

Quality Assurance of Administrative Data Sources

Licenced Heavy Goods Vehicles Statistics on Domestic Road Freight

Summary

The Department for Transport (DfT) is responsible for collecting information regarding the activity of UK registered Heavy Goods Vehicles (HGVs) within the United Kingdom (UK) and internationally and the goods they move and publishing subsequent statistics on road freight.

Road freight statistics are an indicator of the performance of the road freight industry and of the wider economy. They are important for providing an understanding of freight transport trends; for example, they support a better understanding of the load on the road network, impact on emissions, and the effectiveness of the freight industry.

The collection of statistics on road freight is governed by European Regulation (EU) No 70/2012 and is used to support EU competition policy and regulation of the European market. Statistics have been collected in this area since the 1950s, using powers under the Statistics of Trade Act 1947.

The DfT collect this data through three continuous road freight surveys:

- the Continuing Survey of Road Goods Transport (Great Britain) (CSRGT GB)
- the Continuing Survey of Road Goods Transport (Northern Ireland) (CSRGT NI)
- the International Road Haulage Survey (IRHS)

All three of these surveys rely on the use of “Administrative data” to produce their statistics which can be defined as:

“Data that are collected for administrative or operational purposes, with statistical use being a secondary purpose.”

This document focusses on the administrative data used by the Continuing survey of Road Goods vehicles in GB and Northern Ireland (CSRGT GB and CSRGT NI, respectively, referred to as CSRGT onwards for simplicity):

Vehicle licensing records of HGVs which have a goods vehicle taxation class (i.e. vehicles with a gross weight of over 3.5 tonnes which are used for carrying goods).

This licence information is collected and compiled by the Driver and Vehicle Licensing Agency (DVLA) and supplied to the DfT.

This document seeks to provide users of the road freight data with information regarding how this administrative data source is used in the survey and what impact DfT considers this to have on quality of the final statistical output. The Department for Transport is keen to receive views of users of Road Freight Statistics about this administrative data and we invite you to contact us at:

ROADFREIGHT.STATS@dft.gsi.gov.uk

Assessment of the data:

Our assessment of the administrative data was carried out using guidance as provided in the Quality Assurance Toolkit, developed by the United Kingdom Statistics Authority (UKSA) but reflects the opinions of DfT. The toolkit is designed to help statistical producers inform users about the quality assurance arrangements for administrative data.

The toolkit uses a matrix approach to assessing administrative data which compares risk surrounding quality concerns of the administrative data (rated as low, medium or high) against the level of public interest in the statistics (rated as low, medium or high). Combining these helps producers of statistics to determine what they feel is an appropriate level of assurance required regarding the impact of the administrative data on the statistics and document this accordingly.

There are 3 levels of assurance that statistics producers should seek to deliver regarding administrative data: basic, enhanced and comprehensive. Information on each of the three levels of assurance are detailed below:

- **Basic assurance:** The statistical producer has reviewed and published a summary of the administrative data QA arrangements
- **Enhanced assurance:** The statistical producer has evaluated the administrative data QA arrangements and published a fuller description of the assurance
- **Comprehensive assurance:** The statistical producer has investigated the administrative data QA arrangements, identified the results of independent audit, and published detailed documentation about the assurance and audit.

Assurance is sought across four areas of practice regarding the administrative data:

- Operational context and the administrative data collection processes
- Communication the DfT has with data supply partner
- The Quality Assurance principles, standards and checks applied by data suppliers
- DfT's Quality Assurance investigations on the data & the subsequent documentation.

Further information about these areas and the levels of assurance can be found on the UKSA's website¹.

How the data are used by DfT

The CSRGT survey is a vehicle-level survey of road freight activity. The CSRGT GB collects information on domestic journeys made by GB-registered HGVs transporting freight by road while the CSRGT NI collects information on domestic and international journeys made by NI-registered HGVs transporting freight by road.

Registered keepers of HGVs are asked to provide details of all freight trips within a specified survey week (7 day period) and are legally required to provide this information to the DfT under legislation set out in the Statistics of Trade act 1950. Not all registered keepers of HGVs are asked to report their activity - only activity from a selected number chosen via a stratified random sample is collected. This information is then grossed up to produce an overall picture of domestic road freight activity undertaken by UK registered HGVs.

In order to know which HGVs should be included in the survey and obtain contact details for the registered keepers, DfT consult a list of all licenced HGVs in the UK which are taxed as goods

¹ <https://www.statisticsauthority.gov.uk/monitoring-and-assessment/monitoring/administrative-data-and-official-statistics/>

vehicles. This list comes from vehicle licencing records which are held by the Driver and Vehicle Licensing Agency (DVLA).

The Driver and Vehicle Licensing Agency is an executive agency of the Department for Transport (DfT), which has responsibility for maintaining the registration and licensing of drivers in Great Britain and the registration and licensing of vehicles, together with the collection and enforcement of Vehicle Excise Duty (VED) in the United Kingdom.

The DVLA database will include a new record every time a new (or used) vehicle is first registered in the United Kingdom. A wide range of variables are captured, including the make, model, body type, taxation class of the vehicle, plus some basic information about the keeper of the vehicle, including the postcode of the keeper.

This information is transferred to the DfT who use it to select registered keepers of HGVs to be included in the CSRGT survey. The selected HGV keepers are given survey questionnaires to complete for freight trips undertaken during the survey period and this information is then used by the DfT to produce the road freight statistics.

Level of assurance

The road freight statistics series are required under European legislation which means they are technically deemed to have a high level of public interest. However, while the statistics themselves are extremely important they are not as high profile as, for example, GDP or employment statistics nor do they have the same potential impact if errors are found in the data. Therefore should the statistics ever have reason to be withdrawn (e.g. because of the discovery of an error in the data) while this would have local ramifications for the department, there wouldn't be severe wider political or economic considerations as suggested by the matrix.

With regards to quality, the relatively small role the administrative data play in producing the statistics (their use as a sampling frame rather than contributing to the outputs themselves) and the very low risk of the data being incomplete mean that their quality poses less of a risk.

Consequently DfT consider the level of assurance required for this series is **basic** and assurance appropriate to that level or above has been ascertained and set out in this document.

In general DfT is satisfied that the quality of the data received from DVLA combined with the separate checks DfT conducts on the data mean that the statistics output are of a sufficiently high standard as required to meet National Statistics status.

Practice area 1: Operational context & administrative data collection

The DVLA database will include a new record every time a new (or used) vehicle is first registered in the United Kingdom. A wide range of variables are captured, including the make, model, body type, colour and size of the vehicle, plus some basic information about the keeper of the vehicle, including the postcode of the keeper. More information about what is captured and how this is done can be found in the V355 Guidance Notes for Completion of the V55 Forms document (available here:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/516919/V355X5.pdf).

The initial registration of new vehicles is usually carried out by vehicle manufacturers at the point of sale. In around 97 per cent of cases, this process is carried out using the Automatic First Registration and Licensing (AFRL) system – in this case all of the vehicle particulars will be transferred directly from the manufacturers' records, leaving only the details of the first vehicle keeper to be entered into the electronic system at the point of sale. Any vehicle that has been imported into the United Kingdom by an individual or has been restored and brought back onto the road will be registered by the keeper. In these cases, along with any new vehicle that is being registered outside of the AFRL system, the information is keyed in manually by DVLA staff which was also the method used prior to the introduction of AFRL in 1998. Where a vehicles is being registered manually the following process is followed:

- The vehicle keeper completes a V55 form (although different variants exist for new and used vehicles, manufacturers and private individuals but all ask for the same info).
- The form is then submitted to DVLA along with evidence it's new – a customs form (if it's imported), plating certificate for HGVs and type approval evidence (where necessary).
- DVLA staff then verify all the information matches the supporting documents and manually key the data from the form into the vehicle record.
- Once complete, a V5C registration Certificate (log book) is issued to the customer with most of the information on it. This allows the customer to highlight if any information is incorrect and this to be changed.

When a vehicle changes hands, the old and new keepers are obliged to inform DVLA immediately, using different tear-off sections of the Vehicle Registration document. Commercial fleet operators may register with DVLA to use a dedicated 'fleet' management service to streamline their interactions with the agency. A vehicle keeper must keep their vehicle continuously taxed if they wish to use it on the road and strict automatic penalties are issued place if this in not adhered to. Vehicles being driven without the proper taxation risk being identified to police through a widespread network of Automatic Number Plate Recognition (ANPR) and other such technology making it difficult to operate an unlicensed vehicle over a long period. If the keeper wished to take the vehicle off the road (and hence stop paying tax) they must make a Statutory Off-Road Notification (SORN) to DVLA, which is also subject to regulation. The entry, operation and exiting of a vehicle from UK roads is a tightly controlled process meaning the records of vehicle licences provide good coverage and a reliable sampling frame for the CSRGT surveys.

Potential changes in the context or collection arrangements of the licence data, especially cessation of the collection or changes in the licence types could have a noticeable impact on the DfT's ability to identify licenced HGVs and run the CSRGT surveys. However such changes are unlikely without changes to legislation and are therefore deemed to be a low risk. It is also highly unlikely that major changes would come about without DfT's knowledge and input, given that vehicle licensing is directly within its area of responsibility, although statistical requirements may need to take second place to operational necessities.

Practice area 2: Communication with data supply partners

Vehicle licensing data extracts are provided to the DfT by the DVLA every quarter, under a long standing and well established data transfer arrangement. The main purpose of this arrangement is to allow DfT statisticians to produce a range of National Statistics on vehicle licensing in its own right, and to provide data for a range of secondary statistical purposes, amongst which is the design and sample selection of road freight surveys. The DVLA supplies extracts of all live vehicle registration records (whether currently licensed or not) on encrypted DVDs delivered by courier. There is no end date to this arrangement, and the data supply will continue as it is unless changed with the agreement of both parties. The DfT Vehicle Statistics team work closely with DVLA colleagues responding to FOIs, on other analytical and policy support tasks, and periodically exploring the potential for development of the data extracts, all of which serve to maintain an up-to-date mutual understanding of the data and the uses we put it to. While this arrangement does technically pose a risk that the data required may not be supplied, this risk is minor owing to the professional relationship that exists between DVLA and DfT however work is underway to formalise this arrangement through a data sharing agreement.

The Road Freight Statistics team use these data on a quarterly basis to select the sample for the CSRGT survey. The population of licensed HGVs taxed as goods vehicles is extracted from the DVLA database and a stratified random sample of HGVs is selected from this population. The stratum are based on the region in which the HGV is registered, the type of HGV artic/rigid and also the weight of HGV. Stratifying by the region in which the HGV is registered is done to gain adequate coverage across the UK however this may introduce bias due to the fact that HGVs can operate beyond the region in which registered.

The registration marks for HGVs selected to be part of the sample are sent back to DVLA for the purposes of adding address information which is not included in the normal quarterly data extract, for data protection reasons. DVLA then provide a list of addresses which are used to send out the survey forms. This is currently supplied on an encrypted disk sent to the DfT by DVLA, though processes are underway exploring whether this data can be transferred by secure encrypted e-mail.

Should the data not be supplied or the records contain missing variables, this would be raised immediately with named contacts at DVLA to investigate or resolve this.

End users of the Road Freight statistics data are continually invited, through the DfT website, contact details in statistical publications and tables and have opportunities through face-to engagement at events to comment on the statistics, give feedback on them and the processes used in their production. Any suggestions or comments made are taken seriously and responded to in a prompt manner.

Practice area 3: QA principles, standards and checks applied by data suppliers

It is a legal requirement for all vehicles used on the roads to be licensed. DVLA carry out enforcement checks on the road, as do the police, and evaders are prosecuted. Road freight operators are subject to further regulation and the Driver and Vehicle Standards Agency (DVSA) carry out additional roadside inspections of goods vehicles – therefore it would be particularly difficult to operate an unlicensed goods vehicle for any length of time. DfT estimate Vehicle Excise Duty Evasion (i.e. unlicensed vehicles) once every two years, based on a random sample of over one million vehicles seen on the road. In 2015 it was estimated that 1.4 per cent of all vehicles in use on the road were currently unlicensed, and 0.9 per cent of HGVs. The DVLA database is therefore regarded as an authoritative list of all licensed and SORN vehicles.

However, there will be some errors in some of the specific details of individual vehicles which will in turn impact on the DfT's ability to include those vehicles in its survey.

DVLA carry out annual Traceability Surveys which look at whether it is possible to trace the owner of any given vehicle, based on the address and contact details to which the vehicle is registered. These surveys are themselves subject to a range of uncertainty, but DVLA's "most likely" estimate from the 2015 survey was a 92.6 per cent traceability rate, although it is recognised that this may over-estimate the traceability rate. This means that for every 100 registered vehicles on the DVLA's database, it was not possible for DVLA to establish contact for approximately 7 of those. This does not mean that the registered keeper's details are incorrect, just that they could not be contacted. The impact of this on the road freight survey may be a slightly lower response rate and potentially some minor bias in the survey results if untraceable owners share key characteristics in their freight activity. However this is unlikely to be of major impact nor enough to distort the overall results.

As noted above, in practice area 1, 97 per cent of vehicle particulars at first registration are provided from the manufacturer's information systems via AFRL, so are highly likely to be correct – and most vehicle particulars, apart from keeper details and tax status, are unlikely to change over the life of the vehicle. DVLA continuously monitor the accuracy of manual coding of vehicle database transactions by their staff – for example the overall accuracy rate in September 2016 was 99.1 per cent (when over 1 million transactions were recorded).

Timing is also a factor on the quality of the Road Freight data. Owing to the time currently allowed for transactions to complete in the DVLA database, the quarterly snapshot, provided by DVLA is effectively 1½ - 2 months old by the time it is available for the road freight team to select vehicles from it for inclusion in the survey. Around 0.9 per cent of vehicles selected from the snapshot for inclusion in the sample cannot be subsequently matched by DVLA back to their records – an effect of the snapshot becoming outdated and vehicles become unlicensed etc. After matching, the Road Freight statistics team send questionnaires out to the vehicles owners identified on a weekly basis over the 3 month survey quarter. As such there is a lag of up to 6 months from the time the sample is initially drawn to when the last vehicles are surveyed at the end of a quarter allowing the survey to become increasingly outdated, increasing the risk of non-response. The sample is drawn as late as possible to try and achieve a balance of data which is up to date but also delivered so as to allow timely production of the statistics. Any changes made to the population, after the sample is drawn, will be reflected the time the next sample is drawn. Further information on this and the ways DfT attempts to mitigate this risk is given in the next section.

Practice area 4: Producer's QA investigations & documentation

The survey forms are dispatched in clearly marked envelopes, bearing the DfT crest and the phrase "On Her Majesty's Service" in bold font to distinguish them from "junk mail" and are accompanied with detailed guidance and supporting documents which explain each question clearly, provide definitions and explanations of terminology and contain a helpline for applicants to contact for further assistance. Keepers who fail to return the survey forms within the requested time period are issued up to 3 reminder letters and follow up phone call from a member of DfT reminding the keeper of the legal need to respond and offering assistance in completing the forms should the keeper require it. These features reduce the amount of non-response as well as the scope for errors in data provided which in turn will increase the quality of the data being provided.

The average response rate (based on over a 10 year period [2005 to 2015]) was 81% and can be taken to be relatively representative of a typical response rate. Specific response rates that the CSRGT surveys achieve are published annually on Gov.uk and can be found at:

<https://www.gov.uk/government/statistical-data-sets/rfs01-goods-lifted-and-distance-hauled#table-rfs0142>

The use of the licencing data in the Road Freight Statistics is solely to provide the DfT with a sampling frame. The main issue that may affect the quality of the freight statistics data is having incomplete records as this would mean the survey is not covering the entire population of HGVs. This would lead to inaccurate statistics and levels of freight activity being underreported.

The high degree of completeness of the DVLA registration data was covered in the previous section.

Consistency checks are carried out on the quarterly DVLA extracts when received by DfT to ensure they have been extracted, transmitted and loaded correctly, and that the data appear consistent with previous extracts. Further checks are completed when the sample is selected from the DVLA database for the CSRGT surveys to ensure the number of HGVs selected are consistent with previous numbers. Similar checks are completed when deriving grossing factors so that sample cases from the two surveys can be weighted to produce estimates for the UK-registered HGV population.

The other main characteristic of the DVLA data which may impact on the quality of the surveys is the quality of the vehicle keeper address information. This was also covered in the previous section.

Acronyms

AFRL – Automatic First Registration and Licensing
ANPR – Automatic Number Plate Recognition
CSRGT – Continuing Survey of Road Goods Transport
DfT – Department for Transport
DVLA - Driver and Vehicle Licensing Agency
DVSA - Driver and Vehicle Standards Agency
EU – European Union
FOIs – Freedom of Information (Requests)
HGV – Heavy Goods Vehicle
SORN – Statutory Off-Road Notification
UKSA - United Kingdom Statistics Authority
VED - Vehicle Excise Duty