

## Data Profile: Linked Consumer Registers

### Introduction

This data profile describes a dataset of reports held by the CDRC which has been created by original CDRC research. The linked consumer registers contain the names and addresses for adults in the UK in annual snapshots between 1997 and the current or previous year. They are compiled from reported outputs from linking public versions of the electoral role and consumer registers (supplied by DataTalk, CACI and a new supplier). The data represent a near complete coverage of the adult population at an individual level.

The paper outlines the structure of the dataset and the variables that are available, highlights some potential uses of the data and provides some initial assessment of data quality.

### Scale and Extent

Field	Value
Data Providers	CACI UK Ltd, DataTalk Research Ltd, new supplier
Analytical Units	Individual
Data Format	CSV
Temporal Extent	1997-current
Geographical Extent	UK
Variables	Names, Addresses, entry year, exit year
Observations	Circa 50 million each year

### Citation Information

The following statement should be included when citing the use of this dataset:

“The data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC

[Project Number], ES/L011840/1;  
ES/L011891/1”

### Data Classification and Access Summary

These data are classified as Controlled and are available only upon approved application. To make an initial application, please visit: <https://www.cdrc.ac.uk/data-services/using-our-data/>

### Content

Field level metadata is provided in the table overleaf. Data are provided for each individual and record only names, the address and the duration of their occupancy at the address during the 20 year period. The data pertains to persons aged 17 and above only with the majority of inputs originating from public versions of the electoral register.

The number of cases in each dataset are outlined in the table below.

### Novelty

The Consumer Registers empower researchers to undertake individual level research on the adult UK population. As a set of geographic datasets available for several years, they can provide detailed estimates of local population change. They are also an invaluable tool for sample design. Research at the CDRC has used this data base to create ethnicity estimations (using the CDRC ethnicity estimator names software) and also applied novel linkage analysis to produce estimations of origin-destination internal migration flows.

### Quality

As the data is inputted from several different organisations, it is possible that some names and addresses are inconsistently formatted between datasets. Despite applying bespoke rigorous address matching and name matching methodologies, it is still possible that some cases may not have matched. Thus the number of unique addresses is slightly overestimated.

## Data Profile: Linked Consumer Registers

---

Addresses are recorded as address lines (in a separate table). This might make address matching to other data quite difficult as the number and composition of address lines varies by addresses, and between different versions of the data too.

In addition, it is very difficult to determine the completeness of the data. The register has near complete coverage of the adult population for every year. However, for some addresses anonymised residents are imputed where a house sale is known to have occurred but no new households were detected (roughly 6% of addresses in 2016). In addition, the occupancy of households are brought forwards by a few years where no new data are obtained for an address. Data lags occur because individuals do not volunteer data every year. Thus many records for the latest two years may be out of date.

It is also possible that adults who reside in multiple addresses may have duplicate entries within the data.

### Representation and Bias

Whilst the data providers have attempted to compile registers which are both as complete and accurate as possible, there are data biases that should be considered. Firstly, the electoral register is known to sufficiently under-represent the following groups: the younger age groups, the non-white British population & those in rented accommodation.

Further to this, research by the CDRC using implied ethnic groups from names identified that there was a slight over-representation of White British names. Roughly 84% of individuals were classed as White British, compared to 81% from the 2011 Census for the UK. It can also be identified that there are on roughly 1.4 million more females than males in each of the datasets (based on their titles). Research also identified that areas with higher proportion of adults in rented accommodation had the greatest underrepresentation within cities.

### Related Datasets

CDRC Population Churn Index  
<https://data.cdrc.ac.uk/dataset/cdrc-population-churn-index>

CDRC Modelled Ethnicity Proportions  
<https://data.cdrc.ac.uk/dataset/modelled-ethnicity-proportions>

## Data Profile: Linked Consumer Registers

### Data Triangulation: data sources used to establish provenance

Source	Variable	Spatial granularity of comparator	Temporal granularity of comparator	Note(s)
Address URN	Address Identifier	Individual	Annual	A separate lookup table provides the full address strings for each output (postcode is kept separate)
Forename	Forename	Individual	Annual	
Surname	Surname	Individual	Annual	
Entry Year	The first year the person was recorded at the address	Individual	Annual	The earliest data collection period is 1997
Exit Year	The last year the person was recorded at the address	Individual	Annual	The latest data collection period is the current or previous year.

