Understanding Society: Linked Driver and Vehicle Licensing Agency (DVLA) Dataset, 2013-2015

USER GUIDE

Version 2, June 2023

Contents

1. Introduction

Data linkage is one of the key innovative features of *Understanding Society*, the UK Household Longitudinal Study (UKHLS), which allows researchers to develop and implement new research agendas. It includes linkage to administrative data sets in key policy areas as well as other external data sources. *Understanding Society* currently pursues data linkage in a wide range of topic areas, including education, health, economic circumstances, and transport, as well as area-level characteristics.

The Driver and Vehicle Licensing Agency (DVLA) administrative data contains vehicle statistics (date of manufacture, registration, acquisition and loss, engine information, model, CO_2 emissions etc.). However, these administrative data lack information on related behaviours, attitudes and beliefs, relationships and family circumstances, economic and health statuses, and life events, which only survey data can supply.

The combined *Understanding Society* and DVLA data will allow analysis of car ownership and use in the context of detailed individual, family, and neighbourhood-level information. It will provide a much wider scope for robust and policy relevant analysis than is currently possible with these datasets in isolation.

2. Information on the DVLA data

DVLA data contains detailed information on vehicles accessible to Study participants. Vehicle statistics contain but are not limited to: brand name, year of manufacture, year of first registration, price band, engine size, vehicle mass, CO_2 emissions, and fuel type.

3. Consent - method and outcomes

The consent question "**vreglink**" is part of the UKHLS wave 5 (2013-2015) demographics module and was asked of all individuals aged 16 or older in all countries of the UK who: (1) had a full UK driving licence, (2) had access to a car or van for personal use, and (3) had at least one car or van for personal use registered in the UK. Of eligible individuals (27,543), 21,804 consented, 5526 declined consent and the remaining 213 either refused to answer the question or responded "don't know". If an individual agreed to consent then they were asked for the registration number of their (first, second, third) registered vehicles. The total number of vehicles reported was 26,974. In total there were 22,798 unique vehicles reported across 15,264 households.

Participants were provided with information about data linking/sharing so that they could make an informed decision as to the benefits/risks of consenting to having their data shared

and linked. This included reassurances as to how the link would be made, which types of data would be shared, and examples of potential research uses. Participants were also informed of how the linked data would be handled, provided with further information on the website, and given the contact details of the Scientific Leadership Team based at ISER, should they have any queries.

4. Matching procedure

Understanding Society respondents with a full UK Driving Licence were asked for their consent to access publicly available information about their vehicle, and for those who gave their consent, the vehicle registration was collected. The information collected was used to access public records from several sources about the vehicle. It's important to note that the car registration details were collected between 2013-2015 and the publicly available data about the vehicle in 2018 which included current information on MOT and road tax. While the data to time invariant factors are correct, the MOT and road tax data may not reflect the information current in 2013-15.

The first public data source is the DVLA vehicle enquiry service, containing the vehicle make, date of first registration, year of manufacture, cylinder capacity, CO_2 emissions, fuel type, euro status, export marker, vehicle status, vehicle colour, vehicle type approval, wheelplan and revenue weight. However, this service did not provide the model of the car, which was obtained together with the fuel type and vehicle colour from the DVLA check MOT service. The model of the car was important as it allowed access to further information from a third-party car database cars-data. However, inconsistencies across the model and/or fuel type between the DVLA databases emerged. These differences occurred mainly for cars with personalized number plates where the vehicle had been replaced but the owner had retained the original registration number.

The "**match_type**" variable identifies matched and unmatched cases between: (1) the DVLA public records and (2) the information from cars-data. 45.82% of consenters' vehicle information could not be matched across the two databases. 50.12% of records were matched based on a 'broad match' relying on 4 variables (vehicle brand, CO_2 emissions, cylinder capacity and fuel type). 4.01% (701 observations) were matched on a 'narrow match' based on 5 variables (vehicle brand, CO_2 emissions, cylinder capacity, fuel type and vehicle model).

The final match rate was 75.73% (17,265 unique registration numbers). It was not possible to match 5,497 registration numbers and a further 36 numbers were invalid and removed before matching. 318 matched observations were removed in the data cleaning process, being deemed duplicates (they occurred in households who reported more than one vehicle of the same brand and model which were first registered on the same date). The final (vehicle) sample size is 16,947.

The consent and linkage process potentially leads to non-random selection of individuals for whom linked data is available. We are working on guidance on deriving analysis-specific weights, which will cover issues specific to linked data. Guidance will be made available soon.

5. Data/variables included in the linked dataset

The DVLA data file (**DVLAcars**) can be linked to the UKHLS mainstage data using the **hidp** variable (household identifier) included in each file. The vehicle information is released at the household level because multiple individuals within a household may have access to the same car, thus avoiding duplicate information. Should individual level analysis be necessary, individual household member usage can be deduced from the variables "**ncars**" (number of cars/vans owned by household) and "**caruse**" (access to a car or van that each individual can use whenever they want).

The released dataset contains 33 variables. The number of observations (vehicles) is 16,947, among which 12,593 are unique household identifiers. All the frequencies and summary statistics, as well as descriptors of the variables, can be found at: https://www.understandingsociety.ac.uk/documentation/linked-data/environment.

6. Data access

The linked data is available from the UK Data Service, the details for which can be found at: https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8832. It is classified as End User Licence (EUL) and is available in Stata, SPSS and Tab delimited formats. Full details of the access requirements and the application process can be found at: https://www.ukdataservice.ac.uk/get-data/how-to-access.aspx.

7. Citation

If you use Understanding Society data you must cite every study that you use.

The bibliographic reference for this study is as follows:

University of Essex, Institute for Social and Economic Research. (2021). *Understanding Society: Linked Driver and Vehicle Licensing Agency (DVLA) Dataset, 2013-2015.* [data collection]. UK Data Service. SN: 8832, http://doi.org/10.5255/UKDA-SN-8832-1.

All works which use or refer to these materials should acknowledge these sources by means of bibliographic citation. To ensure that such source attributions are captured for bibliographic indexes, citations must appear in footnotes or in the reference section of publications.

<u>Citing this User Guide</u>
When citing this User Guide, you can use the citation of this particular version quoted below.

Institute for Social and Economic Research (2021), Understanding Society: Linked Driver and Vehicle Licensing Agency (DVLA) Dataset, 2013-2015, User Guide, Version 2, June 2023, Colchester: University of Essex.