Vehicle Excise Duty evasion statistics: Notes & Definitions

1. Introduction

This document provides guidance on the [Vehicle Excise Duty Evasion Statistics](). Comparable statistics in this series are available for 2007 onwards. They are published every 2 years by the Department for Transport (DfT).

2. Survey design and methodology

The following method has been used for Vehicle Excise Duty (VED) surveys from 2007 onwards, with only relatively minor changes from year to year. These changes are summarised in **Section 3** below.

During June in each survey year, over 1 million valid sightings of vehicle registration marks are collected at sites around the UK by contractors working on behalf of the Department for Transport. Collection is conducted at 256 sites, distributed across the regions of Great Britain and within Northern Ireland.

In Great Britain, 4 sites (1 of each of 4 different road types) were selected in each of the 49 police force areas outside London. In London, 16 sites (4 of each road type) were selected. The road types covered in each area are: Urban A roads; Rural A roads; Urban Minor roads; and Rural Minor roads. In addition, 2 motorway sites are chosen in each country or region, with the exception of London, where 4 motorway sites were selected. As well as these, 20 sites were selected in Northern Ireland to cover the different road classes and to give a large enough survey sample to give reliable results. These sites have remained the same since 2007 to make year-on-year results as comparable as possible, although the classification of some of the roads has changed over time. Therefore, the distribution of road types described above no longer applies in all cases. In particular, the DfT's method of classifying roads as rural or urban has changed since the sites were originally selected.

At each site, the contractors are required to collect data for one twelve-hour period on a weekday and for six hours on a weekend day. On weekdays, surveying was between 8.00 am and 2.00 pm and between 3.00 pm and 9.00 pm, while on weekends surveying was for one of these periods only. Weekend surveying is predetermined to give equal numbers of morning and afternoon counts and equal numbers of Saturday and Sunday counts; otherwise the surveying would be down to the discretion of the contractors. The precise location of each site was agreed between the survey contractors and the Department for Transport.

Collection was carried out by positioning a video camera at the roadside and then using Automatic Number Plate Recognition (ANPR) software at the contractor's office to collect details of the observed vehicles' registration marks from the video footage. The ANPR software automatically records each vehicle's registration mark from the video footage. Each automatically produced registration mark is then manually checked against the video footage of that vehicle. The number plates for motorcycles are collected by additional cameras at the roadside, as motorcycles do not have forward facing number plates.

The collected data are then returned to the Department where those registration marks in an invalid or foreign format or where the vehicle was noted as displaying trade plates are removed. The remaining records are passed to the Driver and Vehicle Licensing Agency (DVLA) in order to identify which registration marks are licensed when they were seen and which were unlicensed. (The DVLA took over the registration of Northern Ireland vehicles in 2014, prior to this, records are also passed to the Driver and Vehicle Agency Northern Ireland at this stage). Once these matched data are returned, a further quality assurance check was made by comparing the ANPR-produced registration mark with the video image of all vehicles reported to be unlicensed. Any misread registration marks identified through this process are removed from the survey dataset (see also **Section 5** below).

The number of records included in the final survey dataset is provided in published table [VEH0501](VEH0501).

3. Changes to the general method

Sample size reduction from 2011 onwards

In line with Government guidelines to reduce costs from 2010/11 onwards, the Department worked with the survey contractor to find ways of delivering a reduced cost survey whilst maintaining a high standard of data. In order to achieve this, all non-motorway sites in 2011 were surveyed in one direction only. Motorways continued to be surveyed in both directions. This resulted in the sample size falling from around 1.4-1.6 million to about 1.1-1.3 million vehicles. As all sites are still surveyed and there is no reason to expect a higher rate of evasion in one direction than another, the figures should be as reliable as in previous years.

Method of recording motorcycle registration marks from 2013

Prior to 2013, the manual collection of motorcycle number plates was done by the site technician manually noting down the number, or recording it on a voice recorder and transcribing it later. In 2013, digital cameras were used to record motorcycles at 16 sites. Because these sites were selected on the basis of having larger numbers of motorcycles, they accounted for almost half the motorcycle sample. The original manual methods continued to be employed on the other sites. The digital camera method was considered successful on the basis that (i) clear images resulted (ii) the images enabled full QA checking of apparent unlicensed cases (iii) fewer records were rejected as being 'obviously invalid' than in previous years (i.e. a motorcycle plate as recorded in the survey corresponding to some other vehicle type in DVLA records).

Revisions to the Northern Ireland evasion rates

An improved weighting methodology was introduced for the overall evasion rate in Northern Ireland in 2011, and retrospectively applied to all estimates from 2007, bringing the overall rates for Northern Ireland closer to those for Great Britain. The rates provided for the private and light goods vehicles and goods vehicles tax classes in Northern Ireland are not affected.

Revisions to the traffic weighting

It was realised after publication of the 2013 survey results that changes over time to the classification of road type at the survey sites had not been carried through to the survey weighting methodology. The 2013 results were therefore re-calculated using the up-to-date road type classifications, so that the survey results were weighted to the correct traffic totals. This made no difference to the overall national results, but made small differences to some of the of the vehicle type, region and road type breakdowns.

Changes in handling of low observation numbers for particular tax classes at survey sites

For the production of the 2019 estimates, a correction to the methodology has been made relating to sites at which motorcycle evasion was not measured specifically and also an adjustment for all vehicle types as to how data from sites with low numbers of observations for particular tax classes was used.

For motorcycles prior to 2017, even at sites for which motorcycles were not monitored, the overall motorcycle traffic expected at the site was still used in the calculation of overall evasion rates. This had the effect of decreasing motorcycle evasion rates for those years by roughly one percentage point. The method has now been changed to only consider sites for which motorcycles were actively monitored.

In addition, further analysis has shown that for sites with low numbers of vehicle observations in particular tax classes, there can be a significant impact of relatively few observations causing the calculated overall evasion rates to be quite volatile. As a consequence, for 2019 results, it has been decided that each survey day at each site would only be included in the analysis if there were at least 10 observations of that tax class. **Table 1** summarises the impact on the number of observations used for analysis for each tax class in 2019.

Table 1: Number of surveys used for 2019 estimates with the proportion of observations used, United Kingdom

Tax Class	Surveys used [1]	Observations used (%)
Private and Light Goods	511	100.0%
Goods	272	98.9%
Motorcycles	103	96.7%
Bus	259	94.4%
Exempt	414	99.2%
Other	48	59.7%
Total	511	100.0%

1. There were 512 surveys conducted, one weekday and one weekend count at each site, but one survey was not used at all due to only having 8 observations.

For this reason, it is considered that there is now a discontinuity in the calculated evasion rates estimated for motorcycles between the 2017 and 2019 surveys.

These methodology changes had a negligible impact on overall evasion rate estimates, and there is no discontinuity between years for these estimates or other tax classes.

Revisions of previous figures

Data for 2015 and 2017 have been revised slightly reflecting corrections to the weighting factors used in regional estimates – this mainly affects Northern Ireland data, and therefore the UK estimates – but does not make significant change to the trends displayed. Revised figures are clearly marked in published tables and no other figures were affected.

4. Derivation of evasion estimates

The exact methodology used to derive the estimates within this report is complex and is outlined in detail within a methodological review carried out by Southampton University. This review is available at:
https://www.gov.uk/government/publications/vehicles-statistics-guidance.

Derivation of rate of unlicensed vehicles in traffic

1. The rate of unlicensed vehicles in traffic for each tax class as observed at each survey site is calculated.
2. The rates in (1) are weighted by traffic flow information specific to that site, drawn from national traffic census sources.
3. These weights are used to calculate an average unlicensed rate for each tax class within each road type and region.
4. The rates in (3) are further weighted by traffic flow information specific to each road type and region.
5. These secondary weights are used to calculate an estimated unlicensed rate in traffic for each tax class.

Derivation of rate and number of unlicensed vehicles in active stock

6. The number of times each individual vehicle was seen within the survey is calculated.
7. These data are used within a statistical model based on the negative binomial distribution, using the frequency of repeat sightings, to estimate the relative mileage of licensed and unlicensed vehicles within each tax class. A table showing the results produced by this model is given in the table below. From 2010, the distribution used to estimate relative mileage for the *Goods*, *Motorcycles*, *Bus* and *Other* tax classes includes the sightings from all ANPR-based surveys (i.e. 2007 and later). This provides a larger sample which produces a more robust relative mileage (and hence 'in-stock') estimate.
8. These relative mileage estimates are combined with the rates of unlicensed vehicles in traffic to produce an estimated rate of unlicensed vehicles in active stock for each tax class.
9. The rates calculated in (3) are used together with the known number of licensed vehicles to produce the estimated number of unlicensed vehicles in active stock within each tax class.
10. The numbers calculated in (4) are summed to give the overall estimated number of unlicensed vehicles in active stock.

Derivation of revenue lost from unlicensed vehicles

11. The average cost of a yearly licence within each tax class is calculated. These costs are published in table VED0503.
12. The values from (1) are multiplied by the estimated numbers of unlicensed vehicles within each tax class to give the estimated revenue lost from unlicensed vehicles within each tax class.
13. The values calculated in (2) are summed to give the overall estimated revenue lost from unlicensed vehicles.

Confidence Intervals

Where possible, 95% confidence intervals have been provided for the most recent estimates. These are calculated using exact binomial confidence intervals as outlined in 'Tian et al, 2009, *A comparative study of confidence intervals for negative binomial proportion*. Journal of Statistical Computation and Simulation, 79(3), pp 241-249'. As they are exact confidence intervals the figures are asymmetrical. This means that the upper and lower limits are not equally spaced from the central estimate.

The confidence intervals give a range of values around the central estimate. These mean that if the survey was carried out many times (collecting a different set of number plates, perhaps at different locations or on different days), 95% of the surveys would give an evasion rate between the upper and lower value. In practice, this can be interpreted to mean that it is highly likely that the actual evasion rate in the whole population (as opposed to estimated rate for just our sample) will fall somewhere between the upper and lower limits found in the published tables. This is by no means certain, though, and there is a 5% chance that the actual evasion rate does not fall within the confidence limits produced from this survey.

The only way to know the actual evasion rate for the whole population would be to record every single vehicle movement on the roads throughout the year. Clearly this is not practically possible, either from a technical or financial point of view. Therefore, the survey sample figures presented in the table are the best estimates that can be produced given the practical restrictions.

As the 'in-stock' calculation combines the 'in-traffic' figures and relative mileage figures, it is not possible to provide 95% confidence intervals on the same basis. Therefore, upper and lower limits have been estimated by multiplying the upper limit from the 'in-traffic' figures by the upper limit from the relative mileage figures, and the same for the lower limits for both estimates. These have been used in tables VED0201, VED0202 and VED0301.

5. Effect and treatment of misread registration marks

Effect of misreads on evasion estimates

The use of ANPR (Automatic Number Plate Recognition) technology for the vast majority of the survey provides a good understanding of the effect of misread registration marks on the survey results.

To establish the nature of this effect, random samples of 1,000 to 5,000 vehicle registration marks collected through the various surveys have been taken and, for each instance, the registration mark recorded manually compared to a still video image of the vehicle as provided through the ANPR system. Where a misread was found to have occurred, the corrected registration mark was compared to a copy of the DVLA database to establish whether the actual vehicle sighted was licensed or unlicensed when it was seen.

The presence of misread registration marks, if left untreated within the survey data, would serve to artificially inflate the survey's evasion estimates. This is clearly demonstrated by the table below which shows the evasion rates that would be expected under different scenarios if misreads were left untreated within the survey data.

Table 2: Expected evasion rate without treatment of misreads

Actual evasion rate	Misread rate				
	1%	2%	5%	10%	20%
0.5%	0.7%	0.9%	1.5%	2.6%	5.0%
1.0%	1.2%	1.4%	2.0%	3.1%	5.5%
1.5%	1.7%	1.9%	2.5%	3.6%	5.9%
2.0%	2.2%	2.4%	3.0%	4.0%	6.4%

Treatment of misread registration marks in ANPR-based surveys

As every registration mark collected through the ANPR software is manually checked by the contractor within their data validation processes, it is likely that the majority of misreads are corrected before the survey data are submitted to the Department. One of the quality standards required by the Department is that a random sample of the data supplied by the contractor are >98% accurate (tested by the Department by visual inspection against the original image, supported by comparison with DVLA records).

However, the analysis shown above demonstrates that even these relatively small levels of misreads can significantly inflate evasion estimates and, therefore the Department carries out an additional stage of quality checking by comparing the registration marks and video images of all vehicles reported to be unlicensed after matching to the DVLA database. Any registration marks found to have been misread through this process are removed from the survey dataset.

While this additional process removed those remaining misreads incorrectly matching the registration mark of a separate 'unlicensed' vehicle, it will not have removed those erroneously matching 'licensed' vehicles - creating the potential for under-estimation of evasion. However, the extent of this under-estimation is likely to be negligible.

In the case of motorcycles, data were collected manually by roadside enumerators in all cases until 2011, and in about half of cases in 2013 onwards. It was therefore not possible to directly check the registration marks that had been recorded in the survey for these vehicles, although registration marks which matched to vehicles that were not motorcycles in the vehicle register were assumed to be misreads. As such, these data may contain several misreads which would be likely to have an inflationary effect on evasion. This inflationary effect may have been significant for motorcycles, but due to the relatively small number of registration marks that were collected in this way, it is very unlikely to have had a significant effect on the overall estimates for all vehicles.

Treatment of misread registration marks in previous surveys

Roadside surveys were carried out before 2007, but in these the majority of registration marks were collected manually and so could not be checked for accuracy, as no photographic images were available to do this. As a result, all evasion estimates given in surveys prior to 2007 did not allow for the effect of misreads, and so are likely to have been inflated and should therefore not be compared with those from subsequent surveys.

6. Strengths and weaknesses of the data

Overall, these statistics is considered to provide good quality estimates because:

- they are based on a purpose designed statistical survey, using methods which have been carefully developed and peer-reviewed;

- they are based on a very large sample of more than 1 million observed vehicles per survey;

- the effects of number plate misreads have been considered and additional quality assurance procedures introduced to deal with them;

- they therefore avoid many of the statistical biases that would be likely to arise from using data derived from administrative or enforcement systems;

- the results are discussed with the DVLA and are thought to be consistent with their operational experience.

However, it also has limitations of varying degrees of importance, including:

- the high cost and practical constraints of data collection mean that only relatively infrequent 'snapshots' can be taken, and surveying is concentrated in June to maximise the hours of daylight available, but this means no analysis of seasonal variation can be undertaken;

- the number of sites is limited by cost and equipment constraints, which reduces the efficiency of the very large overall sample;

- motorcycles are much more difficult to survey given their relatively small numbers, different behaviour in traffic and smaller rear-facing number plates. So the results for motorcycles are likely to be less reliable than for other vehicle types;

- the survey does not identify vehicles which are using 'cloned' number plates, i.e. plate which are valid but belong to another vehicle;

- the survey does not pick up any vehicles which are only parked on the public highway without ever being used as it only surveys those in traffic. This which would also be a breach of licensing rules.

7. Uses of the roadside survey data

The estimates of VED evasion derived from this survey are published as VED Evasion Statistics. These statistics are produced primarily to provide the Department for Transport, the motoring agencies and others with evidence to take informed view on road tax evasion and related policies. The data also provide a key source of evidence for auditing the Vehicle Excise Duty account, which is independent of the processes of collecting or enforcing payment.

The same databases of observed registration marks have also been used to derive estimates of foreign vehicle traffic and MOT non-compliance, both published as DfT statistics.

We welcome any feedback on these statistics, to ensure future releases best meet user needs. Feedback can be provided by email to vehicles.stats@dft.gov.uk.