

# Sampling Frame for National Surveys

A Use Case for the TCSS Organisation Register

March 2026

---

The UK Third and Civil Society Sector Database's organisational spine offers a near-comprehensive register of civil society organisations appearing in administrative data across the UK. Several features make it a powerful sampling frame for primary research.

## 1. Population Coverage and Stratification

---

For the first time, it will be possible to draw a random sample of organisations that crosses legal form and organisation-type, and is pulled from the same sampling frame. Researchers can draw samples stratified by organisational type, regulatory jurisdiction, geographic location (region, local authority, rural-urban classification, deprivation decile), industrial classification (ICNPTSO section or SIC code), or organisational age derived from registration dates. This makes it possible to design studies targeting specific sub-populations (e.g., CICs in deprived areas, or health charities in rural England) while still knowing the population from which the sample is drawn.

## 2. Temporal Dimension

---

Registration and removal dates allow researchers to define the population at any point in time — sampling only currently active organisations, or constructing historical cohorts (e.g., all charities founded between 2010 and 2015). This supports both cross-sectional surveys and longitudinal panel designs where organisations are sampled at formation and followed up.

## 3. Size-Based Sampling

---

The linked financial history data provides annual income and expenditure for most charities and nonprofit companies, enabling size-stratified sampling. This is particularly important given the extreme right-skew of the income distribution in civil society: a simple random sample would overwhelmingly draw micro-organisations and miss the relatively few large charities that account for the bulk of sector income and expenditure. Researchers can use income bands (aligned with standard UK charity size thresholds) to oversample larger organisations or to ensure adequate representation across the size distribution.

## 4. Screening for Specific Characteristics

---

The procurement data linkage identifies organisations that have received public sector contracts, enabling researchers to sample specifically from government suppliers — or to construct matched comparison groups of similar organisations that have not engaged in public procurement. Likewise, the financial variables allow screening for organisations with particular income profiles (e.g. high dependence on donations, or persistently loss-making organisations).

## 5. Practical Contact Information

---

The spine includes organisational names, addresses, and postcodes, providing the basic contact details needed to administer postal or in-person surveys for research.

## 6. Survey Weighting

---

The spine as a sampling frame also provides opportunities for robust survey weighting approaches, given the characteristics of the population that are known.

### Design weights

Design weights correct for unequal selection probabilities that arise from the sampling design itself. They are the inverse of each unit's probability of inclusion and are determined entirely at the design stage. Because the spine contains the full population within each stratum, these probabilities are known exactly. Stratified sampling (oversampling larger organisations, rarer organisational types, or specific geographic areas) will require design weights that vary accordingly across strata.

### Non-response weights

Non-response weights adjust for the fact that not all sampled units participate. If non-response is related to characteristics that also predict the outcome of interest, unadjusted estimates will be biased. The spine is particularly useful here because the full set of auxiliary variables — organisational type, income, age, jurisdiction, region, deprivation decile, rural-urban classification, and procurement engagement — is available for both respondents and non-respondents, enabling detailed diagnosis and modelling of non-response. Response propensities can be estimated via logistic regression on these covariates, and the inverse of the predicted response probability used as a non-response adjustment factor applied to the design weight. The richness of the spine means that meaningful predictors of non-response can be incorporated into the adjustment, reducing bias that would go undetected with a sparser sampling frame.

### Calibration weights

Calibration (or post-stratification) weights make a final adjustment so that weighted sample totals reproduce known population margins. Where design and non-response weights bring the sample closer to representativeness, calibration explicitly benchmarks the weighted sample against external population totals — and the spine supplies these totals directly. Researchers can calibrate to the known number of organisations by type, by region, by nation, by deprivation decile, or by

income band, using post-stratification. This is particularly valuable for studies producing estimates of total sector income, expenditure, or employment, where even modest deviations from the true size distribution can substantially distort weighted totals.

## 7. Limitations as a Sampling Frame

---

The spine is not a complete census of all civil society activity. It excludes unregistered community groups, informal associations, and organisations below regulatory thresholds (e.g., charities with income under £5,000 that are not required to register with the Charity Commission in England and Wales). Coverage varies by jurisdiction: financial data is less complete for Northern Ireland, and ICNPTSO classifications are only available for registered charities. Some organisations lack postcodes or registration dates, and the deduplication process may not capture all dual-registered organisations perfectly. Researchers should also note that the spine reflects a snapshot in time and that address information may become outdated for organisations that have moved. Researchers should assess the completeness of stratification and weighting variables for their target population before finalising a design, and may need to treat organisations with missing size or classification data as a separate stratum or exclude them with appropriate documentation. Similarly, because the spine is updated periodically rather than in real time, there will be some lag between the sampling frame and the actual population at the point of data collection.

Despite these caveats, the spine represents the most comprehensive openly available listing of UK civil society organisations and is substantially more inclusive than relying on any single register. It is well suited to supporting probability-based sampling designs for surveys, case study selection, and as a universe definition for quantitative research on the sector.

---

*Produced by the [UK Third and Civil Society Sector Database \(https://uk-third-sector-database.github.io/\)](https://uk-third-sector-database.github.io/) project.*