Using MOT test data to analyse travel behaviour change - scoping report

S Cairns, RE Wilson, T Chatterton, J Anable, S Notley and F McLeod
Disclaimer

This report was produced as an output from the Research Councils UK Energy Programme research grant EP/J004758/1, led by TRL, working in partnership with the Universities of Southampton, Aberdeen and the West of England.

In December 2013, the report was updated in the light of findings from the Research Councils UK Energy Programme research grant EP/K000438/1, led by the University of Aberdeen.

The information contained herein is the property of TRL, and the project partners, and does not necessarily reflect the views or policies of the funding body or anyone else who has helped with the research.Whilst every effort has been made to ensure that the matter presented in this report is relevant, accurate and up-to-date, TRL and partners cannot accept any liability for any error or omission, or reliance on part or all of the content in another context.

When purchased in hard copy, this publication is printed on paper that is FSC (Forest Stewardship Council) and TCF (Totally Chlorine Free) registered.

Acknowledgements

Grateful thanks to the Research Councils UK Energy Programme; Daryl Lloyd, Jeremy Grove and Rachel Moyce (Vehicle Statistics Team at DfT); Aidan Naughton (VOSA, now DVSA); Data Sharing Team, DVLA; and Ben Lane (Ecolane, nextgreencar.com) for all their help with this study.

© TRL and partners, 2014
Contents

1 Introduction
   1.1 Overview 1
   1.2 Policy background 2
   1.3 Structure of the report 3

2 Data availability and future possibilities
   2.1 Existing data
      2.1.1 Overview 4
      2.1.2 The first data release 4
      2.1.3 Updates to the original data release 5
   2.2 Data possibilities, as discussed with DfT, VOSA and DVLA
      2.2.1 Overview and other users 6
      2.2.2 Future data possibilities 6
   2.3 Relevant data on vehicles of less than 3 years old 8

3 Data processing techniques
   3.1 Linking vehicles 9
   3.2 Generating intervals 9
   3.3 Using intervals to generate mileage rates 10
   3.4 Understanding data integrity 11
   3.5 Summary of data cleaning activities 12
   3.6 Understanding trends over time 12
   3.7 Limits of the existing approach 13

4 Understanding potential uses of the MOT data – exploring the impacts of vehicle age on mileages driven
   4.1 Overview 15
   4.2 Understanding how vehicle age affects vehicle use for the national vehicle fleet 15
   4.3 Spatial and temporal disaggregation 17
   4.4 Illustration of policy relevance - assessment of the car scrappage scheme 19

5 Combining MOT data with other data sources
   5.1 Overview 21
   5.2 Spatial resolution and scale 21
   5.3 Potential datasets for integration
      5.3.1 The Census 26
      5.3.2 Indices of deprivation 26
      5.3.3 MOSAIC UK or other geodemographic classification systems 27
      5.3.4 Active People Survey 27
5.3.5 Indices of accessibility 27
5.3.6 Annual energy use data 27
5.3.7 Background air pollution concentrations 28
5.3.8 Atmospheric emissions inventory 28

6 Future research activities 31
6.1 Overview 31
6.2 Objectives of the new project 31
6.3 Structure of the new project 31
1 Introduction

1.1 Overview

This report describes the main outcomes of a three month scoping study, completed between April and June 2011, funded by the Research Councils UK Energy Programme, via the EPSRC, to explore the potential to make use of data collected during the MOT test (research grant EP/J004758/1).

The MOT data were first made publicly available in November 2010, with the release of over 150 million test results. The data set is unique in providing national information about vehicle types and usage at a scale which is not available from any other source.

The main project objectives were:

- To develop software tools for the analysis of the MOT data;
- To work with the data providers (in particular, the Department for Transport (DfT), VOSA and DVLA) on assessing the maximum potential that could be made of the MOT data set whilst respecting issues such as data protection;
- To scope the application of the MOT data and the possibilities for triangulating with other data sets; and
- To develop one (or two) small-scale demonstrations illustrating its potential usefulness.

The project team comprised Dr Sally Cairns (Transport Research Laboratory and University College London); Prof. Eddie Wilson (University of Southampton, now University of Bristol); Prof. Jillian Anable (University of Aberdeen); Dr Tim Chatterton (University of the West of England); Simon Notley (Transport Research Laboratory) and Fraser McLeod (University of Southampton).

For various reasons, the publication of this report has been delayed since the completion of the scoping study. It has subsequently been updated to reflect latest knowledge as of December 2013, including insights gained from a large follow-on grant from the Research Councils UK Energy Programme (EP/K000438/1). (Through this grant, a number of the issues raised in this report are being examined in more detail, as set out in Chapter 6.)

In general, there are a lot of developments occurring on this topic, so it should be noted that all information given here is only correct to this date.

Since completion of the scoping study, there have been three further publications:


1 DVLA – the Driver and Vehicle Licensing Agency – holds details about the vehicle keeper, including their address whilst VOSA – the Vehicle and Operator Services Agency – was responsible for the MOT data. In 2013, it was announced that, in 2014, VOSA would merge with the Driving Standards Agency, into a new organisation to be called the Driver and Vehicle Standards Agency.


The Department for Transport is also continuing its own work in this area – as set out in DfT (13/6/13) ‘Experimental Statistics – Analysis of vehicle odometer readings recorded at MOT tests.’

### 1.2 Policy background

There is increasing policy interest, not least driven by the carbon reduction agenda, in initiatives that:

- Reduce car use;
- Reduce car ownership; and
- Encourage people to own cleaner, more efficient vehicles.

These initiatives range from large-scale national projects, such as the current drive to promote electric vehicles, through to much smaller-scale projects, such as schemes making it more attractive to walk or cycle in the local area.

Hence, policy makers are interested in questions such as:

- Are initiatives effective?
- What scale of effects are they having (and over what time frame)?
- How do results vary with geography? (e.g. between regions; between urban and rural areas; between different towns)
- How do results vary with different types of people? (who might be defined by socio-demographic characteristics, type of car owned etc.)
- How do patterns associated with one type of behaviour (e.g. driving) relate to other behaviours that impact on the environment (domestic energy use, recycling, etc.)?
- Is it possible to have any kind of benchmark or control data to assess what would have happened without the initiative?
- What potential future scenarios exist for energy use from transport at spatially disaggregated levels?

To assess these questions, it is necessary to have some kind of reliable data about vehicle ownership and use. In practice, existing methods of assessing this are limited. There are two main techniques used. The first involves surveys of individuals or households – which are always potentially subject to bias in terms of who chooses to reply, the seasonal effects of when surveying is done, potential misreporting (either deliberate or accidental) etc. The second is on-street traffic counts, where counters are notorious for having periods of not working, the counter network is rarely sufficiently fine-grained to be sure that all routes are covered, roadworks can cause substantial disruption, and estimation procedures are required to convert spot counts into some
A further dimension of the debate is that, as discussed in the Eddington report\(^2\), there is an increasing interest in smaller-scale, local transport measures to achieve travel behaviour change, in the belief that these tend to be relatively effective and represent particularly good value for money, compared with larger-scale schemes. However, there is a credibility issue, in that these smaller schemes are particularly unlikely to have the large monitoring budgets that would be needed to assess their worth, based on existing techniques.

The MOT data therefore provides a way of addressing some of these problems. First, it is both relatively comprehensive, and likely to be relatively accurate. Moreover, it is collected as a matter of course, and so, with appropriate analysis procedures, could potentially provide a relatively cheap way of assessing some of the issues described above. Potential applications are discussed further in Chapters 4 and 6.

### 1.3 Structure of the report

This report is structured as follows:

- Chapter 2 describes the data that were already available and a range of possibilities relating to future data availability.
- Chapter 3 describes the ways in which the data have been manipulated and processed to enable analysis.
- Chapter 4 looks at the insights from the existing data set on vehicle age and use, and their relevance to policy debates.
- Chapter 5 explores other data sets for potential integration with the MOT data.
- Chapter 6 describes our future research programme.

---

2 Data availability and future possibilities

2.1 Existing data

2.1.1 Overview

In January 2005, the Vehicle and Operator Services Agency (VOSA) introduced a computerised system for reporting MOT (roadworthiness) test results, which was fully implemented by the beginning of April 2006. Since that time, the results of approximately 35 million MOT tests have been collected every year, and stored in a database by VOSA. The Department for Transport business plan, published 8th November 2010, promised to make available the “detailed VOSA MOT data” and on 24th November 2010, data were released. The dataset consisted of the results of 150 million MOT tests from the beginning of 2005 to the spring of 2010 for vehicles in Great Britain required to take an MOT test. (Vehicles in Northern Ireland are not included).

Some fields, such as vehicle registration number (VRN), vehicle identification number (VIN) and unique VTS (vehicle test station) identities have been withheld from the published data in order to preserve anonymity. However, what remains still contains a wealth of information that is not available in any other dataset.

In addition to the results of the MOT test itself (including detailed reasons for failure), the data include:

- The vehicle odometer (mileage) reading;
- The top-level postal area (letters only from the postcode) of the VTS;
- The vehicle manufacturer, type and engine capacity; and
- The time of vehicle registration.

Meanwhile, there are clearly limits with the data set. In particular, for example, private vehicles are not required to have an MOT test until they are 3 years old; the current data set does not distinguish between vehicles kept by organisations and vehicles kept by private individuals; the amount of spatial information is very limited; and there is no information about the characteristics of the vehicle keepers. Hence, part of the project involved exploring ways to overcome these limitations.

2.1.2 The first data release

In addition to the documentation and look-up tables, the published data releases consist of two main groups of files:

- Testing data results: which contain one data record for each MOT carried out (pass or fail); and

---

3 These vehicles include motorbikes, private buses, cars and other light vehicles (including goods vehicles up to 3.5 tonnes). Detailed legislation on vehicles exempt from the MOT is set out in the Motor Vehicles Test Regulations 1981 regulation 6 (as amended), and in the Road Traffic Act 1988 Section 189. Examples of vehicles exempted from MOT testing include electrically propelled goods vehicles, track laying vehicles, vehicles constructed or adapted to form part of an articulated combination, works trucks, trailers, pedestrian controlled mechanically propelled vehicles and electrically powered pedal cycles.
• Testing data failure items: which can be cross-linked with the testing data results files, and a separate index file. These files enable a detailed examination of the advisories / reasons for failure in any particular MOT test.

This scoping study was primarily focused on the information in the testing data results files. For convenience, this is segmented into files which are organised by the year in which the given MOT tests took place. Each line of each of these files describes a single MOT test. For example, from the 2007 test results file, we had:

1738409|2007-07-07|4|PR|P|85436|MK|VAUXHALL|ASTRA CLUB 8V|WHITE|P|1598|1999

This gives 13 fields:

• "1738409": the identifier of the individual test (a unique integer across all the files provided).
• "2007-07-07": the date of the test.
• "4"; "PR"; "P": these represent the class of the vehicle, the test type and the result.
• "85436": the odometer reading reported by the Vehicle Testing Station (VTS).
• "MK": the two-digit postcode area of the VTS where the test was performed – which may be different to where the vehicle is registered.
• "VAUXHALL"; "ASTRA CLUB 8V"; "WHITE": the make, model and colour of the vehicle according to the DVLA database.
• "P"; "1598": fuel type (here petrol), and engine capacity.
• "1999": year of first use.

An unadvertised feature of the data set was a potential opportunity to follow vehicles from year to year. This was developed and exploited during the scoping study, but has since been replaced by an official linking field, as described below.

2.1.3 Updates to the original data release

Since the original data release, the dataset has been upgraded in two ways:

• More detail has been provided in the ‘first registration’ field (specifically, the year of registration has been replaced with the day of registration); and
• An anonymised vehicle identifier has been introduced, to enable vehicles to explicitly be tracked year on year.

Further releases have also taken place, so that data files are now available for tests up to 9th August 2013.

During discussions, it was noted that the date of first registration is usually the date when the vehicle is first sold to a customer (regardless of whether it has been part of a bulk registration). Mileage readings are usually zero, or very close to zero, at that point, since a significant mileage reading on the odometer would reduce the vehicle value.
2.2 Data possibilities, as discussed with DfT, VOSA and DVLA

2.2.1 Overview and other users

During the project, helpful discussions were held with VOSA, DfT and DVLA. Through these discussions, it emerged that current users of MOT data fall into a number of categories. These include:

- Companies involved in checking the provenance of vehicles (such as HPI and CDL (Cheshire Datasystems Ltd));
- Companies involved in market research (such as Experian); and
- Manufacturers and related organisations involved in understanding vehicle defects (such as Kwik Fit).

The SMMT (Society of Motor Manufacturers and Traders) has been particularly involved in understanding the potential of the data, and developing the tools required to use it.

DfT report that making the MOT data available has resulted in various new applications (see, for example, [http://good-garage-guide.honestjohn.co.uk/mot-results/](http://good-garage-guide.honestjohn.co.uk/mot-results/)).

DfT also report that there has been considerable interest in the new MOT data from the research community, particularly were the data to be made available with more spatially specific information. DVLA data is also specifically used by the police.

The DVLA is primarily responsible for all the information which relates to the registered keeper of any vehicle. They have a number of specific data sets derived from the information that they hold, which they make available (an ‘anonymised data set’, which is primarily used by marketing companies; a ‘bulk data set’, which is primarily used for vehicle verification checks; and a ‘mileage data’ release, which is also primarily used for vehicle verification). Details of variables of potential relevance to this project are discussed in the next section.

2.2.2 Future data possibilities

During the discussions, a number of further data possibilities were explored, as follows.

Obtaining more information on vehicle location

Obtaining more information about vehicle location (or more specifically, the location of the registered keeper of any vehicle) is probably the most important additional information for future research, and was therefore explored in some detail.

DVLA clarified that there are concerns about vehicle theft if thieves were able to identify the location of unusual vehicles. Consequently, the level of resolution of spatial information which they release usually relates to the number of vehicles of the same make/model in any one location. Specifically, DVLA usually provides spatial location information for a given vehicle, providing there are at least 200 vehicles of the same make and model registered to that location. In practice, this means that the level of spatial resolution usually provided is relatively coarse.

There are several possible ways forward. The first relates to potentially obtaining additional information subject to appropriate security arrangements. The second would involve suppressing information in some of the other vehicle fields, such that it would be impossible to cross-reference between the existing MOT data release, and a release with
more specific geographic information. For example, this might include changing the vehicle information to a generic vehicle class, and rounding the odometer readings. Specific discussions would be required to understand the best way to access the data.

*Additional mileage readings obtained at points of change of use*

Currently, people provide information about mileage readings when vehicles change use (including within the first three years of a vehicle's life) on a voluntary basis. DVLA currently provide a summary of data received on a daily basis to interested parties (for a fee). This is not formally stored or consolidated by DVLA – and would be difficult to incorporate with other data sets at present.

*DVLA data on date of birth of keeper or other driver variables*

Although DVLA hold various sorts of information about drivers, it is currently impossible to link the driving licence data and vehicle data together as driving licence number is not needed for registering a vehicle. Hence, linking driver information with vehicle information is not possible. Moreover, some keepers will not even have a UK licence, as a keeper does not need to be able to drive to have a vehicle registered in their name, or they could have a driving licence from a different country.

*Identifying vehicles that are part of the same household*

Only DVLA’s data suppliers (IBM/Fujitsu) hold full address details for vehicle keepers, together with a separate version of the database held by the police. (The DVLA only holds postcode data.) It is possible that, at some point in the future, it might become possible to ask for a revised data set, where vehicles from the same household are flagged as such - or even where vehicles are simply tagged depending on the number of other vehicles registered at the same address. However, under current arrangements, obtaining a data set in this form would require a substantial payment.

*Understanding mileages at the time when vehicles leave the vehicle fleet*

The DVLA generates a quarterly report of the vehicle stock. From this, it would be possible to infer when vehicles leave the vehicle stock (to a 3 month period). However, there is not currently a readily available list of these vehicles that is produced each quarter. There might also be information available from when owners of vehicles that have been taken off the road apply for tax refunds, though the nature and availability of this information is currently unclear.

*Vehicle emissions*

With approval from the DVLA, it would be possible for DfT to add manufacturer values for CO₂, NOx and PM₁₀ for each vehicle to the existing data set⁴.

---

⁴ There is considerable debate about the validity of manufacturer emission values, however they provide the most readily available figures for use at the present time. At some point, it may also be valuable to understand the range of emission values over different driving cycles.
Identifying company-owned versus privately-owned vehicles

In the quarterly statistics that DfT obtain from DVLA about the vehicle stock, there is a field which gives the registered keeper title. This is either personal (e.g. Mr, Ms, Dr, Rev etc) or business related (e.g. Company, Messrs etc.). Using this field, it would be possible to add a field which indicates whether the vehicle was in personal or business ownership, at the time of the previous or subsequent quarterly vehicle stock update.

Vehicle class

As part of this project, a look-up table was created, to enable conversion of vehicle make/model information into vehicle class. This started from a classification system and table of cars currently in production kindly supplied by Nextgreencar.com. This table was then modified and expanded, given that the MOT data uses more specific make/model information, and given the need to code cars no longer in production. Approximately half of the car types could be coded using an automated matching function (working from the Nextgreencar table), whilst the remainder were manually coded by a car expert at the University of Southampton. In total, 32,000 types of cars were coded. The table was not subject to independent verification, but was adequate for some exploratory analysis.

In the future, there are two possibilities.

First, SMMT have a process for classifying vehicles, which is relatively widely used.

Second, DVLA use a ‘body type’ classification field, which would enable vehicles to be identified as being in certain classes.

2.3 Relevant data on vehicles of less than 3 years old

Primarily because of concerns about ‘clocking’ (i.e. falsifying odometer readings to give artificially low mileage readings), the British Vehicle Rental and Licensing Association set up a database, whereby rental and leasing companies supplied them with data about vehicles at points of change in use. This data is passed to three third parties, who then provide a vehicle verification service for those who need it. At present, this includes about 25 million records, relating to about 10 million vehicles. It is being expanded to include data for vehicles at times when they are serviced.

For future work, this dataset could prove invaluable for obtaining insights into mileages travelled by vehicles of less than three years old. Our understanding is that this dataset includes vehicle identification information and registration date; vehicle make and model; and an odometer reading together with a relevant date.

Access to this data would require further discussion.

---

There are various vehicle classification systems, which attempt to categorise the very large number of different makes and models of car into a small number of relatively homogenous categories – using labels such as ‘small family car’, ‘large family car’ etc.
3 Data processing techniques

3.1 Linking vehicles

This chapter describes analysis which has been undertaken with the existing data set.

To estimate mileage rates from odometer readings, it is necessary to link tests for the same vehicle together over time.

At the time of the scoping study, the available MOT data did not include a field labelling individual vehicles explicitly, but we were nevertheless able to link tests as shown in Table 3.1, where the criteria for a block of consecutive records to be considered the same vehicle were that they matched on (i) make, (ii) model, (iii) colour and (iv) year of first use.

Having grouped the data on this basis, an example of the resultant information is given in Table 3.1. (for a Vauxhall Astra, first registration 1999). From this information, we may infer that the vehicle drove 85,436-71,803 = 13,633 miles in the 375 days from 2006-06-27 to 2007-07-07, at an average rate of 36.35 miles per day. Moreover, the test centre location data indicate that it most probably resided in the Milton Keynes area (postcode MK).

However, we cannot say how those 13,633 miles were spread amongst the 375 days, and in practice that distribution is likely to be highly non-uniform. Moreover, we cannot say where those miles were driven.

<table>
<thead>
<tr>
<th>Date</th>
<th>Test Outcome</th>
<th>Test Code</th>
<th>Odometer Reading</th>
<th>Make</th>
<th>Model</th>
<th>Colour</th>
<th>Year of First Use</th>
<th>Location Code</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-06-27</td>
<td>P</td>
<td>4</td>
<td>71803</td>
<td>VAUXHALL</td>
<td>Astra Club 8V</td>
<td>WHITE</td>
<td>P 1598</td>
<td>MK</td>
<td></td>
</tr>
<tr>
<td>2006-06-27</td>
<td>F</td>
<td>4</td>
<td>71803</td>
<td>VAUXHALL</td>
<td>Astra Club 8V</td>
<td>WHITE</td>
<td>P 1598</td>
<td>MK</td>
<td></td>
</tr>
<tr>
<td>2007-07-07</td>
<td>F</td>
<td>4</td>
<td>85436</td>
<td>VAUXHALL</td>
<td>Astra Club 8V</td>
<td>WHITE</td>
<td>P 1598</td>
<td>MK</td>
<td></td>
</tr>
<tr>
<td>2007-07-07</td>
<td>F</td>
<td>4</td>
<td>85436</td>
<td>VAUXHALL</td>
<td>Astra Club 8V</td>
<td>WHITE</td>
<td>P 1598</td>
<td>MK</td>
<td></td>
</tr>
<tr>
<td>2008-07-07</td>
<td>F</td>
<td>4</td>
<td>96592</td>
<td>VAUXHALL</td>
<td>Astra Club 8V</td>
<td>WHITE</td>
<td>P 1598</td>
<td>MK</td>
<td></td>
</tr>
<tr>
<td>2008-07-07</td>
<td>F</td>
<td>4</td>
<td>96592</td>
<td>VAUXHALL</td>
<td>Astra Club 8V</td>
<td>WHITE</td>
<td>P 1598</td>
<td>MK</td>
<td></td>
</tr>
<tr>
<td>2009-08-28</td>
<td>F</td>
<td>4</td>
<td>107094</td>
<td>VAUXHALL</td>
<td>Astra Club 8V</td>
<td>WHITE</td>
<td>P 1598</td>
<td>UB</td>
<td></td>
</tr>
<tr>
<td>2009-08-28</td>
<td>F</td>
<td>4</td>
<td>107094</td>
<td>VAUXHALL</td>
<td>Astra Club 8V</td>
<td>WHITE</td>
<td>P 1598</td>
<td>UB</td>
<td></td>
</tr>
</tbody>
</table>

This block of data gathers together four consecutive years of MOT tests (pass and fail) for an individual vehicle, in reverse-time order. We may read off the reported odometer history at each of the test dates: 2006-06-27, 71,803 miles; 2007-07-07, 85,436 miles, etc..

3.2 Generating intervals

Once the tests were sorted in this manner, we looked at the intervals between consecutive tests for the same vehicle, and these intervals formed the basic unit of our analysis. For example, the eight tests in Table 3.1 yield seven inter-test intervals, as shown in Table 3.2.

In Table 3.2, the test outcomes (pass, fail etc.) are removed, as they are not of primary interest. This data may be linked with vehicle-specific fields (namely |4|VAUXHALL|ASTRA CLUB 8V|WHITE|P|1598|1999) which do not vary from test to test.
In practice many intervals (e.g., intervals 1, 3, 5 and 7 in Table 3.2) have a span of zero days, because they correspond to a failed test and re-test on the same day. The mileage rate is thus either infinite or undefined, and such intervals were removed from the data before analysis continued.

### Table 3.2: Inter-test intervals generated from Table 3.1

<table>
<thead>
<tr>
<th>Interval</th>
<th>First test</th>
<th>Second test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Date $t_1$</td>
<td>Miles $x_1$</td>
</tr>
<tr>
<td>1</td>
<td>2006-06-27</td>
<td>71803</td>
</tr>
<tr>
<td>2</td>
<td>2006-06-27</td>
<td>71803</td>
</tr>
<tr>
<td>3</td>
<td>2007-07-07</td>
<td>85436</td>
</tr>
<tr>
<td>4</td>
<td>2007-07-07</td>
<td>85436</td>
</tr>
<tr>
<td>5</td>
<td>2008-07-07</td>
<td>96592</td>
</tr>
<tr>
<td>6</td>
<td>2008-07-07</td>
<td>96592</td>
</tr>
<tr>
<td>7</td>
<td>2009-08-28</td>
<td>107094</td>
</tr>
</tbody>
</table>

Note that there is redundancy in this data format because most tests appear twice, both as a first test (i.e., 'left-hand end') and as a second test (i.e., 'right-hand end'). However, the advantage of this redundancy is that for a given search date, the set of all intervals which 'span' that date can be found without table look-ups.

The main computational task that we performed was to construct a very large number of such intervals. The methods used to identify contiguous same-vehicle blocks of data yielded approximately 113 million intervals. A suite of software tools were written in MATLAB for constructing, querying and filtering intervals.

### 3.3 Using intervals to generate mileage rates

In order to generate the mileage rate for a particular point in time, we used all intervals that straddled the chosen observation date – in particular, those where the left hand test date fell on or before the observation date and the right hand test date fell strictly after the observation date. This is illustrated in Figure 3.1. For each interval, we computed the difference in odometer readings and the difference in test dates. Dividing the former by the latter produced a mileage rate for each interval, and the average over intervals was taken as the overall average mileage rate.
To run disaggregated queries, for example based on vehicle age or postcode region, only intervals that matched the chosen criteria were used. When disaggregating spatially, it is important to note that the left hand test may not be in the same postcode region as the right hand test. Such an interval can be weighted by one half when computing the overall average mileage rate for a single postcode region, in order to avoid double counting.

Intervals involving a test with an anomalous mileage reading were also filtered out - in particular when the implied mileage rate for the interval was negative or implausibly large, which could skew the results, as described below.

### 3.4 Understanding data integrity

As part of the analytical process, various checks were performed on the data, some of which were facilitated by having reorganised the data into intervals. The results are shown in Table 3.3 – based on the 2010 data set (with over 150 million tests, and a corresponding number of intervals). The implication is that the data are fairly clean.

#### Table 3.3: Insights into the accuracy of the MOT data

<table>
<thead>
<tr>
<th>Type of issue</th>
<th>Records affected</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing ‘first year of registration’ data</td>
<td>166,618 tests</td>
<td>0.11</td>
</tr>
<tr>
<td>Zero odometer reading</td>
<td>1,597,644 tests</td>
<td>1.05</td>
</tr>
<tr>
<td>Anomalous odometer reading (e.g. 99999)</td>
<td>10,284 tests</td>
<td>0.001</td>
</tr>
<tr>
<td>Odometer reading which was lower than the previous reading</td>
<td>2,608,382 intervals</td>
<td>1.71</td>
</tr>
<tr>
<td>Pair of odometer readings which indicate an unfeasibly high mileage between tests*</td>
<td>665,289 intervals</td>
<td>0.44</td>
</tr>
</tbody>
</table>
* Errors in the recording of the odometer reading at the VTS will result in erroneous mileage rates. Manual inspection of the sequences of tests for individual vehicles seemed to indicate that a common problem was either the omission of a digit, or the inclusion of a spurious additional digit, in the odometer reading - which results in a pair of intervals which have respectively abnormally high and abnormally low (or negative) odometer readings. Unfortunately, there is no fool-proof method for identifying these tests and removing them from the analysis. The challenge, therefore, is to remove tests which imply such an abnormally high mileage that they would skew the statistics for the overall vehicle population. Thus, the ad hoc method that we used was to discard all intervals where the implied mileage rate was above a given threshold (in miles per day). We then performed tests to examine the dependency between population-level statistics and this threshold. Clearly, if the threshold is set too low, then genuine examples of high mileage vehicles will be discarded - which is not desirable. So the aim was to choose a threshold which was somewhat above what we considered to be a feasible maximum mileage. In practice, we found that very few vehicles drive more than 100 miles per day on average (although some, such as taxis, do). So our approach was to be conservative and experiment with thresholds some way above this, for example 500 miles per day. Most of our statistics are based on a threshold of 720 miles per day, corresponding to continuous 30 mph driving for 24 hours. However, we found that population averages were not very sensitive to the precise choice of the threshold. For example, the DfT has chosen a threshold of 100,000 miles per year (equivalent to about 274 miles per day). The task, going forward, is to agree a standard choice so that statistics computed by different parties are comparable – or to find a way of using longer sequences to identify suspect intervals, and thereby removing them (hence obviating the need to use a somewhat arbitrary threshold).

3.5 Summary of data cleaning activities

In summary, then, as described above, in performing all calculations, the following data filters were used:

- Exclude negative and zero mileages;
- Exclude mileage readings where the start or end odometer reading was either zero or 999999;
- Exclude mileage readings of 720 miles per day (equivalent to an annual mileage of 262,800);
- Exclude tests within 30 days of each other; and
- Exclude tests with intervals of 500 days or more.

Tests within 30 days were excluded on the basis that they are most likely to relate to retesting. Tests with intervals of 500 days or more were excluded on the basis that the vehicle is likely to have been off the road for a non-trivial part of the test interval.

The limits used were chosen on the basis of balancing what we considered to be reasonable gaps between tests and re-tests, with an examination of the impact of these thresholds on the population-level statistics.

3.6 Understanding trends over time

The methods described above generate an average mileage rate for a particular point in time. By disaggregating by the variable of interest, this can then be used for various cross-sectional analyses.

However, it is also often interesting to examine trends over time. There are a number of problems with using the existing data set to do this.
The first relates to the fact that the ‘straddle’ rate at any one time period is actually an average of the two year period that it is central to. This means that any seasonal effect occurring due to people driving by different amounts during the year is averaged out.

Second, closer examination of the MOT data indicates that vehicle mileage also varies with the time of year at which the vehicle is tested. This may reflect the fact that car purchasing is not uniform over the year, and that different types of people may buy cars at different times\(^6\).

Moreover, because vehicles of different ages are driven different amounts (as discussed in Chapter 4), it is important to understand whether the nature of the vehicle fleet remains relatively constant over the time period of interest.

More specifically, at the time of the scoping study, the dataset available covered all tests 01/04/2006 – 01/04/2010. Assuming every vehicle is tested yearly, then only observation dates in the range 01/04/2007 through 01/04/2009 will include intervals for all active vehicles (of more than three years old). Outside this range, many intervals will be ‘missing’ because the first or last test in the interval is not included in the dataset. This appears to generate anomalous results.

Figure 3.2 provides a plot of the straddle rates, indicating the first and last year problem. To ensure that this was not some peculiar feature of the particular time period studied, other plots were also created where an earlier cut-off date was used. These demonstrated that the same anomalous changes to the trend line occurred.

The same graph is shown in Figure 3.3, but split into groups by year of first registration. Using this approach avoids the ‘beginning/end’ of the time period effect, but any apparent trends will at least partly be due to a vehicle aging effect. (Moreover, for the older vehicles, the graphs are ‘noisy’ due to smaller sample sizes for these groups, meaning that any apparent trend may not be significant.)

Understanding how to adjust for seasonal variation in driving quantities (for all vehicles); adjusting for differences in the seasonal pattern of vehicle purchasing; and, related to this, differences in the composition of the vehicle fleet at different times of the year, and in the first/last year of available data; are all issues identified as needing further research.

### 3.7 Limits of the existing approach

The approach developed in this scoping study is particularly useful for cross-sectional analysis, comparing different parts of the country. In terms of finding new ways of manipulating the data, future research priorities include:

- Looking at longer strings of test results for individual vehicles;
- Moving from rates which are, in effect, an average for a two-year time period, to finer temporal intervals; and
- Understanding the impacts of seasonal variability in order to adjust the data to look at trends up to the most recent MOT tests available.

---

\(^6\) In relation to this point, it should be noted that the time of year at which a vehicle is tested tends to echo the date at which it was first registered, because its first test is three years after that date, and subsequent tests typically occur at one year intervals.
Figure 3.2: Mean daily mileage (2 year averages) at different times

Shaded areas indicate where large numbers of intervals are missing. The shaded period in 2006/7 is greater than in 2009/2010, because full computerisation of test results was only complete by April 2006.

Figure 3.3: Mean daily mileage (2 year averages) at different times, for vehicles grouped by year of first registration (YOFR)

Shaded areas indicate where large numbers of intervals are missing. The shaded period in 2006/7 is greater than in 2009/2010, because full computerisation of test results was only complete by April 2006.
4 Understanding potential uses of the MOT data – exploring the impacts of vehicle age on mileages driven

4.1 Overview

Part of the purpose of the scoping study was to investigate whether the MOT data set was likely to yield information that would be of value to policy makers, or other stakeholders over and above information available from other sources of data. As set out in Chapter 6, there are many potential applications for this data, which will be explored in our follow-on study. Meanwhile, this chapter provides an illustration of the types of insights that are available for just one topic – namely, the effects of vehicle age on vehicle mileage.

All future projections of emissions from transport rely on making assumptions about the composition and use of the vehicle fleet. For example, the National Atmospheric Emissions Inventory relies on inputs from Boulter (2009)[7], which used a snapshot of MOT data for 2006/7 to look at the relationship between vehicle age and use. The analysis carried out for this project complements that work, making it possible to assess whether the relationships identified there hold true or are changing as the age composition of the vehicle fleet alters, as changes in economic conditions affect car purchasing habits, as people change how they use cars (for example, if older cars are driven further than they used to be; if households own more cars, but use each one less; etc.).

The following sections illustrate the types of analysis that are possible with the MOT data.

4.2 Understanding how vehicle age affects vehicle use for the national vehicle fleet

Figure 4.1 and Table 4.2 illustrate how vehicle age affects vehicle use, based on calculations using all available intervals in the data set, whilst excluding data about vehicles registered in 2008, 2009 and 2010, on the basis that the sample sizes in the data release used for this project were too small, and unlikely to be representative[8].

---


[8] In general, there was some variability in sample sizes for each year, which requires further investigation and understanding.
Data generated by only including tests separated by an interval of at least 30 days, and up to 1825 days, with an average implied daily mileage of less than 720 miles per day between tests.

Table 4.1: Daily mileage by year of first registration

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-1990</td>
<td>8.32</td>
<td>3.03</td>
</tr>
<tr>
<td>1990</td>
<td>13.32</td>
<td>8.29</td>
</tr>
<tr>
<td>1991</td>
<td>14.76</td>
<td>9.73</td>
</tr>
<tr>
<td>1992</td>
<td>16.65</td>
<td>11.33</td>
</tr>
<tr>
<td>1993</td>
<td>17.05</td>
<td>12.47</td>
</tr>
<tr>
<td>1994</td>
<td>17.57</td>
<td>13.37</td>
</tr>
<tr>
<td>1995</td>
<td>18.08</td>
<td>14.13</td>
</tr>
<tr>
<td>1996</td>
<td>18.67</td>
<td>14.90</td>
</tr>
<tr>
<td>1997</td>
<td>19.14</td>
<td>15.57</td>
</tr>
<tr>
<td>1998</td>
<td>19.83</td>
<td>16.42</td>
</tr>
<tr>
<td>1999</td>
<td>20.56</td>
<td>17.25</td>
</tr>
<tr>
<td>2000</td>
<td>21.25</td>
<td>17.90</td>
</tr>
<tr>
<td>2001</td>
<td>22.70</td>
<td>19.28</td>
</tr>
<tr>
<td>2002</td>
<td>24.44</td>
<td>20.81</td>
</tr>
<tr>
<td>2003</td>
<td>25.22</td>
<td>21.24</td>
</tr>
<tr>
<td>2004</td>
<td>25.94</td>
<td>21.71</td>
</tr>
<tr>
<td>2005</td>
<td>26.71</td>
<td>22.16</td>
</tr>
<tr>
<td>2006</td>
<td>27.35</td>
<td>22.09</td>
</tr>
<tr>
<td>2007</td>
<td>31.00</td>
<td>24.28</td>
</tr>
</tbody>
</table>
As illustrated above, it is clear that average vehicle mileage does decrease with vehicle age, and the scale of the decrease is non-trivial. Moreover, the scale of this decrease can be quantified.

**Table 4.2: Comparing data from cars with a 10 year age difference**

<table>
<thead>
<tr>
<th>Year of first registration</th>
<th>Mean daily mileage</th>
<th>Median daily mileage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>21.25</td>
<td>17.90</td>
</tr>
<tr>
<td>2001</td>
<td>22.70</td>
<td>19.28</td>
</tr>
<tr>
<td>2002</td>
<td>24.44</td>
<td>20.81</td>
</tr>
<tr>
<td>2003</td>
<td>25.22</td>
<td>21.24</td>
</tr>
<tr>
<td>2004</td>
<td>25.94</td>
<td>21.71</td>
</tr>
<tr>
<td>2005</td>
<td>26.71</td>
<td>22.16</td>
</tr>
<tr>
<td>2006</td>
<td>27.35</td>
<td>22.09</td>
</tr>
<tr>
<td>2007</td>
<td>31.00</td>
<td>24.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year of first registration</th>
<th>Mean daily mileage</th>
<th>Median daily mileage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>13.32</td>
<td>8.29</td>
</tr>
<tr>
<td>1991</td>
<td>14.76</td>
<td>9.73</td>
</tr>
<tr>
<td>1992</td>
<td>16.65</td>
<td>11.33</td>
</tr>
<tr>
<td>1993</td>
<td>17.05</td>
<td>12.47</td>
</tr>
<tr>
<td>1994</td>
<td>17.57</td>
<td>13.37</td>
</tr>
<tr>
<td>1995</td>
<td>18.08</td>
<td>14.13</td>
</tr>
<tr>
<td>1996</td>
<td>18.67</td>
<td>14.90</td>
</tr>
<tr>
<td>1997</td>
<td>19.14</td>
<td>15.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Difference in mileages</th>
<th>Ratio* according to the difference between the means</th>
<th>Ratio* according to the difference between the medians</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.6x</td>
<td>2.2x</td>
<td>1.6x</td>
</tr>
<tr>
<td>1.5x</td>
<td>2.0x</td>
<td>1.5x</td>
</tr>
<tr>
<td>1.5x</td>
<td>1.8x</td>
<td>1.6x</td>
</tr>
<tr>
<td>1.5x</td>
<td>1.7x</td>
<td>1.6x</td>
</tr>
<tr>
<td>1.5x</td>
<td>1.6x</td>
<td>1.5x</td>
</tr>
<tr>
<td>1.5x</td>
<td>1.6x</td>
<td>1.6x</td>
</tr>
</tbody>
</table>

* The ratio figure given indicates the difference in mileage between the older and younger vehicles. For example, on average, vehicles registered in 2006 do a mean average of 27.35 miles per day compared with 18.67 miles per day for vehicles registered in 1996 – i.e. approximately 1.5 times the amount of miles, or 50% more. There is a similar difference when comparing the medians.

In general, therefore, these data indicate that vehicle age has a very substantial impact on mileage driven, with newer vehicles typically driven at least 50% more than those which are 10 years older. Moreover, the median data suggests that these differences become even greater as vehicles get very old (i.e. 20 years+).

There are a range of possible explanations for these observed effects. For example, it may be that vehicles are used less as they get older because of aging; that every new generation of vehicles is driven slightly more than the previous one (perhaps because of enhancements in driving performance); and/or that higher mileage vehicles are retired earlier due to wear.

### 4.3 Spatial and temporal disaggregation

The MOT dataset also makes it possible to examine the impacts of vehicle age in different types of area, such as a comparison of vehicle usage in rural versus urban areas.

As an example, during the scoping study, the postcode areas KY (Kirkcaldy, Scotland) and W (West London) were compared, as the most extreme opposites of rural and urban geographies. (It was a significant drawback that the spatial resolution is so coarse, since KY includes the towns of Dunfermline and Kirkcaldy itself, each of which has a population
of approximately 50,000. KY is therefore not entirely rural, but rather it was the most rural postcode area available.)

The analysis took place based on an observation date of 1st Jan 2008, with intervals used that (i) matched the respective postcode areas at both the ‘left-hand’ and ‘right-hand’ tests; (ii) straddled 1st Jan 2008; and (iii) were disaggregated according to vehicle age, focussing on 1992, 1996, 2000, and 2004 as years of first registration.

Tables 4.3 and 4.4 present sample sizes (N) and average mileage rates (r).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>KY</td>
<td>818</td>
<td>4820</td>
<td>10555</td>
<td>12191</td>
</tr>
<tr>
<td>W</td>
<td>1020</td>
<td>3137</td>
<td>5777</td>
<td>4748</td>
</tr>
</tbody>
</table>

Meanwhile, Figure 4.2 displays the results in the form of the distributions of the (individual vehicle) average mileage rate across the relevant populations.

As might be expected, the average mileage rates are higher in the rural postcode, and, moreover, they decline with the age of vehicles irrespective of the postcode area. Meanwhile, this analysis also makes it possible to explore potential dependencies between vehicle age and geography, and to parameterise these relationships if required. Notably, for newer vehicles, the higher mileage in KY displays itself via a ‘right-shift’ of the distribution. However, the difference between the usage patterns of rural and urban vehicles tends to disappear as the vehicles age. We speculate that this trend might be due to patterns in household vehicle ownership — for example, are the older vehicles mainly second vehicles that are used for certain types of journeys or that are owned by particular types of people? Could it be the case that mileage for certain types of journeys, or mileage by particular social groups, is similar in rural and urban postcodes? Understanding more about this topic was beyond the remit of the scoping study, but hopefully indicates how insights from the MOT dataset might generate new topics for analysis.
Figure 4.2: Comparison of the mileage rate distribution in the rural postcode area KY (Kirkcaldy) versus the urban postcode area W (West London), as a function of vehicle age

Frequencies are computed in intervals of width 5 miles per day. Owing to the relatively small sample sizes in this example, grouping with narrower intervals tends to be dominated by noise.

4.4 Illustration of policy relevance - assessment of the car scrappage scheme

It is also possible to consider how insights on the effects of vehicle aging might affect policy analysis, by considering the example of the 2009/10 Scrappage Incentive Scheme.

The UK ‘Scrappage Incentive Scheme’ (SIS) was introduced in the UK between May 2009 and March 2010. Under the scheme, an incentive of £2,000 was offered for the purchase of a new car, if a vehicle of 10 years + was scrapped. The scheme was reported to have generated nearly 400,000 new car registrations over the period, or about 20% of all new cars registered in the UK (SMMT, 2010⁹).

Although the scheme was primarily introduced to support the automotive sector, there were various potential environmental benefits from the scheme. In particular, the scheme had the potential to remove the most polluting vehicles from the road, meaning that the miles that those vehicles would have done were done in vehicles with much lower emissions. Various analyses of the potential environmental impacts of the UK car scrappage scheme have been carried out (including CCC 2009\textsuperscript{10}, SMMT 2011\textsuperscript{11} and IFS 2010\textsuperscript{12}).

In understanding potential impacts, one concern relates to the ‘rebound’ effect – i.e. that people may drive newer vehicles more than the old vehicles that they replace, for a number of reasons (more pleasant ride; greater confidence in the vehicle’s robustness; less concern about maintenance; excitement of new vehicle; cheaper fuel costs etc.).

Estimating the scale of the rebound effect is inevitably problematic – since it requires assumptions about what people would have done otherwise. (For example, would they have traded up their vehicle for a newer one anyway? Are people who scrapped their cars typical of the people who normally buy new cars and/or likely to develop their driving habits?) A further complicating factor is that people may have altered the type of vehicle purchased. (For example, if larger vehicles were replaced with smaller vehicles, any increase in mileage would potentially have been cancelled out by reductions in engine size.)

Notably, in their analysis, both the CCC and the SMMT assumed that vehicle mileage was unchanged between the old cars scrapped and the new cars purchased. Meanwhile, the IFS analysis did make an assumption about differences in mileage, based on evidence from the National Travel Survey (c.f. IFS, 2010, p73). They assumed that drivers of cars of over 10 years old travelled an average of 10,600 km a year, and those driving new cars drove an average of 13,100 km a year – implying that the newer vehicles were driven about 24% more than their 10 year older counterparts.

Meanwhile, the mileage differences between old and new cars suggested by the MOT analysis is considerably greater (i.e. potentially in the order of 50% more for cars that are 10 years newer). This indicates that understanding the scale and nature of any rebound effect, and the potential applicability (or otherwise) of such figures to the change in vehicles engendered by the scrappage scheme, becomes critical to accurate assessment of the carbon savings from the policy, and probably deserves greater consideration than the issue is given in most of the analyses to date. In particular, it becomes important to understand how far there is a new (or newer) car effect on travel decisions, or whether any such apparent effect is, in practice, a reflection of changes in travel decisions occurring for other reasons.

\[\begin{align*}
\text{CCC (2009) } & \text{Meeting Carbon Budgets – the need for a step change. Progress report to Parliament Committee on Climate Change.} \\
\end{align*}\]
5 Combining MOT data with other data sources

5.1 Overview
For research purposes, the greatest value of the MOT data is likely to emerge when it is combined with information from other sources.

Due to significant concerns about confidentiality, the MOT dataset could not be made available at a household level. Instead, one way forward would be to make the information available at a more detailed spatial scale, thereby enabling information from the MOT dataset to be analysed in parallel with that from other datasets, for particular areas. Key issues, therefore, are the spatial scale at which the MOT data might be released, and other data sets with which the MOT data could potentially be combined. Each of these topics is discussed further below.

5.2 Spatial resolution and scale
Currently, as described in Chapter 2, the DVLA will provide spatial location information for a given vehicle, providing there are at least 200 vehicles of the same make and model registered to that location. In practice, this usually means that locational information is limited to postcode ‘area’ (see Table 5.1 and Figure 5.1).

In contrast, many other data sets provide more specific locational information. The smallest area for which data is normally made available is the census ‘Lower Layer Super Output Area’ (LLSOA). The aim would therefore be to obtain geographical information about the MOT data which would enable analysis at this level. For clarity, Tables 5.1 and 5.2 illustrate the different geographical levels at which data are typically made available, together with the typical number of households usually contained in each category. Figures 5.1 and 5.2 illustrate the issue of scale graphically. Lower Layer Super Output Areas may contain, on average, the same number of households as around 25 postcode units, and around a sixth of the number of households of a postcode sector.
### Table 5.1: Description of postcode levels (UK)

<table>
<thead>
<tr>
<th>Postcode Level</th>
<th>Code</th>
<th>Number of Divisions (approx.)</th>
<th>No. of Households per Division (approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
<td>XX e.g. BS</td>
<td>124</td>
<td>226,000</td>
</tr>
<tr>
<td>District</td>
<td>XXNN e.g. BS5 or BS55</td>
<td>3,000</td>
<td>9,400</td>
</tr>
<tr>
<td>Sector</td>
<td>XXNN N e.g. BS5 6 or BS55 6</td>
<td>11,000</td>
<td>2,600</td>
</tr>
<tr>
<td>Unit</td>
<td>XXNN NXX e.g. BS5 6ER or BS55 6ER</td>
<td>1,762,000</td>
<td>16</td>
</tr>
</tbody>
</table>

### Table 5.2: Main ONS geographies for England and Wales

<table>
<thead>
<tr>
<th>Level</th>
<th>Old ONS Code</th>
<th>New GSS Code</th>
<th>Number of Divisions</th>
<th>No. of residents</th>
<th>No. of households</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Min.</td>
<td>Min.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Max.</td>
<td>Max.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Min.</td>
<td>Max.</td>
</tr>
<tr>
<td>Regions</td>
<td>A</td>
<td>E12000000X</td>
<td>9 (England only)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Non-Metropolitan (Shire) Counties</td>
<td>11</td>
<td>E10000000X</td>
<td>27 (England only)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>London Boroughs</td>
<td>00AA</td>
<td>E09000000X</td>
<td>32 (England only)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Metropolitan Districts</td>
<td>00BL</td>
<td>E08000000X</td>
<td>36 (England only)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Non-Metropolitan Districts</td>
<td>11UB</td>
<td>E07000000X</td>
<td>201 (England only)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Unitary Authorities</td>
<td>00EB</td>
<td>E06000000X</td>
<td>78</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Wards</td>
<td>36UDGN</td>
<td>E05000000X</td>
<td>8,570(^{14})</td>
<td>100</td>
<td>40</td>
</tr>
<tr>
<td>MLSOAs</td>
<td>E0200XXXXX</td>
<td></td>
<td>7,201</td>
<td>5,000</td>
<td>2,000</td>
</tr>
<tr>
<td>LLSOAs</td>
<td>E010XXXXX</td>
<td></td>
<td>34,753</td>
<td>1,000</td>
<td>400</td>
</tr>
<tr>
<td>Output Areas</td>
<td>E00XXXXXX</td>
<td></td>
<td>181,408</td>
<td>100</td>
<td>40</td>
</tr>
</tbody>
</table>

\(^{13}\) Information taken from:


\(^{14}\) This does not include the 18 electoral wards that are below the minimum population threshold for the publication of census statistics. The estimated mean ward population for England and Wales in mid-2011 was 6,600 people.
Figure 5.1: Illustration of different postcode scales
Figure 5.2: Illustration of the different scales of census/administrative data - (the outline of postcode 'Sector' BS5 6 is shown in the bottom left diagram)
5.3 Potential datasets for integration

There are a number of spatially-based datasets that could potentially be integrated with the MOT data for useful analysis. These include:

- The Census data, accessed via the ESRC;
- The Indices of Deprivation from the Department for Communities and Local Government;
- Various geodemographic coding systems such as MOSAIC UK from Experian;
- The Active People Survey run by Sport England;
- Accessibility Indicators, available from DfT for England, and separately for Wales and Scotland;
- Annual energy use (gas and electricity) data from the Department of Energy and Climate Change (DECC);
- Background air pollution concentrations from Department for Environment, Food and Rural Affairs (Defra); and
- Key pollutants from the National Atmospheric Emissions Inventory.

Each of these are discussed in more detail below.
5.3.1 The Census

The Census is likely to be a core data set for integration with the MOT dataset. By providing information about population and household sizes, it would enable the data about car ownership and use to be put into context. It also provides a wealth of other important variables.

Considerable amounts of data are now available from the 2011 Census, and five data releases have been made so far. Table 5.3 shows the main information available from the 2011 census. Additional mid-year estimates of population are also available at MLSOA and LLSOA levels (e.g. http://www.ons.gov.uk/ons/rel/sape/soa-mid-year-pop-est-engl-wales-exp/mid-2012/index.html).

### Table 5.3: Key data available from 2011 Census


<table>
<thead>
<tr>
<th>People</th>
<th>Households</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numbers</td>
<td>Number</td>
</tr>
<tr>
<td>Demographic and social information about everybody</td>
<td>Housing</td>
</tr>
<tr>
<td>Age (calculated from date of birth)</td>
<td>Accommodation type</td>
</tr>
<tr>
<td>Birthplace (country)</td>
<td>Bath/shower/WC, exclusive use</td>
</tr>
<tr>
<td>Carers, unpaid</td>
<td>Cars &amp; vans, availability and number</td>
</tr>
<tr>
<td>Ethnic group</td>
<td>Central heating</td>
</tr>
<tr>
<td>Health, general</td>
<td>Floor level, lowest</td>
</tr>
<tr>
<td>Illness, limiting long term</td>
<td>Rooms, number</td>
</tr>
<tr>
<td>Marital status</td>
<td>Tenure</td>
</tr>
<tr>
<td>Migrants</td>
<td></td>
</tr>
<tr>
<td>Religion</td>
<td></td>
</tr>
<tr>
<td>School children and students</td>
<td></td>
</tr>
<tr>
<td>Welsh language (Wales only)</td>
<td></td>
</tr>
<tr>
<td>Employment and qualifications of people 16-74</td>
<td></td>
</tr>
<tr>
<td>Academic qualifications</td>
<td></td>
</tr>
<tr>
<td>Professional qualifications</td>
<td></td>
</tr>
<tr>
<td>Working/not working (in week before Census)</td>
<td></td>
</tr>
<tr>
<td>Hours worked</td>
<td></td>
</tr>
<tr>
<td>Means of travel to work</td>
<td></td>
</tr>
<tr>
<td>Tenure</td>
<td></td>
</tr>
<tr>
<td>Social grade</td>
<td></td>
</tr>
</tbody>
</table>

5.3.2 Indices of deprivation

Indices of multiple deprivation combine a number of indicators, chosen to cover a range of economic, social and housing issues, into a single deprivation score for each small
area in England. They have been produced by the Department for Communities and Local Government in 2004, 2007 and 2010. Areas are scored at Lower Super Output Area level. There are also six district summary scores for each local authority district. (There are also supplementary indices measuring deprivation amongst children and amongst older people.) This data set has been the basis of a wide range of work looking at social and environmental justice issues. It is often used as a proxy for a combination of ‘ability to pay’ and ‘ability to act’ and, combined with the MOT data (and other data sets) could enable the investigation of a number of socio-economic issues relating to transport and energy use.

5.3.3 **MOSAIC UK or other geodemographic classification systems**

There are various systems which have been developed for classifying postcodes, or larger geographical areas, into particular types of people (or, in practice, the type of person that is dominant in each area). One of the most widely used systems is Experian’s MOSAIC UK data, though there are other segmentation systems which have also been developed. Such segmentation systems could be useful for interpreting data in relation to a range of attitudinal and lifestyle factors. MOSAIC Public Sector currently contains 146 ‘Person Types’ aggregated into 69 ‘Household Types’ and 15 Groups.

Three Experian datasets (household counts, median household income and population projections) are currently available at an SOA level from census.ac.uk for academic use, but it may be necessary to purchase a commercial dataset, or potentially seek to gain access to segmentation work undertaken by a government department.

5.3.4 **Active People Survey**

The Active People Survey is carried out by Ipsos MORI on behalf of Sport England. It is the largest sport and recreation survey ever undertaken in the UK. The results are representative at the scales of: England; the English regions; and individual local authorities in England. It began in 2005/6, with further waves in 2007/8, 2008/9, 2009/10 2010/11, 2011/12 and 2012/13. It includes measures of walking and cycling. It could be relevant for looking at the relationship between active travel and car use.

5.3.5 **Indices of accessibility**

The DfT publish a set of Core Accessibility Indicators from 2007 onwards for each local authority in England. These are also available at the LLSOA level. The indicators show measures of accessibility by public transport, walking and cycling and car to seven common destinations/services. Accessibility indicators are also available for Scotland and Wales. The indicators have been calculated for the Index of Multiple Deprivation, and show travel time to a range of services.

5.3.6 **Annual energy use data**

This data is compiled by DECC at LLSOA level (see Figure 5.4) and provides an indication of spatial differences in domestic energy use across the UK. By combining this dataset with the MOT data, it would be possible to get a much more comprehensive view of household carbon footprints relating to direct energy use across three axes: electricity, gas and road fuel.
5.3.7 **Background air pollution concentrations**

This data is calculated from modelling undertaken by AEA Energy and Environment on behalf of Defra in support of the UK Local Air Quality Management programme. The data is generated at a 1km resolution and provides an indication of background pollution concentrations, away from the direct influence of local sources (see Figure 5.5). This is generally taken to be representative of concentrations that people would be exposed to in residential properties. Previous work (e.g. Mitchell & Dorling, 2003\(^\text{15}\)) has considered the exposure of people to air pollution in relation to a range of socio-economic factors (mainly at ward level). As a considerable proportion of air pollution is caused by motor vehicles, particularly in urban locations, there could be value in relating exposure estimates at residential locations to indicators of local car use in order to explore social/environmental justice issues, and to feed into the development of air pollution communication materials.

5.3.8 **Atmospheric emissions inventory**

This data is calculated on an annual basis for DECC (greenhouse gases) and Defra (conventional pollutants) by AEA Energy and Environment as part of the National Atmospheric Emissions Inventory (NAEI). The data is available as a 1km resolution grid (see Figure 5.5). Emissions are broken down into a range of sources, including road transport. Consequently, the methodologies and results for the inventory could be explored with new information from the MOT dataset on the annual use and pollution characteristics of the UK vehicle stock.

---

Figure 5.4: Illustration of DECC energy use data at LLSOA for Bristol

Figure 5.5: Illustration of data for nitrogen oxide (NOx) emissions and background concentrations (Defra and NAEI). Black lines are Bristol ward boundaries.

Total NOx Emissions
(µg/m³ Bristol 1km x 1km)

Road NOx Emissions
(µg/m³ Bristol 1km x 1km)

Background NOx Concentrations
(µg/m³ Bristol 1km x 1km)

6 Future research activities

6.1 Overview

During the scoping study, a number of different possibilities were developed for future work. As outlined in the introduction, these were subsequently consolidated into a successful research proposal (EPSRC EP/K000438/1). This new project is entitled “MOT: Motoring and vehicle ownership trends in the UK”. It is a 3 year project, which began in October 2012. It is led by the University of Aberdeen, working with the University of Bristol, University of the West of England, and TRL. DfT and DECC are formal project partners. More details of this project are given below, and via this link:

http://www.abdn.ac.uk/ctr/research/currentbr-research-projects/mot/

6.2 Objectives of the new project

The objectives of the new project, as given in the grant application, are as follows:

"The overall aim of the project is to transform the way in which energy and emissions related to car use are quantified, understood and monitored to help refine future research and policy agendas and inform transport and energy infrastructure planning. Our vision is to achieve this through the following objectives:

1. Combine new sources of data to give a spatially and temporally disaggregated understanding of car ownership and use, and associated energy demand and emissions in Great Britain.

2. Develop new methodologies, datasets and research capability to understand the relationships between energy demand and emissions from car use, and a wide range of structural and social factors.

3. Describe and explain the linkages between different fuel uses, energy end uses and energy service demands at the domestic level.

4. Develop a baseline of spatially disaggregated energy demand from car use from which future scenarios can be developed and modelled.

5. Investigate a range of social and environmental justice issues in relation to income, fuel use and price, emissions of pollutants and exposure to impacts.

6. Use this understanding to track changes over time and space in order to evaluate the scale and distribution of the impacts of local transport policy interventions."

6.3 Structure of the new project

The new project is structured into a series of workpackages, as shown in Figure 6.1. More details of each of these workpackages are as follows.

WP 1.1 aims to enhance the capabilities of the existing MOT data set, by adding new fields, and aiming to address some of the issues discussed in Chapters 2 and 3.

WP 1.2 aims to explore the opportunities for analysing the MOT data in conjunction with other spatial datasets (as described in Chapter 5), including acquiring relevant data and developing GIS frameworks and methodologies to link data across different spatial units and to assess year on year variations.
WP 1.3 will provide data to all other workpackages, and also involve developing new mathematical modelling techniques to achieve the spatial and temporal representation of the MOT data at different scales. It also involves exploring different ways of fusing this information with other data sets.

WP 2.1 aims to understand the degree and nature of both temporal and geographic variation in car ownership and use, and how insights from the MOT data compare with those from other relevant datasets, including the National Transport Model. It will involve applying statistical and modelling techniques in order to explain the relationship between the vehicle data and ‘macro’ factors such as population density, type of location (e.g. urban-rural) and urban structure, average incomes, population subgroups and local levels of transport infrastructure. It will also explore the way in which the data could be used to automatically produce benchmarks or control trend data for particular types of areas. It will also include a review of international experience of using MOT data to explore issues relating to energy use, transport demand and climate change.

WP 2.2 aims to analyse patterns of CO₂ emissions and energy use from vehicle use, and to understand how these relate to other data on household energy use and the
contribution of vehicle use to air pollution; in the context of debates about social and environmental justice, and the need to understand the relative role of different types of households and locations to emissions, energy use and air pollution.

WP 3.1 aims to understand what insights the MOT dataset can provide about the effects of particular local travel behaviour change policy interventions on car ownership and use, such as the Sustainable Travel Towns; the Local Sustainable Transport Fund (LSTF) etc.

WP 3.2 aims to assess the spatial and social implications of policies designed to reduce energy and emissions from cars, such as (carbon delineated) road user charging, Vehicle Excise Duty, local congestion charges, low emissions zones, car purchase taxes and car scrappage schemes.

WP 4.1 is concerned with ensuring written and electronic outputs from the project which are directed towards policymakers and users, rather than focused only on the academic community, including development of policy briefings and a project website.

WP 4.2 will involve the organisation of seven events for engagement with key stakeholders and dissemination of outputs, in order to allow a two-way discussion with relevant communities, together with more informal meetings.

WP 4.3 aims to explore the potential for creating a tool that might be useful to a wide range of users who might wish to find out about car ownership and use, and associated energy use and emissions for particular areas and/or specific time periods.
Using MOT test data to analyse travel behaviour change - scoping report

This report describes the results of a 3 month scoping study, completed between April and June 2011, to explore the potential of the data collected during the British MOT test. Some updating of the findings reported here took place in December 2013. The MOT test data were first made publicly available in November 2010, with the release of over 150 million test results, with further data releases since that time. The data set is unique in providing national information about vehicle types and use at a scale which is not available from any other source. This report examines the data that have already been made available, and the various ways in which the dataset could be enhanced. It outlines a set of mathematical techniques which have been developed for initial data analysis and describes some of the results that emerge (notably, for example, the important relationship between vehicle age and vehicle use). It also reports on some of the other data sets that could be analysed in conjunction with the MOT test data, in order to explore issues relating to transport demand, energy use and climate change. Following on from this scoping study, a new research project has been developed, and the last chapter of this report describes the activities planned in that project, which will take forward the work described here.

Other reports:

