# Understanding historical road safety data

Data for road safety casualty statistics are collected by each police force independently, this data is commonly known as STATS19, as that was the original name of the form used for this purpose.

STATS19 data is collected to an agreed specification as described by STATS20 guidance and exchanged using the STATS21 formatting and rules. The specification of STATS19 is periodically reviewed to ensure information collected by the police is relevant to emerging road safety needs and minimises the burden on the police. Reviews are conducted approximately every 5 years. The recommendations from reviews change the specification by adding or removing fields, adding or removing categories in fields or changing the data. The latest STATS19 review was conducted in 2018: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/995117/stats19-review.pdfv>

New specifications are generally named after the year they are introduced in (e.g. 2005 specification, 2011 specification, etc). The 2002 review led to the creation of the 2005 specification, which was used by some forces between 2005 and 2018. The 2008 review led to the creation of the 2011 specification, which was implemented in 2011 and is used by some forces currently. The 2018 review will inform the creation of the next specification which is planned to be implemented in 2024.

To produce longer time series analysis, users may be required to group and manipulate data collected on different specifications e.g. the current vehicle\_type category for cars is 9 which was introduced in 2005 specification, however prior to that code 109 was used.

In an ideal world, all forces would adopt a new specification at the same time. However, real world challenges make this far more complicated a process, as changes require digital infrastructure changes, technical training, and funding. In 2018, police forces were independently reporting to the Department in 4 different specifications or variants. Any 2005 specification or pre 2005 specification data still received is superseded by Local Processing Authorities (LPAs) updates in a more modern specification.

Therefore, forces may continue to report using older specifications until they migrate, which will influence what values are in range for specific fields.

Whilst efforts have been made to homogenise the data from differing format specifications, users should consider the effect of differing format specifications when using this data.

In some variables, the transition is easy to identify (e.g. the car example above), but some are more difficult to identify. For example, in the 2005 specification coordinates were recorded to 5 figures, giving a recordable accuracy to 10m. If visualised, these collision locations would not fit well to any road networks, and would form along grid lines rather than have a strong relationship to the road it was related to.The 2011 specification improved this by increasing the coordinates to 6 figures, giving a recordable accuracy to 1m and adding an additional digit in the northings to allow for grid references in the extreme north of Scotland to be collected.

The 2005 specification (recorded the coordinates to 5 figures) has a trailing zero added to both the easting and northings to make it conform with other 6 figure coordinates.

It was only in 2018 that the last 4 data providers using 2005 specification migrated to the 2011 specification and true 6 figure coordinates could be exchanged.

# Data Quality

Due to the constant improvement of the data pipeline and validation, users may find that modern accidents contain more accurate or consistent information. Significant efforts and funding has gone into helping officers collect data.

Historically, all police officers would have been collecting data via handwritten paper forms, which were not held to the same level of scrutiny that digital records are examined to today. Some police officers also have access to mobile phone applications and better data systems, which validate the data entry as officers are inputting it. For example, officers can use a mobile phone’s GPS to gather accurate coordinate information rather than a written text description of the accident based on nearby points of interest.

DfT and Local Processing Authorities (LPAs) have also improved their validation techniques over time, allowing them to identify issues faster so issues are rapidly turned around to the reporting officer for review. For more recent changes to specifications and data systems, please see our annual report: <https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-annual-report-2020/reported-road-casualties-great-britain-annual-report-2020#strengths-and-weaknesses>

Users should also consider whether sharp changes in trends are a result of genuine observed fluctuations or whether they are a result of different forces changing specifications from year to year. If a field or category is added to a newer specification, then sharp increases can be observed as more forces move to the newer specification. This would result in more observations being recorded, as opposed to occurring.

# Introduction of new non-STATS19 fields

DfT also derive several fields e.g. Index of Multiple Deprivation (IMD) by joining fields in STATS19 data to external databases. This creates additional information without increasing the burden on the officers who collect the data.

Where users have expressed a need for specific information, DfT have introduced new fields. These newly introduced fields have not been backdated as of yet due to limitations in the joining of the data, but will be retained in future datasets e.g. IMD of postcodes will change over time, it is not accurate to apply the IMD scores of 2018 to casualties in road accidents from 1979.

# ONS codes

DfT has also started reporting data to new statistical standards e.g. the use of ONS codes for geographic breakdowns. These were introduced to all forces at the lower geographic level in the road safety database in 1997. Previously, DfT had used its own categories, and allowed for a transition period where both fields would be reported. However, DfT will no longer be maintaining the older 3 digit codes in the event of new geographic boundaries, and will default to the ONS version of the fields. The local\_authority\_district field will still be made available to the public.

DfT recommends using local\_authority\_district for road accidents prior to 1997, and the ONS version for road accidents from 1997 onwards. To carry out longer time series analysis, users may be required to manipulate both fields. Users should also account for changes in boundaries e.g. Cheshire splitting to Cheshire East and Cheshire West in 2009.

# Further Questions

Guidance for police officers on the collection of STATS19 data is documented in STATS20, this includes definitions and example scenarios for specific categories: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/995424/stats20-2005.pdf>

For any further questions or clarifications, please contact [roadacc.stats@dft.gov.uk](mailto:roadacc.stats@dft.gov.uk).