SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

Survey Practice Guide 5: Using integrated non-survey data for evaluating and correcting for non-response

Thomas O'Toole[1], Alexandru Cernat[1], Nikos Tzavidis[2], Natalie Shlomo[1], Joseph W. Sakshaug[3], Peter Barlow[1]

[1]University of Manchester, [2]University of Southampton, [3]Institute for Employment Research (IAB) & LMU-Munich

February 2026

www.surveyfutures.net

Contents

Executive Summary 4
1. What is survey non-response? 11
  1.1. How does non-response occur? 11
  1.2. Consequences of non-response 11
2. How is non-response bias defined? 12
  2.1. Sampling and missingness 12
    2.1.1. Non-response bias 12
    2.1.2. Missing data mechanisms 14
3. Additional data sources for dealing with non-response 15
  3.1. Levels of linkage 15
  3.2. Cross-Sectional and Longitudinal Survey Data 15
  3.3. Sampling Frame Data Integration 16
  3.4. Paradata Integration 17
  3.5. Geospatial Data Integration 17
  3.6. Population-level Aggregate Benchmarks 18
  3.7. Administrative Record Integration 19
  3.8. The Challenge of Non-linkage 19
  3.9. Levels of linkage revisited 20
4. Evaluating non-response bias 20
  4.1. Covariate-based indicators 20
  4.2. Item-based indicators 22
5. Compensating for non-response bias 23
  5.1. Mitigating non-response bias via survey design 24
  5.2. Post-survey adjustments 24
  5.3. Imputation methods 27
6. Recommendations 30
  6.1. What to look for in external data? 30
  6.2. Create reproducible procedures 30
  6.3. Using informative sampling frames 30
  6.4. Using paradata in data collection 31
  6.5. Using adaptive and response designs 31
7. Case Studies 32

Evaluating the Utility of Linked Administrative Data for Nonresponse Bias Adjustment in a Piggyback Longitudinal Survey 32
Study purpose 32
Survey data 32
Non-survey data 32
Methods 33
Advantages 33
Limitations 33
Using linked Hospital Episode Statistics data to aid the handling of non-response and restore sample representativeness in the 1958 National Child Development Study . 34
Study purpose 34
Survey data 34
Non-survey data 34
Methods 34
Advantages 35
Limitations 35
References 36

Executive Summary

What is non-response bias?

- If the units that respond to a survey differ systematically from those that do not respond, survey estimates can be substantially different from the estimates that would have been obtained had all units responded. This difference is **non-response bias.**
    - The magnitude of non-response bias depends both on the response rate and the difference in characteristics between respondents and non-respondents.
- A low response rate does not necessarily indicate poor survey quality.

How can non-response be evaluated?

- To evaluate the nature and extent of non-response bias it is necessary to assess how different respondents are from non-respondents.
- This assessment requires the use of **auxiliary data** – data additional to that collected in the survey. This can come from the sampling frame, from the survey process (paradata) or can be data **integrated** from another source.
- Data can be integrated at **aggregate level**, allowing comparisons of the responding sample with the population, or at **unit level**, allowing comparisons of the complete (gross) and responding sample.
- Unit-level comparisons provide a direct assessment of non-response bias, while aggregate-level comparisons provide an estimate of the combined effects of non-response, sampling error and coverage error.
- Common approaches to evaluation of non-response bias involve estimating **response propensity** for each unit in the responding sample, i.e. the probability that they responded to the survey.

Non-response is the property of a survey; non-response bias is the property of a statistic.

Non-response bias occurs when the achieved (responding) sample is systematically different from the non-respondents.

- *If there is a relationship between response propensity and the survey variable(s).*

What is missing data?

Missing data in general (non-response, in the case of surveys) can be classified according to **Rubin (1976)** into three mechanisms.

Missing Completely at Random (MCAR)	Missing at Random (MAR)	Missing Not at Random (MNAR)
*The probability of non-response is **independent of any observed or unobserved data.***	*The probability of non-response is **explained by the observed data.***	*The probability of non-response is **explained by unobserved data.***

Through data integration, researchers can shift the mechanism of missingness from MNAR to MAR by integrating non-survey data that explains the causes of non-response.

Additional data sources for dealing with non-response bias

Sampling Frame Data

- The **Postcode Address File (PAF)** is widely used as a sampling frame for UK general population surveys.
- A wide range of small area aggregate data can be linked to PAF via either postcode or grid reference, for example Census or administrative data at lower super output area (LSOA) level, or the Index of Multiple Deprivation.
- **Other frames** – for example, DWP records of benefit recipients, or school records of pupils – may contain useful variables that should be integrated at the sample selection stage.

Paradata

- Survey paradata describes data about the survey process.
- It is usually collected for all units in the gross sample and can therefore be used for unit-level assessment of non-response.
- Data about the **contact and response process** is often used in non-response analysis. This can include indicators of the number and timing of call attempts, the elapsed time, or number of reminders sent, prior to response, etc.
- Another type of paradata used for non-response evaluation in the case of interviewer-administered surveys is **interviewer observation data** about the selected dwelling and neighbourhood.

Population-level Aggregate Benchmarks

- **UK census data** provided by the Office for National Statistics is commonly used for weighting and post-stratification of survey estimates.

- **Mid-year population estimates** are updated annually to adjust for births, deaths and migration patterns, among other factors. Distributions are published by age group, sex, region and ethnic group.

Administrative Records

- Administrative records are collected for routine and operational purposes, including:
    - **Health data (**e.g. *Hospital episode statistics*).
    - **Education data *(****Educational records*).
    - **Employment data** (*Employment spells*).
- Such records can be integrated with survey data at either unit or aggregate level, though unit-level linkage typically requires an access request to the data holder.

Geospatial Data

- **Geospatial characteristics** include spatial aggregates of Census or administrative data, but also **contextual geographical features** (e.g. population density, proximity to water or green space)
- Integrated survey and geospatial data are particularly useful when non-response is likely to be related to spatial features or to location itself.

Longitudinal Survey Data

- **Paradata from previous waves**, such as whether a cohort or panel member has responded at previous waves, or the amount of effort/persuasion required at previous waves, are often among the strongest predictors of current and future non-response**.**
- **Prior wave survey variables** are typically strong predictors of current wave variables and are therefore useful indicators of non-response bias if also associated with current wave response propensity.

The challenge of non-linkage
Non-linkage arises primarily from lack of consent, especially for sensitive unit-level data. Linkage errors also occur: deterministic methods are prone to missed matches, while probabilistic methods risk incorrect matches. Aggregate-level linkage avoids consent issues but provides less granular information.

Unit or aggregate level data?
The choice between unit-level and aggregate-level linkage depends on survey content and data availability. Aggregate-level linkage is only useful if comparable variables are available in the survey data, while unit-level linkage is only useful for non-response evaluation if it can be done for the complete (gross) sample.

Estimation of response propensity

Covariate-based indicators use the survey response rate and auxiliary data such as sampling frame data, paradata and integrated data.

Population total comparisons

Compares survey estimates of demographic characteristics to known population totals from external sources, such as censuses or administrative records. Requires only aggregate data.

Discrepancies between survey estimates and population totals can indicate coverage errors or non-response bias, prompting adjustments through poststratification.

Sub-group comparisons

Calculates response rates within specific demographic or other subgroups to identify patterns of non-response. Requires unit-level data.

By examining these rates, researchers can detect whether certain subgroups are underrepresented, which may indicate potential non-response bias.

R indicators

Measures the **representativeness** of the survey response by quantifying the variation in estimated response propensities across the sample.

Higher values indicate low variability.

It can be extended with distance measures to quantify the difference between the sample and the benchmark.

Coefficients of variation

Measures of the **relative variability** of the estimated response propensities in the responding sample.

Higher values indicate more variability.

It can be used to **standardise the variability** of a sample to **compare** across datasets.

These methods assume MAR conditional on the chosen auxiliary variables. Comparisons also depend on the accuracy and strict comparability of integrated population data.

Non-response weighting often results in a trade-off with increased variance.

Compensating for non-response bias

Weighting

Inverse probability weighting (IPW)

Models the **response propensity** – for example, via logistic regression – to estimate the probability of **each unit's participation.**

The **inverse of these probabilities** can be applied as non-response **weights.**

Post-stratification

Post-stratification adjustment consists of comparing survey statistics to **external benchmarks.**

Survey weights are calculated so that the weighted **sample aligns with known population** distributions across **specific, aggregate-level subgroups.**

Raking

Raking **(or iterative proportional fitting)** repeatedly adjusts sample weights so that **weighted marginal distributions match relevant population totals**.

Raking requires only knowledge of population **marginal distributions** but can yield highly variable weights.

Calibration

Calibration weighting follows a similar iterative proportional fitting approach but a variety of possible constraints can be imposed.

Constraints can incorporate multiple levels, for example households and individuals.

Imputation

Imputation involves **replacing missing data** with predicted values based on auxiliary data.

Single Imputation methods fill in missing values with a single estimate, such as:

- **Mean/Median Imputation:** Replacing missing values with the mean or median of observed responses.
- **Regression Imputation:** Predicting missing values using regression models based on auxiliary variables.
- **Nearest Neighbour:** Assigning missing values based on the closest observed data point in terms of similarity on a set of relevant variables.
- **Hot Deck Imputation:** Substituting missing values with observed responses from a random similar unit within the dataset.

Multiple Imputation (MI) creates multiple complete datasets, merging results to produce estimates that account for both within-and between-imputation variability:

- ***Multiple imputation via chained equations*:** Using regression modelling to draw a missing value from a random distribution and assumes each imputed variable is conditional on all other variables.
- ***Multiple imputation via predictive-mean matching*:** Instead of drawing an imputed value from a random distribution, it is possible to draw an observed value from a donor having a similar predictive mean.

Imputation and weighting models need careful specification and may be computationally intensive, especially with large datasets or numerous variables.

Datasets containing imputed values should not be thought of as observed data, but as a way of conducting statistical analyses that adjust for non-response biases. Inference should account for imputed values.

Recommendations

- **Explore alternative data sources that contain information that can explain why people do not participate in the survey. These can be sampling frame data, paradata, population-level aggregate benchmarks, geospatial data or administrative data.**

- **When selecting external data to integrate with survey data, look for consistency with survey data; variable definitions, time reference, geographical coverage and measurement properties.**

- **When considering data and variables to use to adjust for non-response bias, look for variables that explain the missing mechanisms and are also related to the topic of interest.**

- **Use the response rate alongside indicators of representativity (R-indicators) and variability (CVs) to evaluate non-response.**

- **When adjusting for non-response, start with a limited set of well-measured variables and document each step and diagnostics to ensure FAIR inference.**

1. What is survey non-response?

Over the past decade, the landscape of survey data has substantially changed. Response rates in probability-based surveys have seen a steady decline, with a particularly sharp drop-off in the wake of the COVID-19 pandemic. A low response rate does not necessarily indicate poor survey quality (Groves, 2006), but survey non-response alongside other metrics (e.g. sample representativeness and sample composition) serves as a gauge of possible problems with the sample (Maslovskya et al., 2025). This guide explores non-response bias in social surveys and the role of non-survey data integration in 1) understanding mechanisms of non-response bias, 2) evaluating the implications of non-response bias for data quality, and 3) identifying methods that can help to correct for non-response biases.

1.1. How does non-response occur?

Survey non-response can be divided into two types: unit non-response and item non-response. Unit non-response occurs when the unit of interest, typically an individual person or household, is invited to take part in a survey but refuses to participate or does not provide sufficient information for their responses to be useful (UK Data Service, 2025). Item non-response occurs when an individual takes part in a survey but does not complete all questions (or items). This guide will focus primarily on unit-level non-response, which is more relevant for understanding data integration applications.

Survey data can be collected from one of two sampling paradigms, probability and non-probability (Cornesse et al., 2020; Wiśniowski, Sakshaug, Perez Ruiz & Blom, 2020). Probability sampling draws units from the sampling frame (i.e., a list of units in the target population) using a randomised process and enables the calculation of the probability of being sampled for each case. Non-probability sampling involves a selection process, usually determined by cost and time, where the probability of inclusion is unknown. The focus of this guide is on probability-based surveys. While some of the issues and methods discussed can also apply to non-probability samples, this is often a more challenging process due to the lack of a sample frame that represents the population of interest.

Representation, in the context of social surveys, refers to the extent to which a responding survey sample accurately reflects the characteristics of the wider population the study is designed to capture (the target population), typically with respect to key demographic and socio-economic variables. Survey "representativeness" is a term which is used inconsistently in the literature. A representative survey should support inference and estimation at the population level by including all relevant groups from the target population via probability sampling approaches and ensuring appropriate heterogeneity in sub-groups with the net effect of minimising selection biases (Lynn, 2015).

1.2. Consequences of non-response

The implication of non-response bias on data quality is that those who respond to a survey may be systematically different on key survey variables when compared to those who do not respond and consequently may not accurately represent the target population (Groves & Peytcheva, 2008). For example, if some people are too ill to participate in a study, it will likely lead to positively biased health estimates. Another example is that people who are employed full-time may be more difficult to reach, and thus, estimates of employment may be biased downward. When non-respondents differ from respondents on characteristics related to survey outcomes, systematic non-response biases may be introduced into the survey sample, which presents a challenge to both the internal validity and generalisability of survey statistics. The consequences of non-response bias also extend further than sample composition and can lead to a reduced power to detect associations in statistical modelling, wider confidence intervals and increased uncertainty in estimates. The ramifications of non-response bias for reduced power will be discussed in detail in Section 5.

The integration of non-survey data sources with survey data can help to ameliorate the effects of non-response as the new variables may provide information that describes the patterns of non-response and facilitates remedial action. For example, linking survey data with administrative records has been found to provide insight regarding the subgroups within the sample that are prone to non-response; employment variables (such as wage and benefits receipt; Büttner, Sakshaug & Vicari, 2020) and administrative health variables (such as number of missed outpatient appointments and treatment for mental illness; Rajah et al., 2023). These approaches can help to buffer the effects of falling response rates in social surveys.

2. How is non-response bias defined?

2.1. Sampling and missingness

In probability-based social surveys, the two main types of unit non-response are refusal to take part and non-contact with a selected person or sample member. Refusal to take part refers to the case when individuals selected into the sample either neglect or actively choose not to take part in the survey. Non-contact with a sample member refers to the inability of the interviewer/survey to establish contact with the individual in question. In longitudinal surveys, these effects can accumulate across each sweep, as initially contacted and willing respondents drop out of the study (or attrit) over the course of multiple waves.

2.1.1. Non-response bias

The measurement of non-response in survey data traditionally focuses on response rate, which involves the use of a binary response indicator that indicates whether a sampled unit has responded or not (Wagner, 2012). This binary indicator is used to calculate the global rate of response as “the number of complete interviews with reporting units divided by the number of eligible reporting units in the sample” (AAPOR, 2023). The inverse of the response rate is the non-response rate. However, a high response rate does not necessarily imply a lack of non-response bias, nor does a low response rate indicate its presence (Schouten, Peytchev & Wagner, 2017; Groves & Peytcheva, 2008).

Formula for (non) response rate (adapted from AAPOR response rate 1, 2023):

$$RR = \frac{n_r}{n_e}$$

$$NRR = 1 - \frac{n_r}{n_e}$$

Note: n_r = complete interviews with reporting units, n_e = number of eligible reporting units in the sample frame

Response rates are not a reliable indicator of non-response bias (Maslovskaya et al, 2025). Moreover, the response rate is the property of a survey, whereas non-response bias is the property of a statistic (Wagner, 2012). For example, a survey with a 90% response rate may result in bias in the estimate of some statistics related to certain characteristics if the 10% who did not respond differ significantly in relevant characteristics from those who did. Conversely, a survey with a 30% response rate might yield unbiased estimates if the respondents' characteristics are aligned with those of the target population. It follows that the greater the disparity in the relevant characteristic between the responding sample and the sample frame, the greater the extent of the bias in the estimate associated with the sample.

Non-response bias is the property of a survey statistic, and so is statistic-specific. For example, in a Time Use Survey a lower response rate amongst men than women would lead to nonresponse bias in estimates of the amount of time spent doing anything that men do more (or less) of than women - e.g. watching sport, perhaps - but might not lead to bias in estimates of anything that both sexes do in equal amounts, e.g. sleeping. Equally, this bias is dependent on the method of estimation. Bias in the variable mean does not imply bias in the coefficient associated with a linear regression carried out on that variable or vice versa. The non-response bias of a mean value can be calculated by the non-response rate multiplied by the difference between the means for survey respondents and non-respondents (Groves, 2006).

Deterministic formula for non-response bias (Groves, 2006; Koch & Blom, 2016).

$$NRB(\bar{y}) = NRR * (\bar{y}_R - \bar{y}_{NR})$$

Note: $\bar{y}$ = respondent set mean, NR = non-response, NRR = non-response rate, NRB = non-response bias

The challenge in estimating non-response bias is that while the (non) response rate is known, the difference between respondents and non-respondents on a variable of interest is often unknown. According to Groves (2006), the sample is not deterministically divided into respondents and non-respondents as the likelihood of a respondent to participate is conditional on a variety of separate causes (e.g. discrete health events); common causes (e.g. general socio-economic factors); and survey-specific causes (e.g. length and topic) (Groves & Peytcheva, 2008). These characteristics may make a respondent's participation in the survey too burdensome and may lead to higher rates of non-response. As such, non-response bias cannot be observed directly, only estimated.

According to the stochastic formula, non-response bias is proportional to the covariance between the response propensity and the survey variable. The stronger the relationship between the response propensity and the variable of interest, the greater the bias in the statistic. For example, if a survey sought to ask questions about income and only carried out interviews during the day, they would be less likely to survey people who are employed and working during the day. Therefore, these would be strongly related, and significant bias may arise. On the other hand, the same survey may yield less bias in its estimate of the proportion of people who eat meat.

Stochastic formula for non-response bias (Groves, 2006; Bethlehem, 1988; Bethlehem, 2002).

$$NRB(\bar{y}) = \frac{\sigma_{(y,p)}}{\bar{p}}$$

2.1.2. Missing data mechanisms

Missing data in general can be classified according to Rubin (1976) into three 'mechanisms' which influence both their potential impact on estimates and methods that can be used to compensate for them:

Missing Completely at Random (MCAR): The probability of missing data is independent of any observed or unobserved data. For example, paper surveys are lost in the postal system, power outages occur during online survey completion, or participants accidentally click "next" and skip survey questions.

Missing at Random (MAR): The probability of missing data is explained by the observed data. For example, those resident in certain geographical regions may be less likely to respond to a survey (i.e., London), and location data **are available** via the sampling frame (i.e., the PAF). For a longitudinal example, those with poorer health are more likely to attrit in longitudinal panel surveys, and health conditions **are recorded** in the survey data. It is usually possible to control for MAR based on the observed data at the researcher's disposal.

Missing Not at Random (MNAR): The probability of missing data is explained, at least in part, by the unobserved data. For example, older people are less likely to respond to a web survey, and age is **not recorded** in the sampling frame. For a longitudinal example, those with poorer health are more likely to attrit in longitudinal panel surveys, and health conditions **are not recorded** in the survey data. It is generally not possible to control for MNAR because the data required is unobserved.

In most situations, users do not know what kind of mechanisms are causing the missing data. If any of the variables collected in the survey predict non-response, it is possible to show that the data is not MCAR. However, it is impossible to rule out that the missing pattern remains MNAR, as we may be missing some of the causes of missingness.

A common approach to compensate for non-response is listwise deletion or complete-case analysis, which involves excluding any cases with missing values (Little, Carpenter & Lee, 2024). This method is straightforward and easy to implement, often being the default approach in many statistical software packages. However, complete-case analysis relies on the assumption that data are Missing Completely at Random (MCAR), meaning the likelihood of missingness is unrelated to any observed or unobserved data, and is not missing by design. Complete case analysis also necessarily results in a loss of statistical power and fewer degrees of freedom on which to estimate statistical models; therefore, it can be detrimental. In practice, complete-case analysis can lead to biased estimates if the proportion of missing data is not minimal and the MCAR assumption is not plausible (Mukka et al, 2016).

Through the integration of survey and non-survey data, researchers may be able to link additional explanatory variables with survey data, making the MAR assumption more plausible. For example, linking administrative data on health conditions with survey data may help to explain of the health-related component of response propensity. By incorporating non-survey data, researchers may be able to shift the mechanism of missingness from MNAR to MAR, making non-response and non-linkage more amenable to standard post-survey adjustment techniques.

3. Additional data sources for dealing with non-response

Measuring non-response involves understanding the degree of missingness as well as possible mechanisms or effects of it on our data. There are three components to consider in the evaluation of non-response in survey data: the amount of missingness, the differences between responders and non-responders on characteristics which can be observed for the entire sample (including auxiliary data), and the relationships between these fully observed covariates and survey outcomes of interest (Andridge & Little, 2008).

3.1. Levels of linkage

There are two approaches to using integrated (external to the survey) data to correct for survey non-response. One involves merging data to the (gross) sample at the unit-level (and involves subsequent unit-level modelling). The other consists of comparing the responding sample to aggregate external data and making group-level adjustments. Many additional sources of data for dealing with non-response can be used at both the individual and aggregate level, which this section will discuss.

3.2. Cross-Sectional and Longitudinal Survey Data

Surveys fall into two broad categories: cross-sectional and longitudinal. Cross-sectional surveys collect data from different respondents at a single point in time, giving a “snapshot” of the population. For example, the Crime Survey for England and Wales (Office for National Statistics, 2025) repeats this snapshot, each time with a new sample, to facilitate comparability of aggregate-level changes in victimisation levels and attitudes towards crime. Longitudinal surveys follow the same cases over multiple waves, tracking how they change over time. For example, Understanding Society (University of Essex, 2024) is a longitudinal household panel study (following members of the same household over time), and the 1970 British Cohort Study (University College London, 2025) is a birth cohort (following those born in a specific time period). Both cross-sectional and longitudinal surveys are frequently integrated with non-survey data sources, such as administrative records, geospatial characteristics and digital trace data. For more information on sources of integrated survey and non-survey data, see O'Toole, Cernat, Tzavidis, Shlomo & Sakshaug (2025).

Unlike cross-sectional surveys, longitudinal surveys may have previous-wave data for current-wave non-respondents. This is a rich source of information about current-wave non-response patterns. In longitudinal surveys, responses from previous waves can help inform dynamic questionnaire routing and skipping stable items while probing longitudinal changes. Prior wave variables can also serve as indicators for non-response weighting and predictors in multiple-imputation models (which will be addressed in section 4). However, prior wave data is, of course, not available for cross-sectional or first-wave longitudinal surveys.

No statement about the bias or representativeness of response is possible without some information about non-respondents and how they differ from respondents on key characteristics. For UK social surveys, this is problematic as there is typically no (or limited) information available for survey units that do not respond: the most-used sampling frame, the Postcode Address File, is largely uninformative, and there is no population register against which to benchmark. The integration of survey and non-survey data sources, such as auxiliary data or paradata, is often the only way to facilitate comparison between respondents and non-respondents and improve the evaluation and correction of non-response bias.

3.3. Sampling Frame Data Integration

The Postcode Address File (PAF) is the most commonly used sampling frame for surveys of the general UK population. The integration of sampling frame information with survey data can provide researchers with contextual information regarding non-respondents and facilitate the modelling of participation. As discussed in Section 2, estimation of response propensities is key to understanding the nature of non-response bias and to identifying suitable adjustment techniques.

Sampling frame information is generally used at the unit-level and is particularly relevant for cross-sectional or first-wave longitudinal studies, which cannot rely on previous wave comparisons. While the PAF has little auxiliary data, the addresses can be linked to other datasets useful in explaining the mechanism of non-response at postcode and other geographical levels. Most often, this linkage is done at the postcode sector level or the Lower Level Super Output Area (LSOA). This enables linkage with a wide variety of data from sources such as the ONS or the UK Data Archive.

While using this approach can augment survey data with a wide range of information, using aggregate data for non-response correction does come with some important assumptions. In addition to assuming MAR, this approach obscures the variation within geographical areas. For example, using a deprivation index aggregated at the LSOA level to explain why people do not participate in a survey ignores the fact that there can be important differences in income for people within the same area. Furthermore, individual and aggregate level mechanisms can also be different (also known as ecological fallacy). As such, researchers should aim to aggregate the data as little as possible when using this strategy.

3.4. Paradata Integration

Survey paradata is data about the survey process (Blom, 2008). While a lot of paradata may exist for respondents (for example, in the case of a web survey, device type, browser, screen size, time taken to answer each question, etc), to be useful for non-response adjustment, paradata must relate to the entire gross sample. Such paradata can include interviewer observations, contact data and other non-survey item characteristics like the location of the respondent, or whether they have a telephone number or an internet connection. Some paradata are available in public survey data files available via the UK Data Service, but depending on the variables required, researchers may have to contact the data providers directly for access to paradata. For example, the Centre for Longitudinal Studies, who house a number of the UK's leading birth cohorts, have a Data Access Committee (DAC) which can grant access to survey paradata and auxiliary data which is not traditionally deposited in data archives (Centre for Longitudinal Studies, 2025). However, it is also of note that the available paradata can depend on the mode of the survey. Surveys which use interviewers would theoretically have substantial paradata based on interviewer observations, while paradata for self-completion surveys may be much more limited.

The integration of contact data with survey data can be particularly useful in non-response and has two broad uses: to evaluate non-response bias and to monitor fieldwork progress. Contact data can be integrated at the unit-level, is available for both respondents and non-respondents, and is often related to both the survey process and the survey outcome (Kreuter & Kohler, 2009; Blom, 2008). Survey paradata is particularly useful for informing responsive and adaptive survey designs, in which paradata indicators, such as contact and refusal rates, are monitored throughout fieldwork, and resources are redistributed to balance response rates in a cost-effective manner (Groves & Heeringa, 2006).

3.5. Geospatial Data Integration

Geospatial characteristics are the characteristics associated with an individual's residential location or local environment, such as population concentration in the area, whether it is urban or rural or the level of deprivation in their area. Survey data can be linked to geospatial data at various spatial scales (government region, middle/lower super output area (M/LSOA), postcode, gridsquare) provided suitable geographical location variables are available for all survey units (e.g. grid reference or unit postcode). Geospatial data can also be used for aggregate benchmark comparisons. Integrating geospatial data with the survey sample or sampling frame can provide more detail on contextual geographical factors, and stratification variables like the Index of Multiple Deprivation (Wales, Scotland and Northern Ireland have distinct indices). For example, Understanding Society releases detailed geospatial indicators (under special licence) for households, facilitating linkage to external open and administrative sources (Understanding Society, 2025). The linkage of geo-spatial data has become increasingly common in the UK and Ireland in the absence of register data. The Generations and Gender Survey (Gauthier et al., 2025) and the Community Life Survey (Department for Digital, Culture, Media and Sport, 2024) are both examples of surveys which have incorporated geo-spatial data as part of their survey design.

3.6. Population-level Aggregate Benchmarks

External data can often be used as a measuring stick against which a sample's representativeness can be analysed. These population-level parameters not only illustrate the representativeness of the sample but also provide a basis for carrying out remedial action, such as applying weights to the survey. These data can come from any source (for example, administrative data), but they must cover the entire population. In the UK, however, with the unavailability of sources like a population register, the most commonly used data sources for benchmarking are administrative data (see section 3.7 below), population-level surveys like the Labour Force Survey (ONS, 2024) and the UK census.

Census data are often used for weighting and post-stratification of survey estimates and are available via the Office for National Statistics. By integrating population benchmark data with survey data at the aggregate level and aligning sample distributions on age, sex, ethnicity, and region to corresponding census marginals, researchers can correct both coverage and non-response errors in survey data. Census-derived variables such as household composition or population density also enrich survey microdata as covariates in regression or small-area estimation models. These calibrations ensure that survey inferences remain anchored to the most complete population counts available.

While population statistics are typically freely available, due to the devolved nature of the UK Census, data can be difficult to navigate and find, which can be a challenge to both validity and replicability. A useful access point for ONS population data is NOMIS, which contains ONS statistics related to population, society and the labour market at national, regional and local levels. A limitation of adjusting non-response to census totals is the length of time between censuses. The census is carried out every ten years in the United Kingdom, meaning that target population totals may not reflect the population of the UK at the time of survey data collection. The census mid-year population estimates for sex and age of the respondent are produced annually using demographic accounting techniques to adjust for births and deaths and a range of other factors (Office for National Statistics, 2025). Researchers should choose population data which is the closest to their survey data in time to benchmark their survey weights, for example, the 2021 Census would struggle to inform on the population structure in 2025. As such, population surveys such as the LFS are a relevant alternative, even if these surveys may suffer from non-response bias themselves. In comparing survey data to external aggregate-level distributions, one should also be aware of possible measurement and definitional differences between the data sources.

3.7. Administrative Record Integration

Administrative records are primarily collected for routine and operational purposes (for more information, see Survey practice guide: Data Integration; O'Toole, Cernat, Tzavidis, Shlomo & Sakshaug, 2025). The integration of administrative records with survey data is a valuable technique and option for researchers and practitioners in addressing survey non-response. Administrative data is often linked at the unit-level, and as such, may require respondent consent (which adds layers of selection for those who consent to linkage and successfully have records linked). Work by Burton et al. (2025) examined the decision-making process for linkage consent, finding that a productive strategy for promoting informed consent may be to encourage reflective decision-making based on trust in the relevant organisation/data holders, rather than a consequential "gut-reaction". Administrative data can alternatively be used for aggregate-level comparisons, as discussed in section 3.6.

Administrative data sources commonly linked to survey data in the UK are:

- **Health data**, for example, from NHS hospital episode statistics (outpatient, admitted patient care and accident & emergency), NHS cancer data and Office for National Statistics mortality records. Hospital episode statistics have been used via imputation methods to address non-response bias in survey data (Rajah et al., 2023). An example of this approach is provided as a Case Study in section 7. This data is generally very restricted and, as a result, is used rarely unless already available as part of the research design.

- **Education data**, for example, the national pupil database (pupil records in Scotland and Wales) and individualised learner records from the Department for Education. Educational records have also been used to evaluate non-response bias and data quality in survey data (Booth et al., 2024).

- **Employment and income data**, including Benefit receipt, tax credits from the Department for Work and Pensions and PAYE data from HM Revenue & Customs. For example, the Family Resources Survey has been integrated with HM Revenue and Customs data, and while their use in a UK context to correct for non-response bias is limited, there is promise for using employment spells as indicators for survey non-response (Büttner, Sakshaug & Vicari, 2021).

3.8. The Challenge of Non-linkage

In the context of data integration, non-linkage can be a challenge. The main cause for non-linkage is typically linkage non-consent, which disproportionately impacts access to sensitive data, such as health records (Sakshaug, 2022). In this context, the concern is that respondents who do not consent to record linkage may be systematically different from those who do. There is some evidence that these differences tend not to be as socially patterned as for unit non-response, bearing more closely to the respondent's level of trust in the linked data provider (Jäckle, Burton, Couper, Crossley & Walzenbach, 2021). Conversely, aggregate-level data linkage does not typically require respondent consent, and so sources like aggregate geographical or sociodemographic Census or administrative records can be a more pragmatic choice of data for evaluating and compensating for non-response bias (see O'Toole, Cernat, Tzavidis, Shlomo & Sakshaug (2025) for more information on accessibility in linked data).

Another mechanism of missingness in integrated data is linkage error, which is caused by missed and incorrect matches between survey and non-survey units. Linkage errors occur when the matching procedure fails to correctly link respondents between data sources. When the matching procedure is deterministic (i.e. via a unique identifier such as a national insurance number), missed matches are more common, whereas if the matching procedure is probabilistic (i.e. match propensity is estimated based on a range of characteristics), then incorrect matches are more likely to occur (Harron et al., 2017).

3.9. Levels of linkage revisited

The choice between whether to use individual or aggregate-level non-survey data is often pragmatic. Key questions to consider are what level of integration is available in the timeframe and budget of the project, and what measures are available in the survey data. If the survey data includes variables that are likely to be related to unit non-response and for which comparable measures are available from an external source, then aggregate-level comparisons may be informative and can provide a basis for adjustment. But if relevant variables available in external sources are not also included in the survey, then only unit-level linkage will enable their use in non-response evaluation and adjustment, and even then, only if the linkage can be carried out for the whole gross sample (which could be the case if, for example, the external source was used as the sampling frame).

4. Evaluating non-response bias

There have been various indicators proposed to examine non-response bias in survey data. Wagner (2012) outlines three types of indicators of non-response bias: the response rate (as previously mentioned), indicators that include only auxiliary variables, and indicators that include both auxiliary and survey variables. Each of these indicators is reliant on the auxiliary data used in section 3.

4.1. Covariate-based sample-level indicators

Covariate-based indicators use the response indicator and integrated auxiliary data, such as sampling frame data and paradata, and assume MAR. Covariate-based indicators of non-response are selected based on their theoretical or empirical associations with main survey variables, but do not use main survey variables in their computation. Examples of covariate-based indicators include the following:

Comparisons to population totals involve the aggregate-level integration and comparison of survey estimates of demographic characteristics to known population totals from external sources, such as censuses or administrative records. Discrepancies between survey estimates and population totals can indicate coverage errors or non-response bias, prompting adjustments through poststratification (Skalland, 2011). In the UK context, unit-level (individual) measures are often not available; for example, when using the PAF as a sampling frame, researchers are often restricted to aggregate population (and sub-group) totals in comparing sample distributions to population distributions.

Sub-group response rates involve calculating response rates within specific demographic or other subgroups (e.g. age, gender, strata, geographic location) to identify patterns of nonresponse, typically by comparison to the sample frame or linked unit-level data. By examining these rates, researchers can detect whether certain subgroups are underrepresented, which may indicate potential non-response bias; however, this approach is usually limited to univariate or bivariate settings, implying that it is difficult to analyse intersectional response rates where response rates may depend on a multitude of variables (Groves & Peytcheva, 2008). To calculate sub-group response rates, a useful approach is to weight the survey via design weights.

Modelled estimates of response propensities. Rather than simply calculating response rates based on one or two auxiliary variables, a larger variety of auxiliary variables can be used as predictors in a statistical model of response, to provide an estimated propensity for each sample unit. These estimates can be used directly in adjustment – for example in response propensity weights – or as inputs to summary measures such as R-indicators or CVs (see below).

R-indicators measure the representativeness of the survey response by quantifying the variation in response propensities across the sample. The R-indicator is defined as 1-2SD(p), where p is the response propensity. An R-indicator value close to 1 suggests high representativeness (low variability in response propensities), while values closer to 0 indicate potential nonresponse bias. R-indicators can be used for cross-sectional survey data, or for the first wave of a longitudinal survey if there is information known about the non-respondents, and can also be particularly useful for longitudinal data, due to the richness of previous wave survey data (Schouten, Cobben & Bethlehem, 2009; Plewis & Shlomo, 2017). However, it needs to be borne in mind that the R-indicator, as with any measure of non-response bias, depends on the quality of the auxiliary data. To be informative, the covariates should predict both response propensity and key survey measures.

Coefficients of variation (CVs) are an alternative statistical measure of the relative variability of the (estimated) response propensities in the responding sample, expressed as the ratio of the standard deviation of the response propensities to the mean propensity (typically expressed as a percentage, with a higher value indicating more variability), i.e. CV=SD(p)/$\overline{p}$. Similar to R-indicators, they allow for a comparison across datasets or waves of a survey (Moore, Durrant & Smith, 2018; Schouten & Shlomo, 2017). CVs and r-indicator should be used in tandem when assessing data quality.

The strengths of covariate-level indicators lie in their ability to provide a granular evaluation of representativeness, as they can highlight disparities in response patterns across different subgroups. They also benefit from the availability of complete auxiliary data for both respondents and non-respondents, facilitating comprehensive analyses without additional data collection.

However, these covariate-level indicators have limitations. They rely on the assumption that auxiliary variables will be strongly correlated with survey variables of interest, which may not always hold true. Additionally, the quality and granularity of auxiliary data can vary across data sources and over time. The indicators described above tend to provide similar but distinct estimates regarding the presence and size of non-response bias. As such, they should be used consistently and in tandem, and are of particular use in comparison across studies and over time, when the covariates can be held constant (Schouten, Peytchev & Wagner, 2017).

4.2. Item-based indicators

Item-based indicators differ from sample-level indicators as they take into account the observed survey data in addition to the response indicator, sampling frame data, and paradata. Consequently, they can indicate how the effects of non-response may differ between survey items. In surveys that collect a range of variables, it may be difficult to identify an optimal and consistent set of indicators from the survey data, and conclusions may differ depending on the survey variables included in indicator computation (Schouten, Peytchev & Wagner, 2017). Item-level indicators typically require unit-level data, and an explicit model for each variable, estimated using observed data, and rely on the assumption that the missing data are missing at random (MAR; Wagner, 2012).

Comparing the sample with follow-up surveys of non-respondents is an additional option open to data collectors. They can be thought of as auxiliary data and can be integrated with survey data to directly assess non-response bias. The procedure involves re-contacting non-respondents to compare their responses with those of initial respondents. Based on this comparison, it is possible to generate an indicator of response bias. However, hard-to-reach groups may still be hard to interview, and careful consideration should be given to the costs/benefits of follow-up surveys (Glynn, Laird, & Rubin, 1993; Lynn, 2003; Zaslavsky, Zaborski & Cleary, 2002; Peytchev, Peytcheva & Groves, 2010).

Correlations between auxiliary variables and survey outcomes involve examining the relationship between auxiliary variables (for example, paradata or variables from the sampling frame) and key survey outcomes. A strong correlation suggests that the selected auxiliary variables can be effective in adjusting for non-response bias (Kreuter et al., 2010; Maitland, Casas-Cordero & Kreuter, 2009).

Comparison of respondent means across deciles of estimated response propensities, **variation of means across deciles of survey weights and correlations between post-survey weights and survey variables.** Significant variation may suggest that the response propensity is related to the survey variables, indicating potential non-response bias (Olson, 2006).

Comparison of late and early respondents can be used to assess potential non-response bias. Late responders are thought to share characteristics with non-responders and may differ systematically from early respondents. If late respondents resemble nonrespondents, they can be used as a proxy for this group, and their inclusion can help identify non-response bias (Dunkelberg & Day, 1973; Smith, 1984; Lin & Schaeffer, 1995).

The fraction of missing information (FMI) quantifies the uncertainty introduced to survey estimates by missing data by assessing the proportion of variability in an estimate that is attributable to missing data (Wagner, 2010). The FMI approach stems from multiple imputation, replacing missing values with multiple sets of estimated values to account for the uncertainty associated with missing data and non-response (Rubin, 1976). The FMI is calculated as the ratio of the imputation variance to total variance, measuring the impact of missing data on the precision of survey estimates (Andridge & Little, 2011).

Proxy pattern-mixture modelling for survey non-response was developed by Andridge & Little (2011) to assess the degree of departure from the MAR assumption after non-response bias measurement and adjustments. This approach was recently extended to measure the degree of departure from ignorable sample selection for non-probability data sources and refers to unknown inclusion probabilities which may depend on the survey outcomes. Little, West, Boonstra & Hu (2019) called this indicator the standardised measure of unadjusted bias (SMUB), and it can be applied in the context of assessing the sensitivity of estimates to non-ignorable non-response (NMAR).

The strengths of item-based indicators include their ability to assess bias at the level of specific survey statistics, aligning the evaluation directly with the survey's substantive outcomes. Furthermore, by identifying specific subgroups or cases that contribute disproportionately to nonresponse bias, item-based indicators can inform adaptive survey designs, allowing for targeted interventions to improve response rates and reduce bias.

However, item-based indicators also have notable weaknesses. Their accuracy heavily depends on the validity of the statistical models used to relate auxiliary variables to survey outcomes; mis-specified models can lead to misleading conclusions. In surveys aiming to estimate multiple statistics simultaneously, item-based indicators may yield conflicting assessments of non-response bias across different objectives, complicating interpretation. The effectiveness of these indicators is also contingent upon the availability of high-quality auxiliary data that are strongly correlated with both response propensity and survey outcomes, which may not always be accessible.

5. Compensating for non-response bias

Once researchers have a good understanding of the amount of missingness and potential biases present in their data, they can consider how to correct for them. In this section, we summarise the main strategies typically used in practice.

5.1. Mitigating non-response bias via survey design

Non-response bias can be mitigated through the use of targeted and adaptive designs (Lynn, 2017; Schouten et al, 2017; Sladka and Lynn, 2025). These strategies include designs which offer (differential) incentives to boost initial participation (for example, monetary incentive; Peytchev et al., 2020), and the use of multi-mode data collection techniques (for example, face-to-face, telephone and web; Schouten et al., 2021). Additionally, responsive survey designs provide a framework for the data-driven tailoring of data collection procedures to different sample members, often for cost and bias reduction (Schouten, Peytchev & Wagner, 2017). These approaches rely on the use of integrated linked data, such as those from paradata and the sampling frame.

For example, the survey practitioner can use the PAF as a sampling frame, and so has access to the postcode. It may be possible to link the postcode to the lower super output area of the potential respondent. Once there is an area-level identifier, it is possible to integrate a plethora of statistics regarding the demographic make-up, deprivation level and population characteristics which are published at the aggregate-level by the ONS. It may then be possible to tailor the survey to that respondent. For example, if the respondent is in an area with a younger aggregate population, it may be reasonable to use a web-based survey rather than a face-to-face one. Hence, this may produce a higher response rate relative to the resources expended.

5.2. Post-survey adjustments

If non-response bias is present in the survey data, and the achieved samples do not accurately reflect the target population, then adjustments can be made to improve survey estimates. As discussed, data integration can support post-survey adjustments by making the assumption of missing at random more plausible. Practitioners should explore data that can explain the reason for not participating in the study and that are correlated to the key variables of interest.

Through the calculation of survey weights, unit-level data can be adjusted to represent the target population. A difference to note here is between survey design weights and non-response weights (National Centre for Research Methods, 2012). Survey design weights reflect the probability of a unit being selected in the study and depend on the sampling strategy used. Non-response weights are created based on the estimated propensity of a sample unit to participate in a survey, with higher weights being given to individuals who are less likely to participate. A third type of weight, known as post-stratification, is also sometimes created to correct for any remaining differences between sample and population distributions after non-response adjusted design weights have been applied. These differences could be due to residual non-response bias, sampling error or coverage error.

Inverse probability weighting (IPW) is a statistical technique designed to adjust for non-response (Mansournia & Altman, 2016). This approach aims to rebalance the achieved sample to better reflect the target population. An example of IPW is found in the second case study in section 7.

The theoretical underpinning of IPW relies on the MAR assumption. By modelling the response probability using variables that explain the participation patterns, it is possible to estimate the likelihood of each unit's participation and apply the inverse of these probabilities as non-response weights in subsequent analyses (Seaman & White, 2013). To calculate IPW, models such as logistic regression, lasso regression, random forest, XGBoost and neural networks can be employed to estimate response probabilities based on auxiliary information.

Inverse probability weighting is potentially much more effective at combating non-response bias than post-stratification, raking and calibration methods, because the data are unit-level rather than aggregate and because interactions between the predictor variables can relatively easily be assessed and incorporated. However, strongly relevant unit-level variables are more difficult to integrate with survey data than aggregate-level variables. The effectiveness of IPW hinges on the correct specification of the response model and the availability of relevant predicting variables. If important predictors of non-response are omitted or inaccurately measured, the weights may not fully correct for non-response bias. Additionally, extreme weights can increase the variance of estimates, potentially reducing statistical efficiency. To mitigate this issue, some surveys cap extreme weights, for example, replacing all weights above a certain threshold with a fixed maximum value to balance bias reduction and precision (Little, Carpenter, and Lee, 2024).

For example, the British Social Attitudes Survey is a repeated cross-sectional survey in the UK (NatCen Social Research, 2022). The BSA provides inverse probability of response weighting to adjust for differential survey non-response. Like all other cross-sectional PAF-based surveys, this is done at the address level using aggregate-level integrated data. Additionally, the likelihood of non-response by a second adult in a household with two or more eligible adults is estimated based using both area-level data and household-level data regarding household size, income and education in the household. Households with lower estimated response probabilities are consequently given higher weights to compensate for increased non-response within these households. The UK Household Longitudinal Study is a longitudinal household panel study (Understanding Society, 2025) which offers non-response weights to rebalance both the cross-sectional sample collected at each wave, and longitudinal sample who respond at each follow-up wave. These weights include an IPW component, modelling based on previous survey responses and household-level measures, and calibrated to a range of (sub)population totals (e.g. residents of Northern Ireland, female respondents in employment, babies born in the previous 12 months).

Post-stratification adjustment consists of comparing survey statistics to external benchmarks and creating survey weights so that the weighted sample aligns with known population distributions across specific, aggregate-level subgroups. The typical aim of such weights is to correct for any other biases that make the sample different from the population. This includes dealing with issues such as coverage error (when some members of the population have no chance of being selected) and other causes of non-response bias that were not corrected using IPW. Population distributions are typically obtained from sources like the UK Census, mid-year population estimates provided by the Office for National Statistics or another high-quality survey.

Post-stratification involves two steps:

1) Identifying "control totals" from the target population, which the survey data should "match", i.e. there should be measurement equivalence between data sources.
2) Calculating survey weights to adjust for over- and under-representation in the survey sample, based on the "control totals" identified.

Each post-stratum weight is calculated by dividing the population proportion by the achieved survey proportion. For example, if males make up 49% of the population but only 25% of the achieved survey sample, then the stratified weight for males would be 49/25=1.96. Applying this weight to male respondents will rebalance the survey sample to better reflect the target population. However, post-stratification approaches operate at the aggregate level and adjust to a known population parameter. To also effectively account for both sample design and individual-level non-response, , poststratification procedures can be applied to weighted estimates that incorporate design weights with IPW non-response adjustments.

Raking (iterative proportional fitting) is an alternative post-stratification strategy that repeatedly adjusts sample weights so that weighted marginal distributions (e.g. age, sex, region) match relevant population totals, which are integrated with survey data at the aggregate level (Deville & Särndal,1992). Raking requires only marginal distributions, making it an effective approach when the weighting groups are too small for a cross-classification of variables or when cross-classifications are not available at the population level.

Calibration weighting follows a similar iterative proportional fitting approach to the one described in Raking. Calibration weighting is designed to ensure that the data associated with a specific sample meets some specified constraints, typically alignment with estimates from a difference data source. This implies that individuals are weighted so that those who are underrepresented in the sample (compared to population totals) will receive higher weighting. Conversely, a lower weighting will be applied to individuals who are overrepresented in the sample. Calibration weights can be calculated for multiple levels of analysis to ensure representativeness across levels. For example, in the Labour Force Survey (Office for National Statistics, 2024), both individual and household statistics are calculated and the inconsistencies compared. Calibration uses the GREG approach which ensures that all individuals in a household obtain the same survey weight and at the same time ensures that individual benchmarks are maintained (National Centre for Research Methods, 2012).

Multilevel regression and poststratification (MRP) is often used in small area estimation and for non-probability surveys where raking and calibration approaches are not effective (Park, Gelman, and Bafumi, 2004; Downes et al., 2018). The MRP approach fits a hierarchical model to survey responses and reweights cell-level estimates to external benchmarks. The multi-level regression allows for estimates to be derived for individuals in similar areas based on extrapolation from similar areas. As a result, areas with low representativeness in the sample and therefore sparse responses are better identified. The post-stratification of the sample uses auxiliary population data to reduce bias and enhance representativeness (Lopez-Martin, Phillips & Gelman, 2022).

Applying non-response weights to survey data can adjust for bias due to differential response rates, but can also impact the standard errors of estimates. Weighting introduces variability; when weights differ substantially among respondents, they can inflate the variance of survey estimates and lead to larger standard errors. In many realistic survey settings where auxiliary variables are strongly related to response propensity and weakly related to the survey variable of interest, IPW may lead to more biased and less efficient estimates than complete cases due to the greater variance implied via survey weighting (Little, Carpenter & Lee, 2022). While non-response weighting aims to produce more accurate (and less biased) estimates, it often results in a trade-off with increased variance, and so the use and calculation of non-response weighting should be carefully considered by survey practitioners and researchers.

Post-survey adjustment methods assume missingness at random conditional on the chosen auxiliary variables and depend on the accuracy of integrated population margins. Best practice calls for starting with a limited set of well-measured variables, inspecting and trimming extreme weights, monitoring convergence (examining the step-by-step repetitive process by which estimates are arrived at), and thoroughly documenting margins, trimming rules, and diagnostics to ensure transparent, reproducible inference (National Centre for Research Methods, 2012).

5.3. Imputation methods

In survey research, weighting procedures are more commonly used for unit-level non-response, with imputation used to deal with item-level non-response. However, weighting can become cumbersome when bespoke weights are needed for multiple sub-samples of interest, such as with longitudinal survey data, where (cross-sectional) weights may be needed for the respondents at each wave and also (longitudinal weights) for those who responded at various combinations of waves. Another example is where the survey is carried out in multiple stages and there may be loss of participants at each stage of the repeated measurements (monotonic nonresponse). For example, a social survey interview may be followed by a second stage of data collection by a nurse for completing a questionnaire related to health and a third stage of biomarker data collection. One approach to compensate for the monotonic nonresponse is to calculate separate survey weights for each stage of the survey, but increasingly, imputation methods are being used. The creation of weights for many different uses is resource-intensive for survey providers and can be confusing for data users.

When imputation procedures are used to replace a very large block of missing data rather than using survey weights (such as the case of a multiple-stage survey where all the data from a stage is missing for a sizeable number of sample units) or to transfer variables from one dataset to another, this is known as 'mass-imputation'. Although mass-imputation increases the sample size beyond the original respondent dataset, there is additional uncertainty arising from the imputation models that need to be accounted for when calculating variance estimates, see for example, Rubin (1987). In general, weighting methods assume that the missing value for a nonrespondent is the average value of the variable of interest amongst respondents who match the nonrespondent on the variables included in the weighting model. Imputation methods, however, can refine this assumption and use different sets of auxiliary variables depending on the variable to be imputed. Examples where mass-imputation approaches are increasingly being used to compensate for monotonic nonresponse, particularly in multiple-stage surveys, are Park and Kim, 2019 and Chatzi, et al. 2024. An example of mass-imputation for data integration is in Chipperfield, et al. (2012) and more recent developments of using mass-imputation for integrating survey data and non-survey data appear in Kim, et al. (2018) and Chen, et al. (2022).

In the remainder of this section, we assume that imputation is used to address survey (unit) non-response. A good overview of imputation methods in the social sciences can be found in Durrant (2005). An example which uses imputation to address issues of survey non-response can be found in the second case study in section 7.

Imputation involves replacing missing data with substituted values from auxiliary data sources, particularly if there is a proxy variable available in non-survey data for respondents and non-respondents (for example, the administrative data previously discussed), allowing for complete case analyses. Integrated data can inform imputation for unit-level non-response, especially where the integrated variables are closely related to the survey variables. There are two main types of imputation: single and multiple imputation.

Single Imputation methods provide a single value for each variable, such as:

- ***Mean/Median Imputation*****:** Replacing missing values with the mean or median of observed responses.
- ***Regression Imputation*****:** Predicting missing values using regression models based on auxiliary variables.
- ***Nearest Neighbour*****:** Assigning missing values based on the closest observed data point in terms of similarity on a set of relevant variables.
- ***Hot Deck Imputation*****:** Substituting missing values with observed responses from a similar unit within the dataset.

Single imputation is often performed prior to weighting procedures to address item missingness for explanatory variables. While straightforward, single imputation methods can underestimate variability and may introduce bias if the imputation model is mis-specified. For example, regression imputation assumes a linear relationship between variables, which may not hold in all cases.

Multiple Imputation (MI) addresses the limitations of single imputation by creating multiple complete datasets, each with different imputed values reflecting the uncertainty around the missing data. Analyses are conducted on each dataset, and results are combined to produce estimates that account for both within- and between-imputation variability (Rubin, 1987). For example:

- ***Multiple imputation using multivariate imputation by chained equations*:** Using repeated regression modelling, which imputes values based on the appropriate distribution for the missing value and assumes each imputed variable is conditional on all other variables (i.e. each variable included is informative of missingness; Raghunathan, Lepkowski, Van Hoewyk & Solenberger, 2001). However, this approach struggles to account for skewness in the data; for example, positive variables may have imputed values which are negative (Austin, White, Lee & van Buuren, 2021). To mitigate this fact, researchers can place bounds on the imputation to restrict it to the real data values, but this can cause a downward bias in the variance of the imputed data. Moreover, it does not entirely mitigate the effect of skewness in arriving at potential bias in the results.
- ***Multiple imputation for continuous variables with the use of predictive-mean matching*:** Instead of drawing an imputed value from a random distribution, it is possible to draw an observed value from a donor with a similar predictive mean. This method is also referred to as semi-parametric imputation, as the predictive means are model-based but the actual imputed value is taken from a donor (Morris, White & Royston, 2014).

The multiple imputation process is repeated multiple times until the model arrives at consistent estimates. It is particularly useful when the missing data mechanism is Missing at Random (MAR) – more likely when the integrated data is particularly rich - and when relationships between variables are complex. However, imputation models need careful specification and may be computationally intensive, especially with large datasets or numerous variables (White et al., 2011).

An important note in the use of imputed datasets is not to treat imputed values as observed data. Datasets containing imputed values should not be thought of as “real” data, but as a convenient methodology for conducting statistical analyses that adjust for non-response biases. Diagnostic and sensitivity analyses should be conducted using the original data source (National Centre for Research Methods, 2012) and a combining approach for estimation and inference, primarily based on Rubin's Rules (1987), should be used. In practice, this conceptual caveat implies that when investigating response rates in a survey, or when evaluating the suitability of the survey data for a given project, complete cases should be reported alongside imputed values to allow for comparison and to check that the imputation process has not added or exacerbated any biases in the dataset.

The Centre for Longitudinal Studies (CLS) provides an example of how to implement multiple imputation methods, which can be used to restore representativeness in survey data. While imputation methods are model-specific and should be tailored for each analysis, the CLS user guide provides an overview of useful steps for researchers and survey practitioners to consider (see Silverwood et al. (2024) for more information).

As with IPW, multiple imputation assumes missing at random. As such, the process of finding auxiliary data to explain the process of missingness is essential. Here, auxiliary data, such as those discussed above, are essential. Once the appropriate auxiliary variables are identified, the users can follow the process of building a model to explain the missing data, creating multiple datasets with the imputed values and finally, analysing these data and aggregating their results.

6. Recommendations

6.1. What to look for in external data?

The utility of external data integration to address non-response bias depends on the characteristics of both the survey and the external data. Auxiliary variables are useful in compensating for survey non-response when they explain the reasons for non-participation and are correlated with key variables of interest from the survey. As such, practitioners should keep in mind these criteria when exploring alternative data sources to use. They should also consider the quality of the auxiliary data. This means ensuring it covers the same target population, the same geographical area, the same time period and that the data is of good quality (e.g., limited missing data, good measurement of the concept of interest). They should also consider practical issues such as the resources and time needed to link the data.

6.2. Create reproducible procedures

Best practice calls for starting with a limited set of well-measured auxiliary variables, inspecting and trimming extreme weights, monitoring convergence, and thoroughly documenting the process to ensure transparent, reproducible inference.

Integrated data should be as findable, accessible, flexible to different data systems, and reproducible as possible (FAIR; Wilkinson, 2016). Data providers should be as transparent as possible about the linkage process and include detailed and clearly reported technical guides. Access to linked data for data users should be streamlined, the documentation required to access linked data should be kept up to date, and administrative burden should be minimised where possible.

6.3. Using informative sampling frames

Some survey institutes have permission to integrate administrative data with the survey sampling frame, which introduces additional contextual variables to aid in the estimation of non-response bias in the survey sample. For example, a number of the studies run by the Institute for Employment Research in Germany (Institute for Employment Research, 2025) are able to link employment variables to sampling frame information, providing administrative records for respondents and non-respondents. Similarly, Healthy Ageing in Scotland (HAGIS; Douglas, Rutherford & Bell, 2018) includes linked health records for the entire sampling frame via Community Health Index (CHI number). These approaches can help make the MAR assumption more feasible by providing observations for non-respondents, which may be more effective for evaluating and compensating for non-response in survey data.

In the UK, given that the PAF is typically used as the sampling frame for general population surveys, the current best practice is to use it as an anchor for linking surveys with alternative data sources. Typically, this is done at an aggregate level, such as LSOA.

These geographical identifiers can then be used to link survey responses to a wealth of data sources from the ONS and the UK Data Archive. This integrated data can then be used to develop models to explain the mechanisms for missing data and to create weights that can compensate for non-response bias.

6.4. Using paradata in data collection

In addition to linking surveys to external data sources, data collectors can maximise the use of paradata. This type of data is created during the process of collecting the survey. The type of paradata collected depends on the mode of interview used. For example, in a face-to-face survey, interviewers can take notes regarding the neighbourhood and the dwelling of the sampled units. They can also collect information regarding the contact history and the outcomes of each one. While interviewer observations cannot be collected for telephone interviews, some contact history can still be captured. Similarly, web surveys and mail surveys can capture some information about the delivery of the invitations.

Paradata can be used to develop procedures for data collection that may reduce non-response bias (see below), and they can also be included in models for correcting for non-response (like weighting and imputation).

6.5. Using adaptive and response designs

In addition to corrections after the data was collected, researchers can focus on minimising non-response bias during data collection. Here, procedures such as targeted and responsive designs can be valuable. Targeted designs use integrated or frame data to apply different survey protocols to different subgroups. Responsive designs imply a continuous tracking of the field-work process and an ability to adjust procedures based on the performance observed in the field. Here, paradata as well as frame and integrated information can be invaluable. Such data can be used to create R indicators and CVs during the fieldwork period. Based on these, agencies' collecting data can decide to change procedures. This could mean prioritising certain subgroups, for example, using better interviewers, more visits, higher incentives, changing the modes of invitation and interviews or even stopping data collection. These procedures can be combined with post-hoc non-response adjustments (e.g., using weighting) to lead to less bias and uncertainty in estimates.

7. Case Studies

Evaluating the Utility of Linked Administrative Data for Nonresponse Bias Adjustment in a Piggyback Longitudinal Survey

(Büttner, Sakshaug, & Vicari, 2021)
https://journals.sagepub.com/doi/abs/10.2478/jos-2021-0037

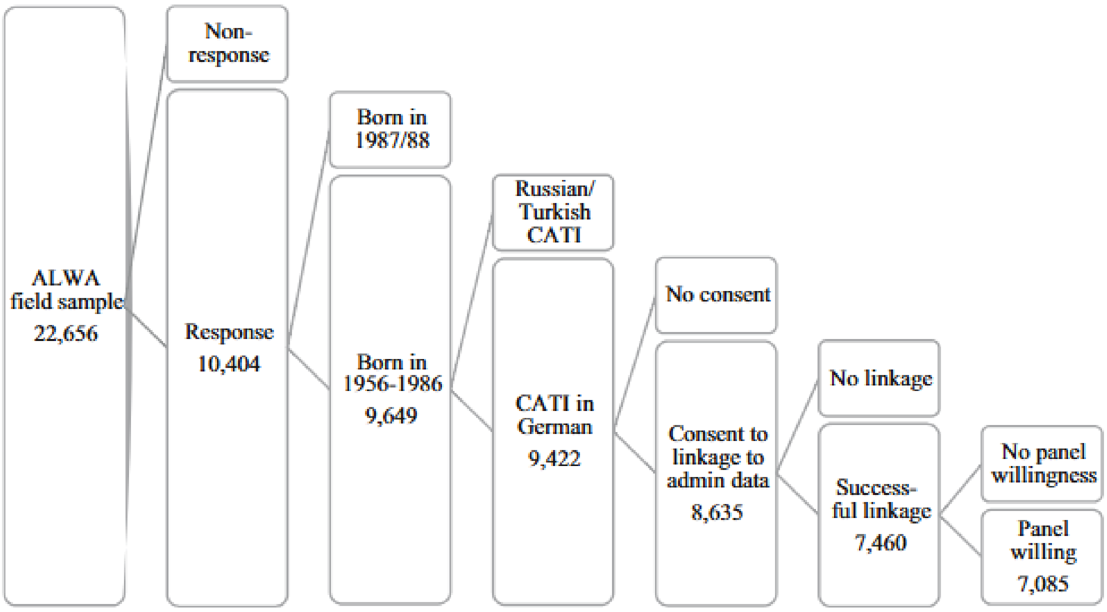
Study purpose

The aim of this study was to evaluate whether linking federal administrative records to respondents in a piggyback longitudinal survey design could improve the monitoring and adjustment of unit nonresponse and attrition bias.

Survey data

The survey data used in this analysis come from the National Educational Panel Study (NEPS) Adult Cohort, a "piggyback" survey built on the cross-sectional survey ALWA. ALWA interviewed 9,649 German adults (born 1956–1986) in 2007–08 by telephone. Of this sample, 93.2% consented to follow-up. After excluding non-German CATI cases and unsuccessful linkages, 7,085 panel respondents formed the baseline NEPS sample. Eight subsequent mixed-mode (CATI/CAPI) waves (2009–17) achieved response rates declining from 73.3% in wave 1 to 43.1% in wave 8. Weights in the original NEPS scheme adjust for sampling design, linkage steps, and panel willingness using paradata (contact attempts), frame variables, prior-wave survey measures, and administrative data.

Figure 1: Flow diagram for selection steps in the NEPS sample.



Non-survey data

The non-survey data used in this analysis are administrative employment records, which were drawn from the Integrated Employment Biographies (IEB) of Germany's Federal Employment Agency. These records provide longitudinal spell-level information on compulsory social-security employment, unemployment benefits (UB I and II), wages, working time, and job changes. Indicators of current labour-market status were calculated (e.g., full-time, part-time, marginal employment, unemployment benefits receipt, commuting) alongside between-wave changes (e.g., job loss, benefit uptake, employer change) for all linked sample units, including respondents who attrit.

Methods

Each selection stage (linkage consent, successful linkage, panel willingness, and wave-specific response) was modelled via logistic regression using two sets of predictors: the original NEPS weighting variables and the original NEPS weighting variables plus the linked administrative variables. Wave *t* weights multiply the inverse predicted propensities across prior stages. Attrition bias was assessed by comparing administrative outcome means under three scenarios: unadjusted, original NEPS weights, and augmented weights. These groups were compared against benchmarks from the full linked sample. Variance effects were summarised via changes in coefficients of variation, and substantive survey estimates were compared across weighting schemes.

Advantages

Incorporating administrative data into weighting led to modest but consistent reductions in attrition bias for labour-market outcomes (for example, current employment, average daily wage) without inflating variance. The linked records capture between-wave events, such as job loss, that standard paradata and prior-wave survey variables miss. Because the piggyback design reuses existing linkages from the ALWA survey, the approach requires no additional field costs for auxiliary data collection and offers ongoing monitoring of attrition patterns across waves.

Limitations

Gains from administrative adjustment were small for most variables and limited to domains covered by the IAB. Linkage consent remains a barrier, and administrative coverage excludes non-social-security employees (for example, civil servants, self-employed), which may bias the auxiliary data. Some relevant events (for example, household moves, composition changes) could not be observed in administrative records. Improvements depended on strong correlations between auxiliary records and survey outcomes.

Using linked Hospital Episode Statistics data to aid the handling of non-response and restore sample representativeness in the 1958 National Child Development Study

(Rajah et al., 2023)
https://cls.ucl.ac.uk/wp-content/uploads/2023/02/CLS-Working-Papers-2023-1-Using-linked-Hospital-Episode-Statistics-data-to-aid-the-handling-of-non-response.pdf

Study purpose

This paper examined the extent to which linkage to Hospital Episode Statistics (HES) can refine the handling of non-response in the 1958 National Child Development Study (NCDS). Indicators predictive of non-response at the age-55 sweep were identified from the HES data, and the utility of integrating these auxiliary variables within multiple-imputation models was assessed.

Survey data

The survey data used in this study were wave 9 of the NCDS, a British birth cohort that began in 1958 with 17,415 participants and has followed them through eleven sweeps up to age 64. At the age-55 sweep, 6,517 cohort members were alive and resident in the UK, of whom 5,786 (88.8%) completed the survey. Early-life measures include the father's occupational social class at birth and cognitive ability at age seven, while later-life outcomes comprise the highest educational qualification and marital status recorded at age 55.

Non-survey data

Auxiliary information is drawn from deterministic linkage to HES across four domains—Admitted Patient Care, Critical Care, Accident & Emergency and Outpatient appointments, using name, date of birth, sex and postcode. Covering the period 1997–2012, the linkage provided 58 aggregated variables. The HES data is at the episode level rather than the individual level; therefore, composite variables were constructed from the episodes associated with each individual. These included counts of admissions and outpatient attendances, counts of missed or cancelled appointments, procedural flags and categorical groupings based on ICD-10 chapters.

Methods

Variable selection was conducted in two stages. First, a LASSO-penalised logistic regression model identified HES indicators most predictive of non-response at age 55. Second, the selected HES variables and established survey predictors were used in a multiple imputation via chained equations (MICE) approach, under a missing-at-random assumption. Two analytical frames were employed: one restricted to participants who consented to HES linkage, used to assess bias in early-life variables against their full-cohort distributions; and a second encompassing all cohort members, comparing imputed later-life estimates against external benchmarks from the 2013 Annual Population Survey. Only individuals who responded to wave eight could respond to linkage. This amounted to 6,593 of the 18,558 total NCDS members.

Advantages

By combining a richly characterised birth cohort with high-quality administrative linkage, the study leverages extensive auxiliary information to probe non-response mechanisms. The data-driven LASSO approach offers a transparent and replicable method for selecting predictive HES indicators. The MICE provides an example for bias correction. The use of integrated health records allowed for the identification of previously MNAR mental health variables, which aided in later imputation approaches.

Limitations

The HES variables are aggregated at a fairly broad level, which may obscure more granular patterns of health-service use predictive of non-response. Because linkage consent was obtained at age 50, individuals who dropped out earlier lack administrative records, limiting the applicability of HES-based adjustments to later attrition only. The benefit of HES auxiliaries over survey predictors proved modest as the survey data already included a broad range of relevant variables.

References

Andridge, R., & Little, R. (2011). Proxy pattern-mixture analysis for survey nonresponse. *Journal of Official Statistics*, 27(2), 153–180.

Austin, P. C., White, I. R., Lee, D. S., & van Buuren, S. (2021). Missing Data in Clinical Research: A Tutorial on Multiple Imputation. *The Canadian Journal of Cardiology*, 37(9), 1322–1331. https://doi.org/10.1016/j.cjca.2020.11.010

Bethlehem J G (1988) Reduction of nonresponse bias through regression estimation, *Journal of Official Statistics* 4(3), 251-260.

Bethlehem, Jelke G. 2002. "Weighting nonresponse adjustments based on auxiliary information." In *Survey Nonresponse*, eds. Groves, Robert M., Don Dillman, John L. Eltinge, and Roderick J. A. Little, pp.275-288. New York: Wiley & Sons.

Blom, A. G. (2008). Measuring nonresponse cross-nationally (*ISER Working Paper* Series No. 2008-41). Colchester: University of Essex.

Booth, C., Crawford, C., Rajah, N., Silverwood, R., & Henderson, M. (2024). Examining sample representativeness and data quality in the linked Next Steps survey and Student Loans Company administrative data (*CLS Working Paper* 2024/1). London: UCL Centre for Longitudinal Studies.

Burton, J., Couper, M. P., Crossley, T. F., Jäckle, A., & Walzenbach, S. (2025). How Do Survey Respondents Decide Whether to Consent to Data Linkage? *Sociological Methods & Research*. https://doi.org/10.1177/00491241251344289

Büttner, T. J. M., Sakshaug, J. W., & Vicari, B. (2021). Evaluating the utility of linked administrative data for nonresponse bias adjustment in a piggyback longitudinal survey. *Journal of Official Statistics*, 37(4), 837–864. https://doi.org/10.2478/jos-2021-0037

Centre for Longitudinal Studies (2025). CLS. https://cls.ucl.ac.uk/data-access-training/data-access/accessing-data-directly-from-cls/

Chatzi, G., Chandola, T., Shlomo, N., Cernat, A. and Hannemann, T. (2024). Is Social Disadvantage a Chronic Stressor? Socioeconomic Position and HPA Axis Activity Among Older Adults Living in England, *Psychoneuroendocrinology*. https://doi.org/10.1016/j.psyneuen.2024.107116.

Chen, S., Yang, S., Kim, J. K. (2022). Nonparametric Mass Imputation for Data Integration. *Journal of Survey Statistics and Methodology*, 10 (1), 1-24, https://doi.org/10.1093/jssam/smaa036

Chipperfield, J., Chessman, J. and Lim, R. (2012). Combining Household Surveys using Mass Imputation to Estimate Population Totals. *Australian & New Zealand Journal of Statistics*, 54(2), 223-238. https://doi.org/10.1111/j.1467-842X.2012.00666.x

Cornesse, C., Blom, A. G., Dutwin, D., Krosnick, J. A., De Leeuw, E. D., Legleye, S., Pasek, J., Pennay, D., Phillips, B., Sakshaug, J. W., Struminskaya, B., & Wenz, A. (2020). A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research. *Journal of Survey Statistics and Methodology*, 8(1), 4–36. https://doi.org/10.1093/jssam/smz041

Department for Digital, Culture, Media and Sport. (2024). *Community Life Survey*. [data series]. *7th Release*. UK Data Service. SN: 2000100, DOI: http://doi.org/10.5255/UKDA-Series-2000100

Deville, J.-C., & Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418), 376–382. https://doi.org/10.2307/2290268

Douglas E, Rutherford A, Bell D (2018) Pilot study protocol to inform a future longitudinal study of ageing using linked administrative data: Healthy AGeing in Scotland (HAGIS). *BMJ Open*, 2018;8:e018802. doi: 10.1136/bmjopen-2017-018802

Downes, M., & Carlin, J. B. (2020). Multilevel regression and poststratification as a modeling approach for estimating population quantities in large population health studies: A simulation study. *Biometrical journal*, 62(2), 479–491. https://doi.org/10.1002/bimj.201900023

Dunkelberg, W. C., & Day, G. S. (1973). Nonresponse bias and callbacks in sample surveys. *Journal of Marketing Research*, 10(2), 160–168. https://doi.org/10.2307/3149821

Durrant, G. B. (2005). Imputation methods in the social sciences: a methodological review. *NCRM Working Paper* Series, NCRM-002. https://eprints.ncrm.ac.uk/id/eprint/86/1/MethodsReviewPaperNCRM-002.pdf

Gauthier, A. H., Kong, S., Grünwald, O., Bujard, M., Caporali, A., Deimantas, V. J., Emery, T., Jablonski, W., Koops, J. C., Rijken, A., & Schumann, A. (2025). The Generations and Gender Survey: A cross-national longitudinal resource. *European Sociological Review.* Advance online publication. https://doi.org/10.1093/esr/jcaf005

Gelman, J. L.-M., Justin H. Phillips, and Andrew. (2022). *Multilevel Regression and Poststratification Case Studies*. In bookdown.org. https://bookdown.org/jl5522/MRP-case-studies/

Glynn, R. J., Laird, N. M., & Rubin, D. B. (1993). Multiple Imputation in Mixture Models for Nonignorable Nonresponse with Follow-ups. *Journal of the American Statistical Association*, 88(423), 984–993. https://doi.org/10.1080/01621459.1993.10476366

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5), 646–675. https://doi.org/10.1093/poq/nfl033

Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(3), 439–457. https://doi.org/10.1111/j.1467-985X.2006.00423.x

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, 72(2), 167–189. https://doi.org/10.1093/poq/nfn011

Harron, K. L., Doidge, J. C., Knight, H. E., Gilbert, R. E., Goldstein, H., Cromwell, D. A., & van der Meulen, J. H. (2017). A guide to evaluating linkage quality for the analysis of linked data. *International Journal of Epidemiology*, 46(5), 1699–1710. https://doi.org/10.1093/ije/dyx177

I-Fen Lin, & Schaeffer, N. C. (1995). Using Survey Participants to Estimate the Impact of Nonparticipation. *Public Opinion Quarterly*, 59(2), 236–258. http://www.jstor.org/stable/2749703

Institute for Employment Research. (2025). IAB - Institut Für Arbeitsmarkt- Und Berufsforschung. https://iab.de/en/the-iab/surveys/

Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F. and Walzenbach, S. (2021) How and why does the mode of data collection affect consent to data linkage?, *Understanding Society Working Paper* Series, No.2021-04. Colchester: University of Essex

Kim, J. K., Park, S., Chen, Y., Wu, C. (2021). Seho Park, Yilin Chen, Changbao Wu, Combining Non-Probability and Probability Survey Samples Through Mass Imputation, *Journal of the Royal Statistical Society Series A*, 184 (3), 941-963. https://doi.org/10.1111/rssa.12696

Kreuter F, Olson K, Wagner J, Yan T, Ezzati-Rice TM, Casas-Cordero C, Lemay M, Peytchev A, Groves RM, Raghunathan TE. Using proxy measures and other correlates of survey outcomes to adjust for non-response: examples from multiple surveys. *Journal of the Royal Statistical Society Series A*. 2010; 173(2):389–407

Kreuter, F., & Kohler, U. (2009). Analyzing contact sequences in call record data. Potential and limitations of sequence indicators for nonresponse adjustments in the European Social Survey. *Journal of Official Statistics*, 25, 203-226.

Kreuter, F., Lemay, M., & Casas-Cordero, C. (2007). Using proxy measures of survey outcomes in post-survey adjustments: Examples from the European Social Survey (ESS). In *Proceedings of the Survey Research Methods Section* (pp. 3142–3149). American Statistical Association.

Little, R. J. A., West, B. T., Boonstra, P. S., & Hu, J. (2020). Measures of the degree of departure from ignorable sample selection. *Journal of Survey Statistics and Methodology*, 8(5), 932–964. https://doi.org/10.1093/jssam/smz023

Little, R.J., Carpenter, J.R. and Lee, K.J. (2024): A Comparison of Three Popular Methods for Handling Missing Data: Complete-Case Analysis, Inverse Probability Weighting, and Multiple Imputation. *Sociological Methods & Research* 53(3), 1105-1135. https://doi.org/10.1177/00491241221113873

Luiten, A., & Schouten, B. (2013). Tailored fieldwork design to increase representative household survey response: An experiment in the survey of consumer satisfaction. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1), 169–189. https://doi.org/10.1111/j.1467-985X.2012.01080.x

Lynn, P. (2003). Developing quality standards for cross-national survey research: five approaches. *International Journal of Social Research Methodology*, 6(4), 323–336. https://doi.org/10.1080/13645570210132848

Lynn, P., (2015). The need for representative survey samples. *Longitudinal and Life Course Studies*. 6 (4), 453-457

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

Maitland, A., Casas-Cordero, C., & Kreuter, F. (2009). An evaluation of nonresponse bias using paradata from a health survey. *Proceedings of the Government Statistics Section of the American Statistical Association*

Mansournia, M. A., & Altman, D. G. (2016). Inverse probability weighting. *British Medical Journal*, 352, i189. https://doi.org/10.1136/bmj.i189

Maslovskaya, O., Lynn, P., Calderwood, L., Durrant, G., Fitzgerald, R., Nicolaas, G., Williams, J. (2025) *Survey Futures Position Statement on Response Rates*. 09 June 2025. Available at https://surveyfutures.net/wp-content/uploads/2025/06/Response-Rates-Posititon-Statement_SurveyFutures.pdf.

Moore, J. C., Durrant, G. B., & Smith, P. W. F. (2018). Data set representativeness during data collection in three UK social surveys: Generalizability and the effects of auxiliary covariate choice. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(1), 229–248. https://doi.org/10.1111/rssa.12256

Moore, J. C., Durrant, G. B., & Smith, P. W. F. (2018). Data set representativeness during data collection in three UK social surveys: Generalizability and the effects of auxiliary covariate choice. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(1), 229–248. https://doi.org/10.1111/rssa.12256

Morris, T. P., White, I. R., & Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology*, 14(75). https://doi.org/10.1186/1471-2288-14-75

Morris, T. P., White, I. R., & Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology*, 14, 75. https://doi.org/10.1186/1471-2288-14-75

Mukaka, M., White, S. A., Terlouw, D. J., Mwapasa, V., Kalilani-Phiri, L., & Faragher, E. B. (2016). Is using multiple imputation better than complete case analysis for estimating a prevalence (risk) difference in randomized controlled trials when binary outcome observations are missing?. *Trials*, 17, 341. https://doi.org/10.1186/s13063-016-1473-3

NatCen Social Research. (2024). *British Social Attitudes Survey*. [data series]. 4th Release. UK Data Service. SN: 200006, DOI: http://doi.org/10.5255/UKDA-Series-200006

National Centre for Research Methods (2012). Weighting. *Practical Exemplars for the Analysis of Surveys (PEAS)*. Retrieved December 8, 2025, from https://www.restore.ac.uk/PEAS/theoryweighting.php

O’Toole, T., Cernat, A., Tzavidis, N., Shlomo, N., & Sakshaug, J. (2025). Data integration: Options for integrating survey and non-survey data, *Survey Futures Survey Practice Guide* no. 1. Colchester: University of Essex. https://surveyfutures.net/practice-guides/

Office for National Statistics. (2023). *Quality indicators for representativeness in administrative data: R-indicators and distance metrics*. ONS Methodology.

Office for National Statistics. (2024). *Labour Force Survey*. [data series]. 11th Release. UK Data Service. SN: 2000026, DOI: http://doi.org/10.5255/UKDA-Series-2000026

Office for National Statistics. (2025). *Crime Survey for England and Wales*. [data series]. 4th Release. UK Data Service. SN: 200009, DOI: http://doi.org/10.5255/UKDA-Series-200009

Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly*, 70(5), 737–758. https://doi.org/10.1093/poq/nfl038

Park, S. and Kim, J. K. (2019). Mass Imputation for Two-phase Sampling, *Journal of the Korean Statistical Society*, 48(4), 578-592. https://doi.org/10.1016/j.jkss.2019.03.002.

Park, D. K., Gelman, A., & Bafumi, J. (2017). Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls. *Political Analysis*, 12(4), 375–385. doi:10.1093/pan/mph024

Peytchev, A., Peytcheva, E., & Groves, R. M. (2010). Measurement error, unit nonresponse, and self-reports of abortion experiences. *Public Opinion Quarterly*, 74(2), 319–327. https://doi.org/10.1093/poq/nfq002

Peytchev, A., Peytcheva, E., Conzelmann, J. G., Wilson, A., & Wine, J. (2020). Modular survey design: Experimental manipulation of survey length and monetary incentive structure. *Journal of Survey Statistics and Methodology*, 8(2), 370–384. https://doi.org/10.1093/jssam/smz006

Plewis, I., & Shlomo, N. (2017). Using response propensity models to improve the quality of response data in longitudinal studies. *Journal of Official Statistics*, 33(3), 753–779. https://doi.org/10.1515/jos-2017-0035

Raghunathan, T. Lepkowski, J.M., Van Hoewyk, J., Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. Survey Methodology 27. *Survey Methodology*.

Rajah, N., Calderwood, L., De Stavola, B. L., et al. (2023). Using linked administrative data to aid the handling of non-response and restore sample representativeness in cohort studies: The 1958 National Child Development Study and hospital episode statistics data. *BMC Medical Research Methodology*, 23(266). https://doi.org/10.1186/s12874-023-02099-w

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. https://doi.org/10.2307/2335739

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons. https://doi.org/10.1002/9780470316696

Sakshaug, J. W. (2022). Reducing nonresponse and data linkage consent bias in large-scale panel surveys. *Forum for Health Economics & Policy*, 25(1–2), 41–55. https://doi.org/10.1515/fhep-2021-0060

Schouten, B., & Shlomo, N. (2017). Selecting adaptive survey design strata with partial R-indicators. *International Statistical Review*, 85(1), 143–163. https://doi.org/10.1111/insr.12153

Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, 35(1), 101–113.

Schouten, B., Peytchev, A., & Wagner, J. (2017). *Adaptive Survey Design* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315153964

Schouten, B., van den Brakel, J., Buelens, B., Giesen, D., Luiten, A., & Meertens, V. (2021). *Mixed-Mode Official Surveys: Design and Analysis* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429461156

Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, 22(3), 278–295. https://doi.org/10.1177/0962280210395740

Silverwood, R., Narayanan, M., Dodgeon, B., Katsoulis, M., & Ploubidis, G. (2024). *Handling missing data in the CLS cohort studies: User guide*. London: UCL Centre for Longitudinal Studies.

Skalland, B. (2011). An alternative to the response rate for measuring a survey's realization of the target population. *Public Opinion Quarterly*, 75(1), 89–98. https://doi.org/10.1093/poq/nfq072

Sladka, V. & Lynn, P. (2025). Targeted procedures for tackling survey non-response: evidence review. *Survey Futures Report* no.5. Colchester: University of Essex. https://surveyfutures.net/reports/

Smith, T. W. (1984). Estimating Nonresponse Bias with Temporary Refusals. *Sociological Perspectives*, 27(4), 473-489. https://doi.org/10.2307/1389038

Statistics Canada. (2025). *A multivariate technique for multiply imputing missing values using a sequence of regression models*. Ottawa: Statistics Canada. https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X20010015857

The American Association for Public Opinion Research. 2023 *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 10th edition. AAPOR.

UK Data Service, (2025). *Survey non-response* — UK Data Service. UK Data Service. https://ukdataservice.ac.uk/glossary/survey-non-response/

University College London, UCL Social Research Institute, Centre for Longitudinal Studies. (2025). *1970 British Cohort Study*. [data series]. 12th Release. UK Data Service. SN: 200001, DOI: http://doi.org/10.5255/UKDA-Series-200001

University of Essex, Institute for Social and Economic Research. (2025). *Understanding Society*. [data series]. 14th Release. UK Data Service. SN: 2000053, DOI: http://doi.org/10.5255/UKDA-Series-2000053

Wagner, J. (2012). A comparison of alternative indicators for the risk of nonresponse bias. *Public Opinion Quarterly*, 76(3), 555–575. https://doi.org/10.1093/poq/nfs032

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4), 377–399. https://doi.org/10.1002/sim.4067

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. https://doi.org/10.1038/sdata.2016.18

Wiśniowski, A., Sakshaug, J. W., Perez Ruiz, D. A., & Blom, A. G. (2020). Integrating probability and nonprobability samples for survey inference. *Journal of Survey Statistics and Methodology*, 8(1), 120–147. https://doi.org/10.1093/jssam/smz051

Zaslavsky, A. M., Zaborski, L. B., & Cleary, P. D. (2002). Factors affecting response rates to the Consumer Assessment of Health Plans Study survey. *Medical Care*, 40(6), 485–499. https://doi.org/10.1097/00005650-200206000-00006