long and involved’ in practise, as this marked the first time it had been attempted.

The SHeS example demonstrates that such an approach can work smoothly once initial data-sharing protocols and processes have been established and has the potential to make fieldwork more efficient. HAGIS demonstrates the potential for a health data-matching approach to be extended to a general population sample in order to facilitate stratification by age groups. For example, it could be used to flag addresses with residents of working or retirement age.

When such broad categories are used, exact matching to individual CHI records is not essential and confidentiality implications are minimised. The question is whether this could improve representativeness compared with PAF stratification using granular Census data, which captures age distribution at Output Area (OA) level. Survey response is known to vary by age groups and age is also a key analysis and reporting variable for many surveys. Being able to screen a PAF sample for broad age categories would in theory enable differential response rates to be reflected in the size of the issued sample. However, it cannot be assumed that the success of the SHeS child boost would be generalisable to other age groups or characteristics, as the Natsal-4 example suggests. The overall accuracy of the demographic and address data in Scotland’s CHI records, and comparable records in the other UK nations, are a big unknown.

6.3. Matching administrative data onto address databases for population estimation

The third category of relevant data-matching studies is the broadest and most frequently found in literature searches: matching UK administrative datasets onto PAF and/or Addressbase for population estimation. The most ambitious of the projects identified is explained in Lansley et al (2019). The aim was to create a ‘linked consumer register’ rather than a survey sampling frame, however the requirements appear similar: ‘an annually updated linked database of the residences of the individuals and households that make up the entire UK adult population’ (Lansley et al. 2019). The initial building blocks of this database 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 matched 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. A two-stage process was used: first, construction of a common address spine followed by attribution of household composition to each address. Cleaning of address fields in the commercial data sources was required to prepare for matching to PAF and Addressbase. Address matching then employed three procedures: rule-based matching, occupier-based matching, and fuzzy matching. The end result was a list of slightly more than 32 million UK addresses estimated to be private residences between 1997 and 2016. The paper acknowledges that this is likely an over-estimate overall and may exclude residences not included in PAF or Addressbase.

The potential usefulness of such a common address spine of residential addresses for survey sampling is considerable, as long as it performed better on the evaluation metrics discussed in the previous section than PAF or Addressbase alone. Lansley et al. (2019) report better match rates to PAF and Addressbase for Land Registry data than electoral registers or commercial databases. 87.5% of addresses in the electoral and commercial datasets could be matched to Addressbase, while an additional 2.6% could be matched to PAF but not Addressbase. 9.1% could only be matched to other electoral or commercial data, while the remaining 1% could not be matched to another data source. On the other hand, 99.8% of property sales data could be matched to Addressbase. As the National Land and Property Gazetteers data is used in the construction of Addressbase, this high figure is not surprising. For sampling purposes, the question is whether the addition of Land Registry data enables residential addresses to be accurately flagged as this data source will not provide up-to-date details of all residents across the whole UK.

The second stage of matching residents' names to the address spine, raised still greater challenges due to inconsistent spelling, inclusion of initials rather than full names, and duplication between data sources. As with address matching, a variety of technical approaches were used to try and deal with this, including imputation of missing records where gaps in an individual's apparent residence at an address were filled in by data for adjacent time periods. Anonymous residents were also imputed for vacant dwellings and new builds, based on property sales data. Of interest for validation purposes is a comparison between the number of individual records finally arrived at for each year, which was compared with ONS mid-year population estimates (MYEs). The mean difference between the linked consumer register total and MYE across the twenty years was 1.8% (Lansley et al. 2019), although in the latter years the register over-estimated the population relative to MYEs. However, the correlation coefficient at district level between the two estimates was over 0.99. While encouraging, this level of similarity does not necessarily indicate that the name and demographic records in the linked register would be sufficiently accurate to draw a named sample from. Indeed, the author's stated potential uses of the register are the production of granular and timely population estimates to supplement the Census and MYEs rather than for sampling. This work offers very valuable information about the technical challenges of data-matching. However, the approach requires further and more detailed validation. 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 the register's accuracy.

Other examples of research using similar data-matching methodologies but focusing on identification and estimation of particular subgroups were found in literature searches, but being less comprehensive in aims are of lesser relevance. Van Dijk et al. (2021) used the linked consumer register developed by Lansley et al. (2019) to estimate residential moves in the UK. Santos et al. (2021) matched Addressbase with Care Quality Commission and NHS data in order to identify care home residents in England. Health records appear the most frequent form of administrative data to be used for subgroup identification and survey enhancement or validation, potentially reflecting the importance of survey research in the health sector in addition to the more established data-sharing protocols. However there are examples in other areas, such as justice in Johnson et al. (2020), social care in Allnatt et al. (2022), and education in Morgan et al. (2020). All three papers stress that data linkage remains relatively novel and under-explored.

Common themes relevant to general population survey sampling frames emerging from research into data linkage are threefold. Firstly, gaining access to administrative data can be a lengthy and complex process, especially for a new purpose, and can vary between nations and public bodies. Indeed, Mansfield et al. (2020) review five different models for child or adolescent data linkage in the UK, recognising that agreement to access administrative data can be a critical barrier to projects. Secondly, 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. Thirdly, data matching is nonetheless receiving increasing attention as a means of using existing administrative data for social research purposes. Methods are clearly advancing and public sector bodies are willing to share data once robust agreements are in place. A systematic literature review on automatic identification of addresses (Cruz et al. 2022) finds increasing use of deep learning, offering improvements on previously established matching methods. There are potential opportunities for general population surveys to apply these lessons when sampling from PAF or Addressbase.

These themes are echoed by review papers on data linkage topics, such as Harron et al. (2020) which provides guidance for assessing data linkage quality in cohort studies. Their work concludes that data linkage strategies should be tailored to the research design, as much information about the linkage process as possible should be retained, and linkage error rates should be estimated. A rapid evidence review of barriers to data linkage for population health research by Moorthie et al. (2022) synthesised ten types of barriers from 28 articles. These were: funding, fragmentation, legal and ethical frameworks, cultural issues, geographical boundaries, technical capability, capacity, data quality, security, and patient/public trust. While some of these barriers will have more salience than others in the survey sampling context, all are worthy of consideration and provide insight into why there are not more existing examples of administrative data linkage to PAF or Addressbase for survey sampling. The benefits of data linkage may appear less significant than the barriers, while PAF is considered a 'good enough' sampling frame.

7. Sampling from Administrative Data

The difficulties of matching administrative data to PAF or Addressbase could be avoided were an administrative database available for use as a general population sampling frame. There are many examples of UK surveys being sampled from administrative datasets, such as the 2021-22 England and Wales Student Income and Expenditure Survey (SIES). Higher education students were sampled from Higher Education Statistics Agency (HESA) and Individualised Learner Record (ILR) databases (NatCen 2023b). Such administrative frames are typically used to sample the specific subgroup(s) that the database exists to track, for surveys of specific relevance to that subgroup. As previously discussed, the UK does not have an administrative database that is intended as a general population register covering all residents. Data is therefore fragmented across different administrative functions. Nonetheless, the data linkage literature discussed above demonstrates that health records have particularly high coverage of the UK general population, the vast majority of whom will have contact with the National Health Service at some point. Indeed, the number of health records often exceeds the resident population size. The question is whether NHS records could function as a general population survey sampling frame.

The best available evidence to try and answer this question is a recent health survey covering the Lothian Health Board area in Scotland (NatCen 2024). This general population survey of adults aged 16 and above was sampled directly from Community Health Index (CHI) records, with no data matching to other sources. For sampling purposes, the frame was anonymised and only includes the agreed sampling stratification variables: age, sex, Local Authority, and deprivation percentiles. Prior to sharing the frame, it was screened to remove cases missing name, address, sex, or date of birth, as well as duplicate records, records for those assessed as having no fixed abode, and those identified as deceased using National Records of Scotland (NRS) data. Once the sample had been drawn, the frame was deleted and name and address data was provided for sampled cases only. Unique identifiers in the sample were scrambled so that the sample could not be matched back onto the frame.

The sample stratification variables provided in the frame were used to evaluate the coverage and quality of CHI records. The number of CHI records in the frame was considerably higher than both 2021 MYEs and 2022 Census estimates of the 16+ population of the four LAs in the Lothian area, an overall difference of 17% from the Census estimate (NatCen 2024). The percentage difference between CHI records and MYEs was greater for City of Edinburgh than the other four LAs, and greater for individuals aged 16-24 than over-25s. Within the sampling strata of age group within LA, the largest difference between CHI records and MYEs was 16–24-year-olds in the City of Edinburgh. More than a third more individuals in this subgroup existed in the sampling frame of CHI records than ONS population estimates. There are multiple potential explanations for this divergence, notably the large number of students and tourists who stay in Edinburgh long enough to use health services and therefore receive a CHI record. Once these individuals leave the city, their CHI record remains in the database but they are no longer considered a resident. For the purposes of sampling, the excess CHI records above 2021 MYEs were assumed to approximate the rate of ineligible cases in the frame for each sampling

stratum. Crucially, anonymisation of the frame meant there was no means of identifying which CHI records could be ineligible.

The 2023 Lothian Health Survey was run by push-to-web mode, with invitations issued by letter asking the named individual to complete an online questionnaire. Consequently, the true rate of ineligible cases in the issued sample could not be identified. The percentage of letters returned as undeliverable (2.6%) or as named person not at the address (5.2%) is a clue to the accuracy of the CHI database (ScotCen 2024a). However, the former returns could indicate inaccurate data entry or addresses that no longer exist; the latter could indicate outdated and 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. These were not removed and responses were received from some such addresses, albeit at a much lower rate than the rest of the sample. When weighting the responses, Scotland's Census 2022 data was used to adjust the response profile as this was judged to be a more accurate population estimate than the sampling frame profile.

There are a number of key lessons from this innovative use of CHI records as a general population sampling frame. When using a novel frame like this, the sampling design must incorporate outcomes of a quality evaluation. Comparison of stratification variables with MYEs was a vital starting point, but future projects should also consider requesting tables of variables not used for sample stratification but important for analysis of survey outcomes. This would allow more granular identification of under-representation or potential ineligibility rates. In future projects, direct access to the sampling frame need not be required, as long as the data controller is willing and able to draw the sample to an agreed specification then share confirmation they have done so (e.g. code used for sampling). Only sharing records for the sampled cases also reduces the complexity of anonymisation. Future use of CHI records for sampling would need to incorporate further exploration of how to identify and screen out duplicate or ineligible cases. Without a robust method for identifying and reducing the ineligibility rate, sampling from CHI records risks introducing bias that cannot be controlled for with weighting. Using selection and non-response weights that adjust the response profile to match the sampling frame profile is unhelpful when the frame is known to be significantly biased.

Ineligibility rates notwithstanding, a CHI records sampling frame offers several theoretical advantages over PAF or Addressbase. A named sample can be drawn, removing the need for random within-household selection that increases the design effect and is impossible to enforce in push-to-web surveys. The sample can be explicitly stratified by age, in the case of Lothian Health Survey incorporating a lower assumed response rate for 16-24-year-olds (NatCen 2024). If a process were established to screen, clean, and validate names, addresses, age, and sex in CHI records, it has the potential to be an immensely valuable general population sampling frame for Scotland. CHI records include institutional residents who are very difficult to access in PAF surveys, as well as basic demographic data that would facilitate the selection of representative samples and the weighting of responses to control for non-response bias. Amid this great potential, it is important not to lose sight of the fact that CHI records (and

administrative data in general) were not collected for the purposes of sampling and any project using them as such must have a strong justification for doing so in order to comply with the UK Data Protection Regulation (GDPR).

As the Lothian Health Survey 2023 only covered four LAs in Scotland, the conclusions regarding the suitability of CHI records for general population survey sampling cannot necessarily be generalised further. The difference in possible eligibility rates between city and rural areas may also exist elsewhere in the UK, but this would need to be confirmed. There is no reason to assume that health care records in England, Wales, or Northern Ireland are of higher quality than Scotland's, though. Further projects of this kind, ideally incorporating additional sample quality evaluation and validation, are needed to determine whether administrative records are suitable sampling frames for a general population survey. At present there are many unknowns.

8. Conclusions

This review has sought to summarise the current status of sampling frame options for UK general population surveys and identify the key questions for the future. UK general population surveys have relied upon the Postcode Address File as a sampling frame for decades and therefore well-established sampling and weighting methodologies exist to minimise the impact of its limitations. Spurred in part by falling survey response rates over the last five years, researchers have started to look at innovative approaches to enhancing or even replacing PAF. The England and Wales 2021 Census used Addressbase as a frame and data-matching methodologies and acceptance are both advancing. Yet Addressbase has yet to match the popularity of PAF as a survey sampling frame and administrative data use remains almost entirely confined to subgroup boosts or subgroup-specific surveys.

At present decision-makers cannot have confidence that alternatives to PAF would offer benefits without incurring significant costs, introducing bias, or breaking time series. To change this, a robust and practical framework for evaluating sampling frame quality is required to compare options. The outline of this can be found in section 5 of this review, but could only be applied where evidence is available to measure quality indicators. Cost and time are potentially significant barriers to this quality evaluation research, which would have to be conducted by the organisations holding administrative data rather than organisations conducting general population surveys. ONS could take an important role in supporting this, having conducted the England and Wales Census 2021 frame evaluation. In the short term, matching administrative data onto PAF or Addressbase offers incremental sample design advantages without requiring wholesale quality assessment of administrative databases. Ultimately, though, the accuracy of administrative data determines its usefulness for any purpose, including sampling. Detailed quality evaluations are the most important research gap identified and until they are available it is impossible say whether administrative data could or should eventually replace PAF as a general population sampling frame.

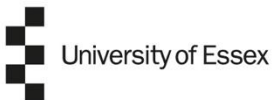
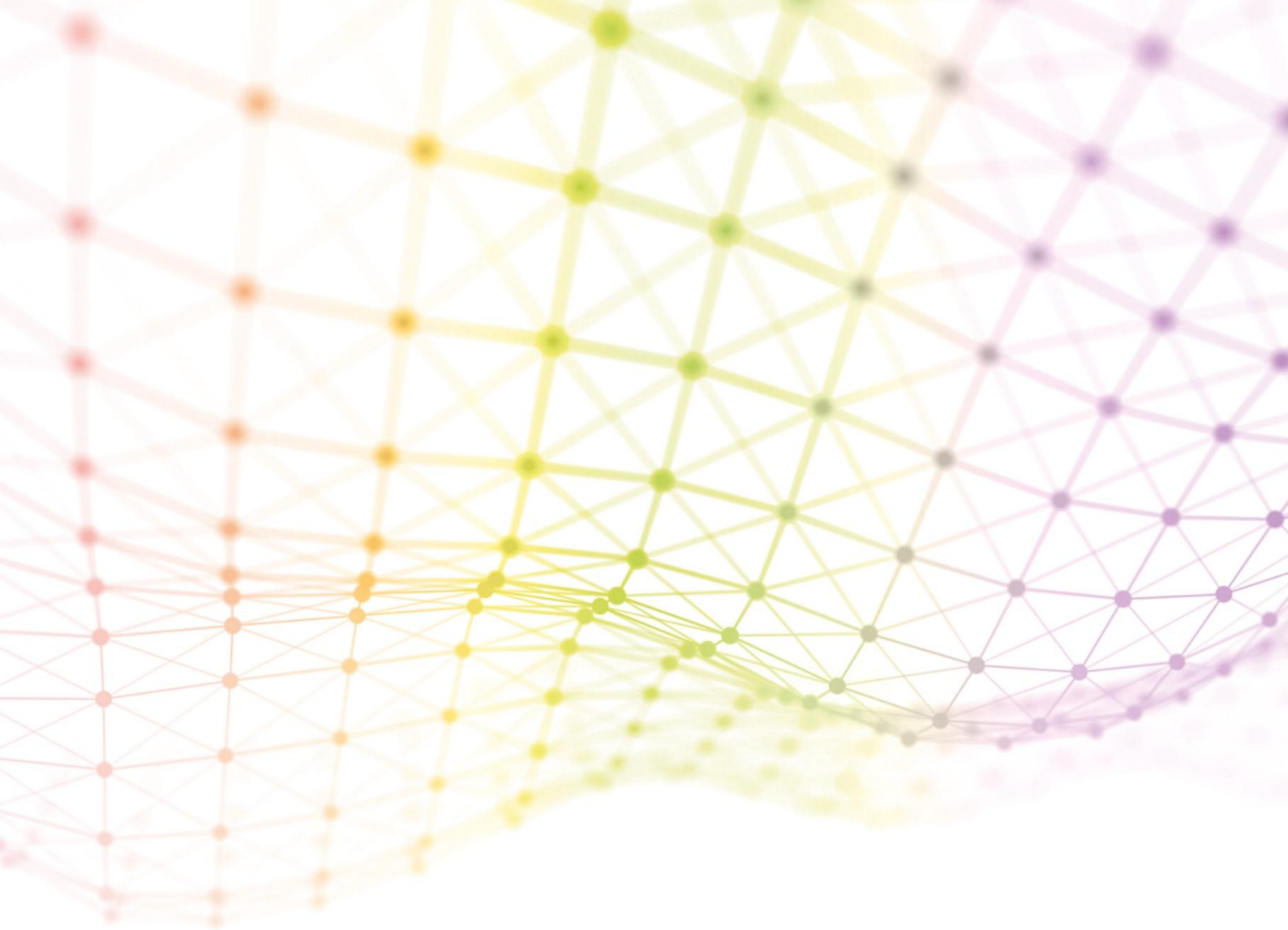
Bibliography

- Allnatt, G., M. Elliott, J. Scourfield, A. Lee, and L. J. Griffiths. 2022. "Use of Linked Administrative Children's Social Care Data for Research: A Scoping Review of Existing UK Studies." *The British Journal of Social Work* 52 (7): 3923–44. <https://doi.org/10.1093/bjsw/bcac049>.
- Bramley, Glen, Filip Sosenko, Jenny Wood, Joel Williams, and Peter Matthews. 2018. *Scoping Project to Investigate the Alternatives for Including Non-Household Populations in Estimates of Personal Well-Being and Destitution*.
- Butcher, R. 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>.
- Cruz, Paula, Leonardo Vanneschi, Marco Painho, and Paulo Rita. 2022. "Automatic Identification of Addresses: A Systematic Literature Review." *ISPRS International Journal of Geo-Information* 11 (1): 1. <https://doi.org/10.3390/ijgi11010011>.
- DfT. 2024. *Chapter 3 Fieldwork Procedures and Response Rate | National Travel Survey 2023 Technical Report*. https://department-for-transport.github.io/NTS_Technical_Report_2023/chapter3.html#ch3.16.
- Dijk, Justin T. van, Guy Lansley, and Paul A. Longley. 2021. "Using Linked Consumer Registers to Estimate Residential Moves in the United Kingdom." *Journal of the Royal Statistical Society Series A: Statistics in Society* 184 (4): 1452–74. <https://doi.org/10.1111/rssa.12713>.
- Douglas, E, T. Wilson, and D. Bell. 2018. *Healthy Ageing in Scotland - Pilot Survey Report*. [https://www.nuffieldfoundation.org/sites/default/files/files/Healthy%20Ageing%20in%20Scotland%20-%20Pilot%20Survey%20Report%20\(1\).pdf](https://www.nuffieldfoundation.org/sites/default/files/files/Healthy%20Ageing%20in%20Scotland%20-%20Pilot%20Survey%20Report%20(1).pdf).
- Douglas, Elaine, Alasdair Rutherford, and David Bell. 2018. "Pilot Study Protocol to Inform a Future Longitudinal Study of Ageing Using Linked Administrative Data: Healthy AGEing in Scotland (HAGIS)." *Public Health. BMJ Open* 8 (1): e018802. <https://doi.org/10.1136/bmjopen-2017-018802>.
- Ernsten, Annemarie, David McCollum, Zhiqiang Feng, Dawn Everington, and Zengyi Huang. 2018. "Using Linked Administrative and Census Data for Migration Research." *Population Studies* 72 (3): 357–67. <https://doi.org/10.1080/00324728.2018.1502463>.
- 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>.
- Harron, Katie, James C. Doidge, and Harvey Goldstein. 2020. "Assessing Data Linkage Quality in Cohort Studies." *Annals of Human Biology* 47 (2): 218–26. <https://doi.org/10.1080/03014460.2020.1742379>.
- Johnson, RD, DV Ford, K. Broadhurst, et al. 2020. "Data Resource: Population Level Family Justice Administrative Data with Opportunities for Data Linkage." *International Journal of Population Data Science* 5 (1): 1339. <https://doi.org/10.23889/ijpds.v5i1.1339>.

- Johnston, Neil. 2024. *Who Can Vote in UK Elections?*
<https://commonslibrary.parliament.uk/research-briefings/cbp-8985/>.
- Junes, Tara. 2022. "Frame errors in EU-SILC: under-coverage." Chapter 4 in Lynn, P. & Lyberg, L. (Ed.s), *Improving the Measurement of Poverty and Social Exclusion in Europe: Reducing Non-Sampling Errors*. Luxembourg: Publications Office of the European Union.
- 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, P, and L. Lyberg, eds. 2022. *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. 1997. "Sampling Frame Effects on the British Crime Survey." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 160 (2): 253–69.
<https://doi.org/10.1111/1467-985X.00060>.
- 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>.
- Mansfield, Karen Laura, John E. Gallacher, Miranda Mourby, and Mina Fazel. 2020. "Five Models for Child and Adolescent Data Linkage in the UK: A Review of Existing and Proposed Methods." *Clinical Review. BMJ Ment Health* 23 (1): 39–44.
<https://doi.org/10.1136/ebmental-2019-300140>.
- Maslovskaya, O., P. Lynn, L. Calderwood, et al. 2025. "Survey Futures Position Statement on Response Rates - the Expert Perspective." *Institute for Social and Economic Research (ISER)*. <https://www.iser.essex.ac.uk/research/news/2025/06/09/survey-futures-position-statement-on-response-rates-the-expert-perspective>.
- Moore, Jamie C., Gabriele B. Durrant, and Peter W. F. Smith. 2018. "Data Set Representativeness During Data Collection in Three UK Social Surveys: Generalizability and the Effects of Auxiliary Covariate Choice." *Journal of the Royal Statistical Society Series A: Statistics in Society* 181 (1): 229–48. <https://doi.org/10.1111/rssa.12256>.
- Moorthie, Sowmiya, Shabina Hayat, Yi Zhang, et al. 2022. "Rapid Systematic Review to Identify Key Barriers to Access, Linkage, and Use of Local Authority Administrative Data for Population Health Research, Practice, and Policy in the United Kingdom." *BMC Public Health* 22 (1): 1263. <https://doi.org/10.1186/s12889-022-13187-9>.
- Morgan, Kelly, Nicholas Page, Rachel Brown, et al. 2020. "Sources of Potential Bias When Combining Routine Data Linkage and a National Survey of Secondary School-Aged Children: A Record Linkage Study." *BMC Medical Research Methodology* 20 (1): 178.
<https://doi.org/10.1186/s12874-020-01064-1>.
- NatCen. 2020. *British Social Attitudes 2019 - User Guide*.
https://doc.ukdataservice.ac.uk/doc/8772/mrdoc/pdf/8772_bsa_2019_user_guide_v2.pdf.

- NatCen. 2022. *British Social Attitudes Survey 2021 User Guide*.
https://doc.ukdataservice.ac.uk/doc/9072/mrdoc/pdf/9072_bsa_2021_user_guide_v4.pdf.
- NatCen. 2023a. *Enhancing the Natsal-4 Sample with NHS Demographic Data: Findings for Scotland and Wales (Unpublished Internal Report)*.
- NatCen. 2023b. *Student Income and Expenditure Survey 2021 to 2022 Technical Report*.
https://assets.publishing.service.gov.uk/media/65674d08312f400013e5d5d3/Student_income_and_expenditure_survey_2021_to_2022_technical_report.pdf.
- 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>.
- 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>.
- NISRA. 2024. *Census 2021 General Report*. <https://www.nisra.gov.uk/publications/census-2021-general-report>.
- ONS. 2020. "Design for Census 2021."
<https://www.ons.gov.uk/census/planningforcensus2021/censusdesign/designforcensus2021>.
- ONS. 2022a. "Electoral Statistics, UK - Office for National Statistics."
<https://www.ons.gov.uk/peoplepopulationandcommunity/elections/electoralregistration/bulletins/electoralstatisticsforuk/december2021>.
- ONS. 2022b. "Population Estimates for the UK, England, Wales, Scotland and Northern Ireland - Office for National Statistics."
<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2021>.
- ONS. 2023a. "Evaluation of Addressing Quality: Census 2021 - Office for National Statistics."
<https://www.beta.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/evaluationofaddressingqualitycensus2021>.
- ONS. 2023b. "Research Outputs Using Administrative Data - Office for National Statistics."
<https://www.ons.gov.uk/aboutus/whatwedo/programmesandprojects/censusanddatacollectiontransformationprogramme/futureofpopulationandsocialstatistics/administrativedatacensusresearchoutputs>.
- Ordnance Survey. 2025. "AddressBase | Data Products | OS." Ordnance Survey.
<https://www.ordnancesurvey.co.uk/products/addressbase>.
- 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. 2025. "Multiple Residence Data - PAF Complementary Product." *Powered By PAF*.
<https://www.poweredbypaf.com/postcode-address-file/multiple-residence/>.

- Santos, Filipe, Stefano Conti, and Arne Wolters. 2021. "A Novel Method for Identifying Care Home Residents in England: A Validation Study." *International Journal of Population Data Science* 5 (4): 1666. <https://doi.org/10.23889/ijpds.v5i4.1666>.
- ScotCen. 2024a. *ScotCen Unpublished Internal Document Regarding Lothian Health Survey*.
- ScotCen. 2024b. *ScotCen Unpublished Internal Report Regarding the Scottish Health Survey 2023*.
- ScotCen. 2024c. "The Scottish Health Survey 2023 - Volume 2: Technical Report." <https://www.gov.scot/publications/scottish-health-survey-2023-volume-2-technical-report/>.
- Scotland's Census. 2025. "Enumeration." Scotland's Census. <https://www.scotlandscensus.gov.uk/about/2022-census/collection-approach/enumeration/>.
- 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>.
- 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>.



www.surveyfutures.net