Department for Transport

Experimental Statistics: Road Condition SCANNER data report - Methodology Note

Introduction

This document supports the experimental statistics on the condition of roads in England: https://www.gov.uk/government/statistics/road-conditions-in-england-to-march-2020-and-march-2021

It is part of Road Network Size and Condition series.

In this document

Further information

Public enquiries

020 7944 3095

roadmaintenance.stats@dft.gov.uk

Media enquiries

020 7944 3066

This note provides the methodology and definitions used for the experimental road condition statistics using SCANNER survey data. It also includes useful information on the source of the data.

Accessing data

Department for Transport with the assistance of Transport Research Laboratory (TRL) were able to access the SCANNER data, which underlies the official statistics reported by DfT for a number of local authorities (LAs). The data that TRL provided is processed output data they use to the SCANNER data as part of their role as auditor under the existing data requirements.

DfT collected consent from LAs to access this data through the last two annual data collections, which covered data for 2017/18, 2018/19, 2019/20 and 2020/21.

Data source

SCANNER surveys: these are automated surveys carried out using SCANNER vehicles. LAs commission them to assess the surface condition of their 'A', 'B' and 'C' road network and in some cases on their unclassified ('U') road network as well.

Since 2007/08, the SCANNER data which LAs have shared with DfT has only included the length of road surveyed and the proportion of roads classified as red (roads which should be considered for maintenance). In recent years this has expanded to roads classified as green and amber also. The new data which DfT has accessed for this report is the data which underlies this information, and provides details of the condition of each road section surveyed.

The surveys measure a number of parameters, with results reported for every 10-metres of road. The outputs from a core set of these parameters are combined to produce a single figure to indicate surface condition at a specific location, the Road Condition Index (RCI).

Sections 3.4-3.5 in the Technical Note detail the parameters measured and how they are used, and section 3.6 for how the parameters are used to calculate an RCI value.

The 'core' parameters that make up the SCANNER RCI are Maximum rut depth (left / right), Moving average longitudinal profile variance (3m / 10m), Texture depth and Whole carriageway cracking. There are a large number of other parameters within the data. Some more information on how the scores for the parameters have been derived can be found in the Technical Note. This analysis focuses on the RCI as an overall measure of condition rather than looking at individual parameters.

Within the dataset each row of data reflects a 10m subsection of road and will have an RCI value, in cases where the results are valid. Instances where an RCI value would not be produced would be where one or more of the core parameters required for the UK pavement management system (UKPMS[1]) national weighting set are missing and produce a non-zero result, these subsections are excluded from the analysis. The main variables included within the output dataset are as follows:

- **TOID** – Topographic Identifier, a road section reference.
- **Easting/Northings** – coordinates for the start point of a 10m subsection.
- **Chainage** – cumulative length of the 10m subsections according to sections of the network surveyed.
- **Lane** – This reflects what lane has been surveyed based on direction (right or left). SCANNER surveys are primarily undertaken on lane 1 (outer lane or slow lane) of the carriageway.
- **Class** – road classification ('A', 'B' & 'C' roads).
- **UR** – the urban/rural flag for the road section.
- **Survey date** – reflects the financial year that the survey was undertaken.
- **RCI** – Road Condition Indicator, this is the main measure and is calculated using specific core parameters, which combine to give an overall measure of surface condition.

Data coverage

Classified 'A', 'B' and 'C' roads in England were within scope for this project, as existing data requirements require SCANNER data to be collected for these roads. Data for unclassified roads was out of scope, because there is no mandatory requirement for LAs to collect data for these roads using SCANNER, and as a result there is low coverage of these roads.

The data was collected for four financial years; 2017/18, 2018/19. 2019/20 and 2020/21. Surveys are carried out on a two year cycle so 2017/18 and 2018/19 was combined, as was 2019/20 and 2020/21, which created two distinct time periods.

1 https://ukrlg.ciht.org.uk/ukrlg-home/guidance/road-condition-information/data-management/uk-pavement-management-system-ukpms/

Data from 123 out of 151 Local Authorities (LAs) for 2017-19, and 85 LAs in 2019-21 have been included in this analysis. Data was included only where permission was obtained from the LA, the data was held by TRL, the data was complete and passed validation checks. Due to changes in how survey's are carried out in London from 2019/20, all London LAs are excluded in 2019-21.

- **The methodologies used in this report will differ from those used in the official statistics.** Processing rules used by LA UKPMS systems to supply information for the official statistics will vary by LA.

Caution should be taken when comparing the figures over time, as there are a different subset of LAs in each time point, which may impact the results especially where more granular breakdowns are shown.

Producing the Experimental Statistics

The statistics produced from the underlying SCANNER data are experimental. They differ to the main statistics published which primarily focus on roads categorised as red and are based on LA's UKPMS national outputs.

The experimental statistics have been produced from the underlying SCANNER data accessed from TRL. It was deemed that the most appropriate way to present the statistics would be to combine two years together as a single time point within the analysis, due to the way the SCANNER data is collected over a two year cycle. In order to combine the two years of data the approach taken has followed rules as close as possible to that of LA UKPMS systems.

Where a road was surveyed twice in a two year period, the most recent data was taken as the RCI score for this road section. Road sections were matched by section label, class, LA, lane and the easting/northing coordinates to allow identification of multiple records for each road section.

The time periods have been combined to give two distinct time periods with no overlap, to better show any changes to network condition over time. This means that the data has been presented for 2017/18 and 2018/19 together for the first time point, and 2019/20 and 2020/21 has been combined for the second time point.

The statistics presented are based on the RCI. The RCI score for each 10m subsection is used to classify the data into the broader categories of condition (i.e. red/amber/green roads) but also into smaller discrete sub-categories to show the distribution across the range. The RCI values are also used to create the various descriptive statistics presented in the statistical release. The analysis uses all recorded RCI values (i.e. 0 and above) with any blank values treated as invalid scores.

Road Condition Indicator (RCI)

The main measure of condition shown in this report is the Road Condition Indicator (RCI). This is made up of several parameters, such as cracking and rutting (see Technical Note), which combine to give an overall measure of the state of the road and an indication of surface condition.

RAG condition categories

Each 10 metre road section has been assigned a condition category based on the RCI value. The thresholds used to determine the categories of condition are as follows:

- 0 to less than 40 = green (good condition - no further investigation or work is likely needed at this time).
- 40 to less than 100 = amber (likely to be some deterioration – work may be needed sometime in the future).
- 100 or more = red (likely to be in poor condition – further investigation may be required to determine whether this section of road should be considered for maintenance. The LA will decide whether any intervention is made).

Breakdown by LA size and urban/rural

Along with national level analysis, data was analysed by LA size, whether an LA was urban or rural and region. To assess the impact of LA size, the size of each LA managed highway network was ranked, based on the length of their road network, from smallest to largest and grouped into four categories (quartiles):

- First quartile: Smallest - within the smallest 25% of LA managed highways.
- Second quartile: Smaller than average - within the smallest 25% to 50% of LA managed highways.
- Third quartile: Larger than average - within the largest 50% to 75% of LA managed highways.
- Fourth quartile: Largest - within the largest 25% of LA managed highways.

LAs have been categorised as urban, rural or mixed based on the proportion of their road network that is urban or rural. If more than 66% of the entire LA managed road network length is urban or rural, then the network is defined as the corresponding category (i.e. 66% urban = urban). LAs that do not reach this 66% threshold for either urban or rural are classed as mixed.

Interactive Map

► Some local authorities were excluded from the analysis but their information has been displayed on the map. This included cases where the experimental statistics were not comparable to the previously published official statistics (based on local authority UKPMS returns); where data was incomplete as a partial survey had been undertaken over the two year period or where data was not provided at the time the analysis was undertaken.

► Not all road sections will be shown on the map as local authorities are only required to survey 90% of their ‘A’ roads in both direction, 85% of their ‘B’ roads in both directions and 80% of their ‘C’ roads in a single direction over a two year period, for national reporting.

► Where the road has been surveyed but an RCI score is unavailable these road sections have been displayed on the map in grey to show the full extent of the network surveyed.

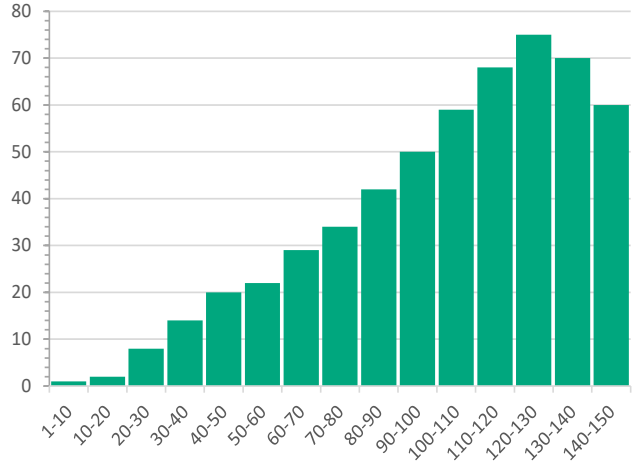
Statistical Measures

A set of statistical measures have been produced to help interpret the SCANNER data and present a more complete picture of road condition in England.

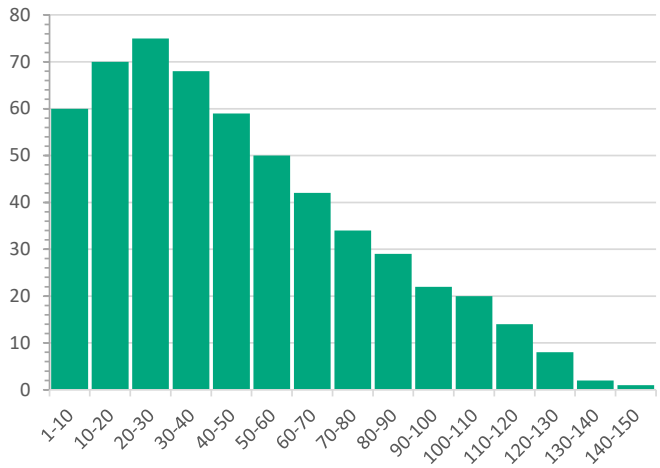
Measure of Skew

The skew was presented to aid interpretation of the distribution of the RCI scores. Skew was measured using the Fisher-Pearson coefficient of skewness.

Skewness is a measure of the symmetry of a distribution around the mean, looking at the deviation from the normal distribution. It tells us if there is clustering of scores at one end of the distribution and a distribution is skewed if there is asymmetry causing a tail on one side of the mean.



Left-Skewed (Negative Skewness)



Right-Skewed (Positive Skewness)

A negative skew indicates that the tail on the left side is longer than on the right side (left-skewed), and a positive skew indicates the tail on the right side is longer than on the left (right-skewed).

- ► If the skewness is between -0.5 and 0.5, the distributions fairly symmetrical i.e. values follow a normal distribution.
- ► If the skewness is between -1 and -0.5 or between 0.5 and 1, the distribution is moderately skewed.
- ► If the skewness is less than -1 or greater than 1, the distribution is highly skewed.

Standard Deviation

The standard deviation measures the dispersion or variation of the values in the data relative to the mean. In the report, the standard deviation was used to look at how spread out the RCI values were around the mean. If the RCI values are further from the mean, there is a higher deviation within the set of data and this is represented by a higher standard deviation score. Conversely a lower standard deviation score will reflect that the RCI values are closer to the mean. This measure helps show whether the number of extreme values is changing over time.

Mean (Average)

SCANNER data is aggregated into 10m sub-sections of road. The average RCI is the mean score of a 10m sub-section of road within the category being assessed. The RCI values of the 10m sub-sections of road are summed and divided by the number of road sections surveyed to give the mean value. This measure provides a more detailed look at the average quality of the road network, than the traditional red, amber and green categories, as it allows assessment of condition along the full distribution of potential RCI scores.

Specific Issues

Covid-19

The data presented for 2019-21 will partly have been collected during the coronavirus (COVID-19) pandemic, although road maintenance did continue this may have been affected by changes to local authority priorities during covid. There was also reduced road traffic during this period,as shown in road traffic estimates.

London data

Road condition surveys for London boroughs have been coordinated by the London Borough of Hammersmith and Fulham under a programme known as Roads 2000. This was fully funded by TfL however was withdrawn in 2019/21. This means that London boroughs would need to plan their own surveys. In 2019/21 no SCANNER data for London boroughs was provided.

Details of all issues and limitations to SCANNER surveys can be found in the Technical Note.

Limitations

There some key limitations to consider when drawing conclusions from the SCANNER data analysis:

- ► The methodologies used in this report will differ from those used in the official statistics. Processing rules used by LA UKPMS systems to supply information for the official statistics will vary by LA. Therefore, figures presented within the SCANNER report may not perfectly match those presented in the official statistics.

- ► This report only includes SCANNER road condition survey data, there are two key limitations to this;

    - ▸ Not all LAs use SCANNER surveys and more are moving to alternative technologies over time. This means that not all LAs can be included in this analysis which may skew the results, but checks have been made to ensure we have a representative sample.

    - ▸ The data presented in the experimental analysis is not directly comparable to the official statistics. The official statistics include figures reported from alternative technologies to SCANNER, whereas the experimental statistics are solely based on SCANNER data.

► The purpose of the RCI is not to give a definitive assessment of whether a road requires maintenance. It is a tool to provide LAs with an indication of surface condition to help identify roads which may require planned maintenance. The LA will ultimately decide whether any intervention is required after taking into consideration a range of evidence and local factors.

► There are limitations to the RCI, as it is an overall measure of condition based on a few core parameters, and in certain cases the RCI value can be heavily influenced by a specific defect. This can result in sections of road that appear better or worse to the road user relative to the RCI value and RAG category shown on the interactive map. For example, roads categorised as green may be viewed as poor from the perspective of a road user as this defect alone does not influence the overall RCI score. Alternatively, roads that are categorised as amber or red may appear in better condition than some roads categorised as green.

► The output data used was produced from the raw HMDIF files TRL received from the LAs, however TRL match the data to a 2009 version of the Integrated Transport Network model (ITN) for consistency between years for their own auditing purposes. This means that there could be some road classifications and urban/rural definitions within this data which are out of date, although roads tend not to change classification frequently. The road classification determines how the RCI is calculated so an incorrect road classification will effect the RCI. For more information on how the network model is fitted to the underlying SCANNER data please see the published guidance documentation on the CIHT website.

►This older version of the network also affects the urban/rural classification of the roads. The data provided by TRL contains urban/rural classifications for each record based on the 2009 ITN. These classifications are more likely to change over time than road classifications and the urban and rural definitions have also been redefined since 2009. Due to these limitations the definition of urban/rural was defined by the whole network size of an LA. This allows general comparisons of LAs that are mostly urban or rural but does not allow link level comparisons.

► Comparisons of road condition across different road types should be made with caution due to differences in the RCI calculation. For rutting and cracking there is no difference in the weightings for each road class however for longitudinal profile variance (LPV) and texture depth there are differences:

- ▸ For LPV there is a difference in the thresholds. ‘A’ roads are weighted against higher standards than ‘B’ roads, which are in turn weighted to a higher standard compared than ‘C’ roads. So for a given measurement for LPV, the score will generally be greater if that road is an ‘A’ road compared to if it is a ‘B’ road.

- ▸ For texture depth the hierarchy is more complex as the thresholds and the importance factors vary, it is also based on road classification and urban and rural classification.

► The data included in the analysis is based upon whether it was deemed useable and follows UKPMS processing where possible, however the proportion of roads categorised as red may differ to the results in the main statistics submitted by the LAs due to the differences highlighted above with the output data provided by TRL.

► Although each vehicle is accredited for accuracy and readings fall within the accepted boundaries of the SCANNER specification for road condition, there is still variability between the results that each machine delivers. It can lead to small changes in the figures over time that are for reasons beyond the condition of the road, and above the expected range of variability that already exists within the data.

► The figures for the two time periods shown use a different subset of LAs as this is based on those LAs that provided their consent and were accessible, which may impact the overall results. Not all LAs in England provided their consent for DfT to access their underlying SCANNER data and some do not use SCANNER to collect their road condition data. It means the regional figures are based on the subset of LAs included in the analysis that make up that region.

► The Easting and Northing co-ordinates within the data have been used to plot the data points on the map. The level of accuracy of the co-ordinates varies, but they should be accurate within a few meters.

Glossary of technical terms

A glossary of technical terms and additional information about road condition reporting can be found in the Technical Note.

To hear more about DfT statistics publications as they are released please follow us on Twitter via our @DfTstats account: http://www.twitter.com/DfTstats. TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates