TRIP

TARGETING ROAD INJURY PREVENTION

Report Prepared by Dr Jo Barnes
Contributors James Nunn Dr Emily Petherick Professor Andrew Morris Matt Staton

August 2021

Executive summary

Prior to 2012 the UK saw a sustained reduction in road casualties where deaths from road collisions nearly halved. However, since then there has been a general plateauing of road deaths per year and incidents of serious injuries have also followed this trend. In 2018, 6.5% of all national killed and seriously injured casualties were fatalities, but in Cambridgeshire it was 7.6%. Additionally, the rate in 2018 for fatalities per 100,000 population in Great Britain was 2.8 for Cambridgeshire this rate was 5.9. The Cambridgeshire and Peterborough Road Safety Partnership sought to explore new ways that road safety interventions can be delivered to reduce serious and fatal injuries resulting from collisions. The notion being to target specific drivers who are responsible for the collisions based on geodemographic profiles. This study is a proof-of-concept study exploring the available data and methods involved to enable routine use of geodemographic profiling in road safety interventions.

Aim

The aim of the study was to inform innovative approaches to road safety interventions to reduce the numbers of killed or seriously injured in road collisions.

Methods

A series of methods were used to identify Cambridgeshire drivers who were culpable of causing the collisions and if they were different to non-culpable drivers.

- The police collision database (STATS19) was linked with hospital trauma audit research network (TARN) data for a five-year period to identify Cambridgeshire resident drivers who were involved in clinically defined serious injury collisions.
- All drivers were culpability scored and categorised as being culpable, contributory, or non-culpable for the collisions. To achieve this the STATS19 variables were mapped on to an existing tool.
- Full geodemographic profiles were appended to the drivers with a culpability score.
- Analysis of the data investigated the culpability and geodemographic profiles of the drivers and explored differences in Cambridgeshire drivers to inform road safety interventions.

Results

The study identified 564 drivers involved in a serious injury or fatal collision on the Cambridgeshire road network and had a culpability score. The mean age of drivers was 43 years (SD17) and most were male (434, 77%). For these drivers, the significant factors impacting on the odds of being culpable were their age, being under 26 or over 76 showed higher odds compared to the mid-aged

(46-55years) as did being the rider of a motorcycle compared to cars. Being the driver of an agricultural vehicle or goods vehicle showed lower odds of being culpable compared to cars as did residing at an address with an Acorn Type designation of 6 ('financially comfortable families') compared to the most frequent Type 23 ('owner occupiers in small villages'). When considering only Cambridgeshire resident motor vehicle drivers (367 (65%)) the significant factors impacting on the likelihood of being culpable, were being the rider of a motorcycle compared to a car or an Index of Multiple Deprivation (IMD) in the 6th decile compared to the most frequent IMD 5th decile. Similarly, to all drivers those Cambridgeshire resident drivers living at an address with an Acorn type designation of 6 had lower odds compared to Type 23 of being culpable. In general non-resident drivers were involved in more fatal collisions (49%) compared to Cambridgeshire residents (42%). The use of risk indexation was explored for the geodemographic Types to identify if there were any Types overrepresented in the study sample compared to the population of Cambridgeshire. Overrepresentation on the risk index determines the extent to which a Type is found culpable compared to the general population of Cambridge. Type 41 culpable drivers were high frequency for fatal collisions and overrepresented compared to the general Cambridgeshire population (risk index >300). Type 41 describes 'Labouring semi-rural estates.' For serious collisions Type 23 were high frequency and overrepresented and had a risk index >200. Interestingly there were some Types underrepresented on the risk index specifically Type 10 (Better-off villagers) for fatal collisions and Type 5 (Wealthy countryside commuters) for both fatal and serious (MAIS3+[1]) collisions, suggesting lower risk of culpability. This would be interesting to explore further with larger datasets to understand how typical the over or underrepresentation of culpable drivers is.

Conclusion

Overall, the methods have allowed for culpable drivers causing clinically defined serious injuries to be identified residing in Cambridgeshire. STATS19 has been mapped to a culpability tool for the first time and is being validated for use on a large dataset. The results indicate the potential if using this methodology to identify drivers causing collisions and to use this knowledge to target specific road safety interventions. However, the sample was small and any inferences in the data need to be made with caution as the focus has been on serious and fatal injury collisions and not those with minor or damage only outcomes and are limited to Cambridgeshire.

Road Safety implications

This method would enhance road safety professionals' opportunities to develop targeted innovative road safety interventions at the culpable drivers. However, the automaticity of determining culpability from STATS19 variables is required before the method can become user friendly in the

[1] MAIS3+ refers to a clinically defined serious injury for example fractured femur.

real world. It would also be beneficial to explore the nuances of the geodemographic Types identified in the study with residents from the profile Types. This would identify whether the profile descriptions have any similarities with residents and determine the best method of delivery of road safety interventions. This would enhance their effectiveness at reaching the target audience and subsequent reduction in serious injury and fatal collisions.

This study was undertaken as post graduate research for the award of Doctor of Philosophy (PhD).

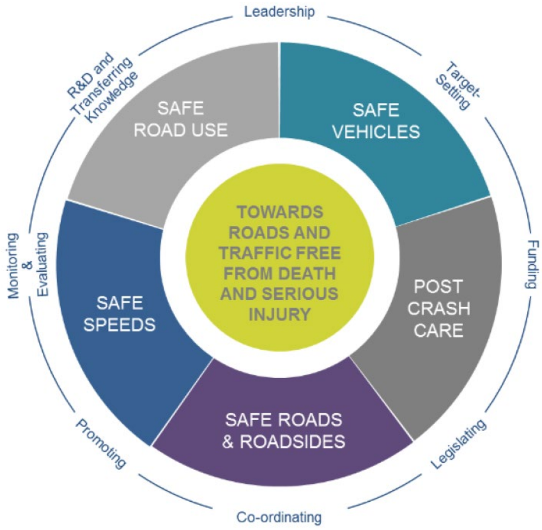
Contents

Tables and Figures

Introduction

Road traffic collisions are multi- faceted-complex events (Wagenaar and Reason, 1990, West, 1997), although most collisions are recorded as driver error (DfT 2019b). In 2015 the UK adopted the safe system approach to road safety which recognises humans are fallible and will make mistakes and cause injury. This human fallibility requires the overall system around the individual to be lenient and forgiving whereby the road, legislation and vehicle provide support to prevent collisions and therefore injury (Figure 1) (Parliamentary Advisory Council for Transport Safety, 2016). Where the system fails any severity of injuries sustained should as low as possible and post injury care as part of the overall system should be efficient and fit for purpose. Road safety is a shared responsibility between numerous stakeholders including road planners, vehicle manufacturers, emergency care providers and road users to take appropriate actions to ensure that road collisions do not lead to serious or fatal injuries (ITF/OECD, 2008).

However, the extremes in driver behaviour, for example excessive speeding will not be absorbed by this approach hence safe road use is also integral to the system working.



(The Royal Society for the Prevention of Accidents, 2018, p. 4)

Figure 1: The five essential elements of the Safe System approach

Prior to 2012 the UK saw a sustained reduction in road casualties, where deaths resulting from road collisions nearly halved from 3,221 in 2004 to 1,754 in 2012. However, despite committing to a safe system approach this progress has plateaued between 1,713 and 1,793 road deaths per year in the subsequent six years to 2018 (Department for Transport, 2019a). The incidence of serious injuries has followed a similar rising trend in recent years and incur high societal and human costs as well as being associated with long term psychological and physical outcomes (Craig et al., 2016; Guest et al., 2016).

Previously injury reduction targets used the police collision dataset (STATS19) to define the term serious injury which was problematic relying on an 'at scene' subjective assessment by the reporting police officer. However, since 2015 the clinical injury severity of MAIS3+ using the Abbreviated injury severity scale (AAAM 2008) has been used to define serious injury across Europe (IRTAD 2011). One of the problems associated with the shift to using this definition in the UK is that there is no single database providing enough information about the collision and clinical injuries without having to manipulate data to approximate the MAIS3+ severity definition. Thus, there is a need to use health sector data for meaningful injury classification to complement police data and to provide an optimal means of defining serious injury (Broughton et al., 2008; IRTAD, 2011). The benefit of having a European definition of serious injury enables comparability in setting and measuring road safety targets. To prevent death and mitigate serious injury on the roads and achieve injury reduction targets has relied on many national road safety education campaigns for example, 'Think bike' or lowering speed limits as well as improving the national road network but there has been little targeting of road safety measures at the drivers culpable of causing serious or fatal collisions. These measures which prevent death and prevent and mitigate serious injury may be quite different from measures to prevent crashes in general (European Commission 2018).

In 2018, 6.5% of the national killed and seriously injured (KSI) casualties were fatalities, but in Cambridgeshire it was 7.6% (50 fatalities and 660 serious injury. Additionally, the rate in 2018 for fatalities per 100,000 population in Great Britain was 2.8 for Cambridgeshire this rate was 5.9 (50 fatalities and a population of 847k) (Cambridgeshire Insight, 2019). This almost double rate of fatal and serious injuries is a concern for Cambridgeshire and a different approach to road casualty reduction warrants further exploration. The notion of trying to target specific road safety interventions at drivers who cause serious and fatal collisions is explored in this study to tackle the rising casualty figures in Cambridgeshire. This study links 5 years of police collision data (STATS19) and hospital trauma data (TARN) to identify culpable drivers involved in serious (MAIS3+) collisions and fatal collisions. Culpability of drivers will be explored to determine whether there are different characteristics or geodemographic profiles to non-culpable drivers that could be used to target innovative road safety measures.

Literature review

Collisions and culpability

The analysis of STATS19 data has been used widely to explore numerous research questions including collision frequency, collision severity, the predicting of different severities at different sites (Wang, Quddus and Ison, 2011), child injuries (Jarvis et al., 2000), collisions crash-speed relationships (Imprialou et al., 2016), the relationship between deprivation and collision risk (Graham, Glaister and Anderson, 2005; Edwards et al., 2006), exploring if graduated driving licence could reduce casualties (Jones, Begg and Palmer, 2013) and geographic distribution of road casualty injuries (Steinbach, Edwards and Grundy, 2013) amongst many others. However, STATS19 is not without fault as like any dataset it requires careful recording of the many factors to understand and analyse how these factors interact (Cercarelli et al., 1992) and has been found to have issues with data quality (Imprialou and Quddus, 2017), and under-reporting of collisions (Roberts et al., 2008; Broughton et al., 2010; International Transport Forum, 2011; 2018; Yannis et al., 2014; World Health Organization, 2018). Generally when exploring the drivers involved in collisions there are a number of factors that can be recorded and typically the majority of drivers are recorded with a contributory factor of driver error or reaction in some 67% of collisions in 2018 (DfT 2019b) and most frequently cited is failure to look properly. Driver behaviour encompasses a multitude of human factors, such as, attitudes to drug or alcohol consumption and risk taking (Bernhoft, 2011), through the impact of intoxicants (Mathijssen and Houwing, 2005) to the impact of emotional state or fatigue (Fell, 1976) which need to be taken into consideration when considering the causes of collisions.

Although STATS19 records contributory factors that relate to the driver, culpability is not recorded which would be beneficial if road safety interventions need to reach this target audience. Culpability refers to any driver actions undertaken or not prior to the collision that in the opinion of the reporting police officer caused the collision to occur in full or contributed in some way. Therefore, to be non-culpable the driver could not have avoided the collision whatever driving action they took (af Wåhlberg, 2002). Culpability is not straightforward to determine, where subjectivity may influence individual observation and hence judgements of individual culpability (Köhnken and Brockmann, 1987).

There are three tools in the literature that have been devised and used to determine culpability in different study populations, Terhune (1983), Robertson and Drummer (1994), Brubacher, Chan and Asbridge (2012). The tool most often used in the literature and deemed the least subjective of the options (Brault and Dussault, 2002; af Wåhlberg, 2009) was the tool devised by Robertson and Drummer (1994). However, it is not without its challenges as the culpability tool must be capable of differentiating individual collision circumstances rather than just identifying active or passive involvement (West 1997, West and Hall 1997). To date these tools have not been used to assign culpability on STATS19 collision data.

Injury severity

STATS19 records collision injury severity as either minor, serious, or fatal but under reporting of serious injury is an often-cited issue (Cryer et al., 2001). Thus, the move to having a clinically defined level of serious injury (MAIS3+) ensures measurable and comparable road safety injury targets can be set in the UK and Europe. The MAIS 3+ severity is derived from the Abbreviated Injury Scale (AIS 2005-2008 update) (AAAM 2008) used widely in research and routinely in some countries to collect trauma data on patients admitted to hospital. Injuries need to be coded by trained coders using the

AIS dictionary which provides a six-point ordinal rating of injury severity to individual injuries sustained by the patient (Table 1).

Table 1: Abbreviated Injury Scale description of injury severity

AIS severity description	Severity ordinal scale	Example injuries
Minor	1	Bruises, fractured finger
Moderate	2	Fractured wrist
Serious	3	Fractured femur
Severe	4	Fractured ribs with lung contusions
Critical	5	Large contusions to the brain
Untreatable	6	Catastrophic spinal cord injury at the neck

The serious injury definition (MAIS3+) describes any collision in which a person sustained at least one injury coded as serious, severe, critical, or received untreatable injuries. This is defined as MAIS3+, indicating there is a maximal abbreviated injury severity of 3 or above. One of the problems in road safety research is that AIS is reliant on clinical coding and is not routinely available in the STATS19 police collision data.
The Trauma Audit Research Network (TARN) in the UK is a database that clinical codes and records traumatic injuries using AIS in the UK and some European Countries (TARN). As with using other hospital records (Pérez et al., 2016) the data held on the TARN dataset will not capture all MAIS3+ injuries as some patients may not be hospitalised for sufficient time to meet the TARN entry criteria.

Data linkage is used to incorporate collision and injury data but to varying success. The Department for Transport undertook record linkage with Hospital Episode Statistics (HES) data to explore the injuries however the linkage rates were poor (DfT 2012). Additionally, HES clinical data is coded using the International Classification of Diseases (ICD-11) (WHO 2016). ICD is a generic coding tool that codes and captures all hospital admission events and is not specific to injury data that enables a severity of injury to be determined at the MAIS3+ level. There have been computer algorithms applied to the ICD data (ECIP 2006, Clark et al 2009) and more recently an expert derived map (AIS08_ICD) was commissioned by the EU to enable the identification of MAIS 3+ injuries from ICD codes (Loftis et al 2016, Zonfrillo et al., 2015). However, there are some issues when using the tool based on the selection of the original number of ICD codes used to generate a MAIS3+ score (Barnes et al 2020). Other studies have linked several datasets that required special permission and ethical consideration but broadly what is evident from data linkage is the disparity between STATS19 injury severity descriptors and that of clinically derived assessment (Morris et al.,2006, University of Leicester 2005).

Targeting collision prevention

There is a plethora of literature on road safety interventions of which Elvik et al (2009) provide a comprehensive description. There have been certain strategies for example the ‘three E’s’ which frames intervention at education, enforcement, or engineering (Groeger 2011). Additional E's have also been described, including economics, emergency response, enablement, and ergonomics, situated in the safe system approach (Plant, McIlroy and Stanton, 2018). Years of European research have contributed to a decision-making tool to help road safety experts identify appropriate interventions (Thomas et al., 2016). These interventions tend to be recommended at a national or even European level rather than identifying specific areas that could benefit.

Geodemographics entails the segmentation of society into smaller groups based on combinations of data available in the public domain, including census and commercial data, financial activity, purchasing history or survey responses (Burns et al., 2018). The use of geodemographic profiles can be more accurate at predicting behaviour and attitudes than other conventional demographic information such as gender, age or occupation (Webber and Burrows 2018). Members of the population tend to be more like the people who reside around them than people of their age and gender living in other areas (Webber and Burrows, 2018). Geodemographic profiling has been applied to STATS19 data on a population basis and is available through the MAST system (http://www.roadsafetyanalysis.org/mast-online/). Elements of social marketing are also being adopted as strategy by the Department of Transport (2015) but these are based on STATS19 casualty data rather than culpability or injury severity data. There is little experience of how to effectively use culpability and injury severity data to target road safety strategies.

The literature suggests that there is an opportunity to explore the concept of targeting road safety interventions at drivers culpable for causing the most serious injuries on the roads. This would provide Cambridgeshire County Council with a different approach to road safety interventions if the drivers could be specifically targeted rather than relying on national or large local measures.
This study links five years of the existing datasets STATS19 and TARN to identify a sample of drivers in Cambridgeshire who have been involved in a serious collision, identified clinically at the MAIS3+ severity, and explore the differences in culpability and geodemographics.

Aim and objectives.

The overall aim of the study is to inform innovative approaches to road safety intervention to reduce the numbers of people killed or seriously injured in road traffic collisions. This was to be achieved through several objectives exploring different methodologies to identify whether drivers culpable of causing serious injury collisions are different to non-culpable drivers in Cambridgeshire.

- Use data linkage to link TARN and STATS19 data to clinically define serious (MAIS3+) road traffic collisions.
- Apply culpability analysis to all drivers involved in MAIS3+ and fatal collisions using STATS19 data.
- Apply geo-demographic profiling to Cambridgeshire drivers involved in MAIS3+ and fatal collisions.
- Analyse the linked data to investigate the culpability and geo-demographic profiles of drivers to explore any differences in Cambridgeshire resident drivers that might influence injury prevention strategies.

Methodology

This study was reliant on various methods to identify a ‘study sample’ for data analysis. The aim was to identify drivers involved in a MAIS3+ or fatal road collision who were categorised as culpable or not and have a geodemographic profile. The datasets available for the study were the police road collision data (STATS19) and Trauma Audit Research Network (TARN) data.
Cambridgeshire County Council supplied the full (anonymised) STATS19 dataset for the study, for the period 1st April 2012 to 31st March 2017. The STATS19 dataset comprised 175 variables incorporating facts about the collision, vehicle and casualty as well as providing additional text or ‘narrative’ about the collision characteristics recorded on scene by the police officer. The narrative is not routinely released as part of STATS19 but was requested for this study.
Cambridge University Hospitals Trust provided anonymised Trauma Audit Research Network (TARN) data encompassing the period 1st April 2012 to 31st March 2017 for all road traffic collision patients attending Addenbrookes Hospital. However, to enable data linkage the first four digits (postcode district) were also supplied. Inclusion into the TARN dataset is based on the trauma patient achieving certain criteria including a length of stay of 3 days or more or, admitted to a high dependency unit, or death of the trauma patient during the hospital admission. This included patients transferred into the Trauma centre at Addenbrookes Hospital from surrounding hospitals.

Data sharing agreements were drawn up and approved by Loughborough University and the data providers for the purposes of undertaking this research. Ethics approval was also granted by Loughborough University to conduct the research. Data was transferred and stored securely and encrypted as part of the data sharing protocols.

The series of methods were used in this study to identify drivers resident in Cambridgeshire who were involved in MASI3+ and fatal collisions (Figure 2).

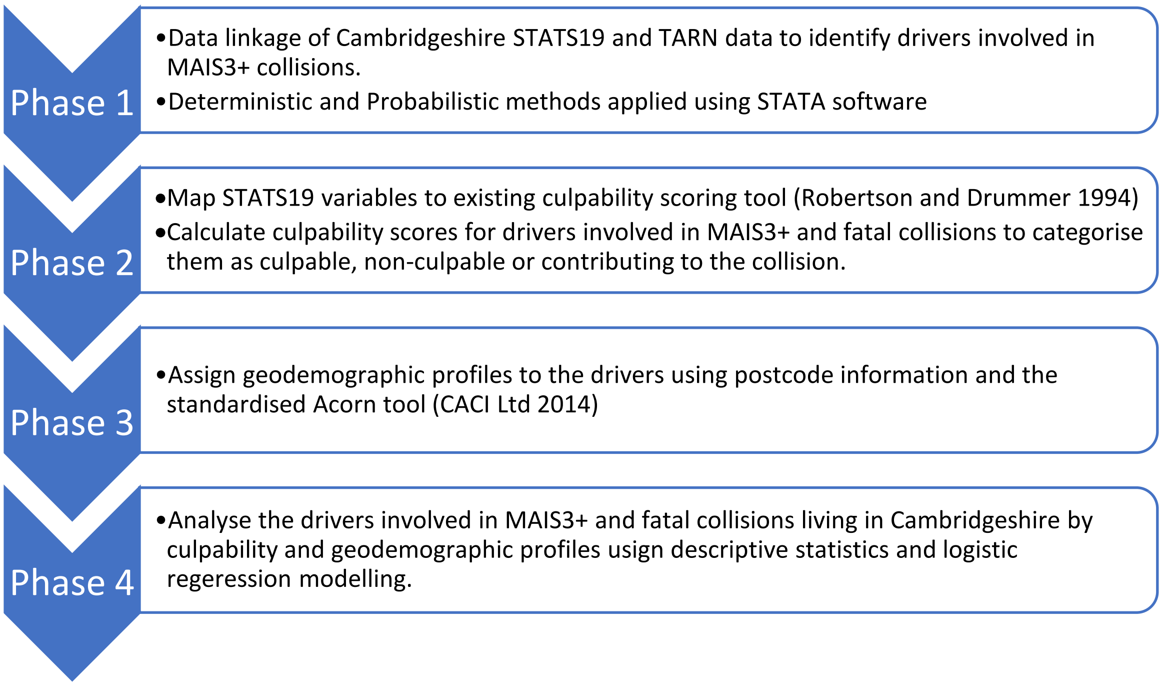


Figure 2: Methods used to identify Cambridgeshire drivers

Phase 1: Data linkage

STATS19 and TARN data for the Cambridgeshire area during the period 1st April 2012 to 31st March 2017 were linked using STATA software. The process used to implement the data linkage was the 'Data Integration Protocol in Ten-steps (DIPIT) (Dipnell et al., 2014). DIPIT follows a clear and systematic process to minimise the production of erroneous outcomes moving from initial data cleaning and organising to recording all stages of the linkage process and reviewing the linked data to ensure it is fit for purpose (Appendix 1).
There were no unique common identifiers in either the two datasets to expedite the linkage, therefore quasi-identifiers were used comprising four common variables (incident date, age, gender and the first string of the home postcode). A deterministic approach was used initially followed by a probabilistic approach (Wasi and Flaaen, 2013, Acock, 2016, Tromp et al., 2011, p. 565). The probabilistic approach fixed the gender, postcode and incident date and only adjusted the weighting for age (+ or- 5 years) to minimise the risk of obtaining false positives (Hagger-Johnson et al. 2017). The initial linkage process matched a total of 324 collision records using deterministic and probabilistic methods (Table 2). At this point there was a total of 412 motor vehicle drivers identified as being involved in a MAIS3+ serious injury collision in the Cambridgeshire area between 1st April 2012 and 31st March 2017.

Table 2: Results of the data linkage process

	Linked collision records	Linked MAIS3+ collision records	Linked MAIS3+ collision records involving motor vehicles	Number of motor vehicle drivers involved in MAIS3+ collisions
Deterministic	295	256	230	399
Probabilistic	29	12	8	13
Total linked records	324	268	238	412

At this point fatal collisions were included in the dataset to explore the number of collisions identified as MAIS3+ casualties from the linkage but were from fatal collisions. Duplicates were removed and any MAIS3+ casualty was removed that was in a fatal collision thus defining only those collisions at MAIS3+ level and those collisions at a fatal severity. This resulted in a total of 660 motor vehicle drivers involved in 368 collisions.

Phase 2: Culpability scoring

No culpability scoring tool is available to analyse STATS19 data and as a result existing tools were identified and applied to the dataset. The main culpability tool used in the literature is the Robertson and Drummer responsibility tool (1994) and was also used in this study. It comprises of eight factors (mitigating categories) which are assessed on a collision basis to assign a score to denote level of culpability. The categories include the condition of the road; the condition of the vehicle; the driving conditions, the type of accident; witness observations; road law obedience; the difficulty of task involved and the level of fatigue (Appendix 2). Each mitigating factor is given a score of one to four, where one is not mitigating through to four mitigating. To be valid there needs to be five or more mitigating factors present to obtain a score. The driver score starts at zero (and the scores are added together, therefore, the lowest score a driver can obtain is eight and the highest is 32. The scoring system works as such:

- culpable (scores 8-12)
- contributory (scores 13-15)
- nonculpable (scores >15)

STATS19 variables, contributory factors and the crash narrative were mapped to the Robertson and Drummer categories by the researcher and the process was validated by three collision experts to ensure the process and mapped outcomes were reliable. The experts were also tasked with scoring 10% of the collision records (n=40) to ensure inter-rater reliability.
It was not possible to map each category to STATS19 due to lack of information thus requiring assumptions to be made. The culpability tool contains a witness observation category which is not present in STATS19, but a police officer will consider witness observations at the scene and record these as contributory factors if appropriate. Therefore, many of the assessment criteria contained within the witness observation section for the scoring tool are contained within the contributory factors allocated to the individual collision report in STATS19.
Culpability scores were calculated for all drivers (n=660) involved in either a MAIS3+ (n=360) or fatal collision (n=300) (Table 3).

Table 3: Results of the culpability scoring using STATS19 variables and the Robertson and Drummer responsibility tool (1994)

Injury Severity	Culpable	Contributory	Non-culpable	Unable to score	Total Number of motor vehicle drivers
MAIS3+ linked data	231	27	98	4	360
Fatal	159	18	123	0	300
Total culpability	390	45	221	4	660

Phase 3: Geodemographic profiling of culpability scored drivers

The Acorn consumer classification geodemographic profiles were used which bases its profiles on a combination of social factors, behaviours, and interactions, allocated to individual postcodes (CACI Ltd, 2014). For Cambridgeshire there are 19706 valid postcode-to-geodemographic profiles in the Acorn system.

Postcodes of drivers with a culpability score (n=656) involved in MAIS3+ and fatal collisions were validated against the Office for National Statistics (2011) census data to ensure a correct postcode and categorise the drivers into local authorities. Of the 656 drivers, 92 did not have valid postcodes and of the remaining drivers with a valid postcode 367 were Cambridgeshire residents and 197 drivers had residential postcodes outside of the Cambridgeshire area. These 564 drivers with a valid postcode were subsequently mapped to the Acorn classification system postcodes and the corresponding geodemographic profile attached to them.

Acorn has three levels to the profile, with 'type' being the finest level of granularity providing a detailed profile compared to the group and category levels. Figure 3 illustrates an example of a geodemographic profile, 'Type 23', its group and the number of options in each level of granularity.

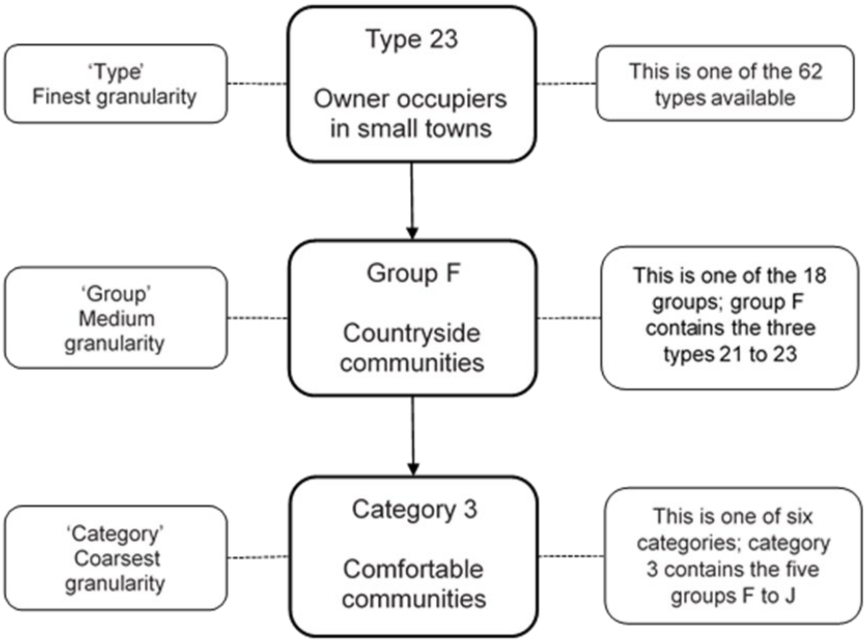


Figure 3: Example of an Acorn profile

Phase 4: Analysis Results

Descriptive analysis

Following the methodological process, a total of 564 drivers involved in a MAIS3+ or fatal collision in Cambridgeshire with a culpability score and geodemographic profile were identified. Some 367 of these (65%) were residents of Cambridgeshire. There were only 45 motor vehicle drivers considered to be contributory in the collision and these were collated with the culpable drivers for ease of analysis.

Table 4 presents the descriptive analyses and driver and collision characteristics. The majority of drivers were male (434,77%) and the overall mean age was 43 years (SD 17). Most were car drivers (366, 67%) and most were scored as being culpable or contributory to the collision (384, 68%).

Table 4: Characteristics of Cambridgeshire and non-Cambridgeshire drivers with a culpability score and Acorn profile

	Non- Cambridgeshire resident drivers (n=197)	Cambridgeshire resident drivers (n=367)	Total (n=564)
Gender (%) **Chi^2 = 1.81 *p* = .179**	M = 158 (80%) F = 39 (20%)	M = 276 (75%) F = 91 (25%)	M = 434 F = 130
Mean age (SD)	43 (18)	43 (16)	43 (17)
Fatal **MAIS3+ severity** **Chi^2 = 7.49 *p* = .006 df = 1**	97 100	137 230	234 330
Culpable and Contributory combined **Non-culpable** **Chi^2 = 0.94 *p* = .331 df = 1**	129 68	255 112	384 180
Vehicle type - **Motorcycle** **Car** **PCV/Agricultural*** **Goods Vehicle** **Chi^2 = 22.95 *p* = .000 df = 4** **Fisher's exact = .000**	 30 112 <10 50	 65 254 <10 37	 95 366 <10 87

*Number not disclosed as n <10

Geodemographic profiling

The distribution of Cambridgeshire drivers in the Acorn categories shows a higher frequency in the 'comfortable community' (n=132) residents and very few in the non-private households (<5). The distribution of Acorn categories and groups can be seen in Figure 4.

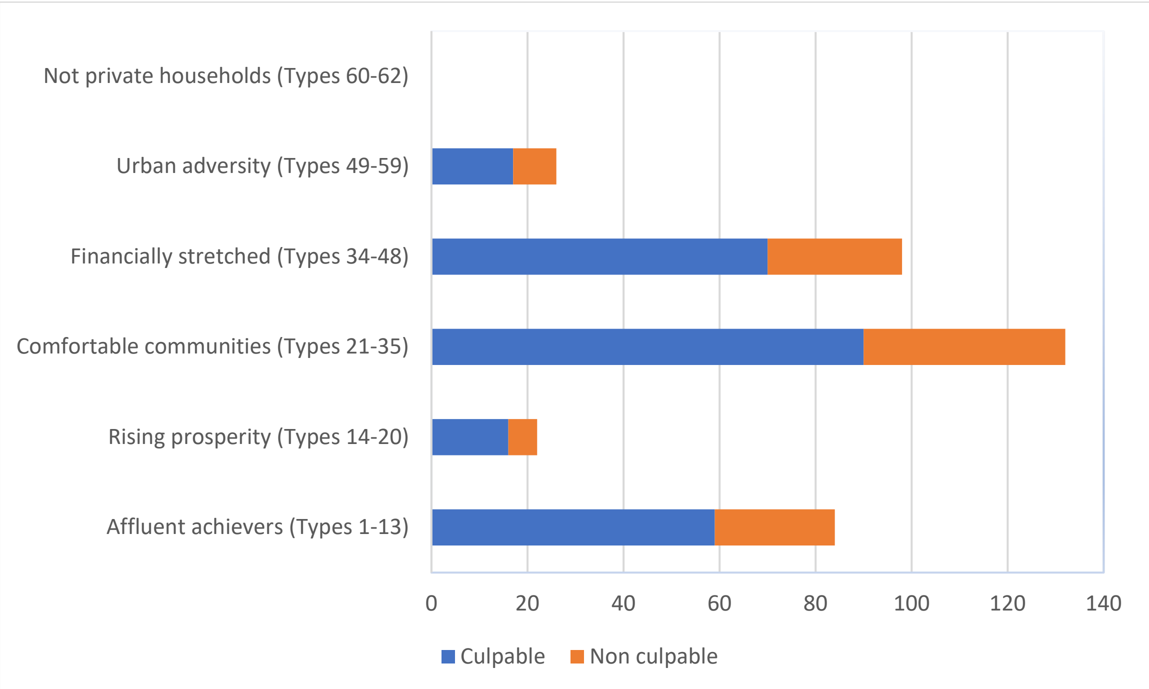


Figure 4: Frequency of Acorn categories for all Cambridgeshire resident drivers (n=367)

Overall, there were 47 Types represented by the Cambridgeshire Drivers, a full list of Acorn Types is presented in Appendix 3, but due to small numbers are reported at the group level.

Risk index

Risk indexation is often used within geodemographic analysis in numerous contexts, to present the relationship between a particular sub-population and the whole population (Anderson, 2005; 2010; Ashby and Longley, 2005; Farr, Wardlaw and Jones, 2008; CACI Ltd, 2014; Quddus, 2015; Loo and Anderson, 2016). Risk indexation was used to explore the Acorn Types in the fatal and MAIS3+ collision sample compared to the Acorn Types in the general Cambridgeshire area. The absolute numbers in the sample are small, and the figures can only illustrate the potential of using this method on large scale data to identify culpable drivers for targeted road safety campaigns.
A risk index was calculated to generate a ratio score whereby scores over 100 in an Acorn Type suggest over-representation in the driver collision population compared to the overall Cambridgeshire population (Appendix 4). Therefore, the higher the risk index the more over-represented the Acorn type was in the motor vehicle driver data and conversely scores below 100 show under-represented Acorn types. Due to the large number of Types reported the most frequently recorded Types have been used in the risk index which account for 50-60% of drivers in each injury collision group.

Figures 5 and 6 uses the top ten most frequently recorded Acorn types for culpable drivers residing in Cambridgeshire for fatal and MAIS3+ collisions to produce the risk index graphs. Numbers are not reported due to the small n <5 frequency in some Types.

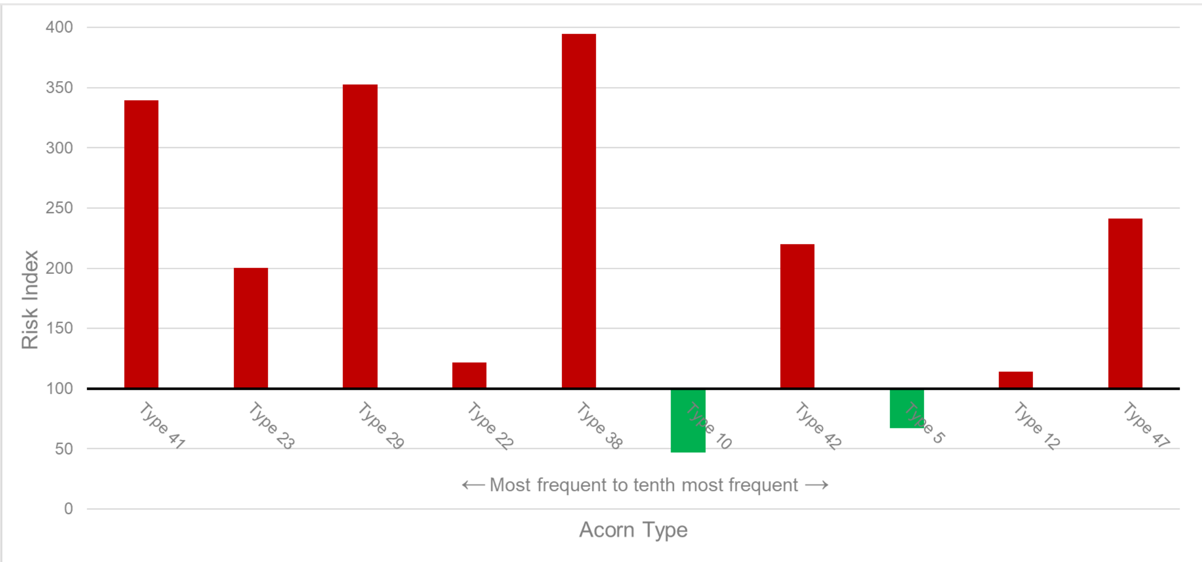


Figure 5: Risk index of Acorn Types for culpable drivers involved in fatal collisions.

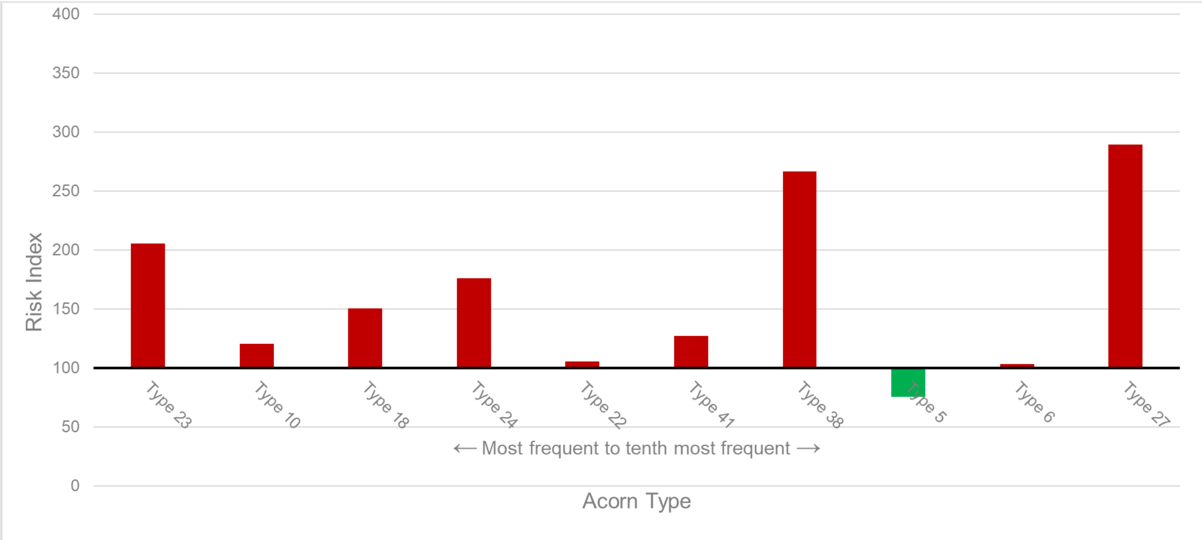


Figure 6: Risk index of Acorn Types for culpable drivers involved in MAIS3+ collisions.

From the small dataset the graphs suggest that there are some Acorn Types of interest. Type 41 culpable drivers in fatal collisions is overrepresented and has a higher risk (> 300 index) compared to the general Cambridgeshire population. Type 41 describes 'Labouring semi-rural estates.' For MAIS 3+ serious injury collisions the most frequent occurring Type and overrepresented compared to the general Cambridgeshire population was Type 23 with a risk index >200 (Owner occupiers in small towns and villages). Interestingly there were some Types where the risk index was lower specifically Type 10 (Better-off villagers) for fatal collisions and Type 5 (Wealthy countryside commuters) for both fatal and MAIS3+ collisions. This would be interesting to explore further with larger datasets to understand how typical the over representation of culpable drivers is.

Logistic regression analysis

Logistic regression was used to analyse the sample of Cambridgeshire drivers (n=367) and their likelihood of being culpable or not. The results found that motorbike drivers were more likely to be culpable than car drivers (OR 5.05 (95% CI 1.84 - 13.89) p.002) and Acorn Type 6 drivers were less likely to be culpable compared to the most frequent Type 23 (OR 0.12 (95% CI 0.02 - 0.62) p.012). There were some other notable findings, although not significant, suggesting there might be a tendency for greater collision culpability if drivers were either under 26 years or over 76-years, and lesser collision culpability if they drove goods vehicle or were Types 10, 22, 33, or 42 (see Appendix 3).

Discussion

Overall a sample of drivers with a culpability score involved in a MAIS3+ or fatal collision could be identified using the methodological processes outlined above. The use of data linkage is not new but importantly has enabled drivers of clinically serious (MAIS3+) collisions to be identified from the STATS19 and TARN linked dataset.

There have been improvements in the Police reporting of collision injuries with the adoption of the CRASH[2] or COPA[3] electronic collision reporting systems used by some Police Forces. These are Injury Reporting Systems offering a drop-down menu of injury descriptions that are predefined for the

[2] Collision Recording And SHaring (CRASH)

[3] Case Overview and Preparation Application

reporting officers as serious or slight. Although they do not provide clinically defined serious injury definition, they do offer the opportunity for consistent reporting. Additional statistical modelling work has been conducted on a national scale to adjust collision statistics to better reflect actual numbers of seriously injured on the roads (Braunholtz and Elliott 2019). Again, these are not clinically defined serious injuries and the main option for accurate reporting at the national level is to use HES data to generate MAIS3+ definitions from the ICD 10 codes (DfT 2015). All ICD 10 codes have been expertly mapped to the Abbreviated Injury Scale allowing for the identification of serious injuries (MAIS 3+) (Barnes et al 2020, Loftis et al 2016).

The sample of Cambridgeshire drivers was broadly like the UK's driver profile of predominantly male and drivers of cars (DfT 2019b, DfT 2019c). Culpability scoring of the drivers was a novel approach used to identify drivers who were culpable or contributory to the collision through their driving actions or decisions made. Previous studies have used culpability scoring to determine the impact of drink and drug driving on collision culpability (Terhune 1983, Robertson and Drummer 1994, Brubacher, Chan and Asbridge 2012). All have been developed from the original Terhune (1983) model and each revision has tried to remove the subjectivity in the tool with the latest iteration (Brubacher, Chan and Asbridge 2012) designed to work with bulk data from the Canadian national collision database. The most reported model in the literature is the Robertson and Drummer responsibility tool (Salmi, Orriols and Lagarde, 2014) which was used in this study. The challenge of using the tool was mapping STATS19 variables to this Australian tool where differences exist e.g., road surfaces and weather conditions. To enable a map there were many 'combination of STATS19 variables and /or contributory factors used to code any one category in the original tool. Notably the 'narrative' about the collision was used for some of the mapping from STATS19 to the tool to identify specific factors, for example the struck or striking vehicle. However, despite the narratives usefulness for this study it is not normally available for analysis in the released STATS19 data. In agreement with other studies, assigning culpability is not straightforward to determine and subjectivity of the researcher may have influenced the judgement of culpability (Köhnken and Brockmann, 1987). What was interesting in this study was the suggestion that only a few variables in STATS19 might be required to ascertain driver culpability in the collision. This provides the opportunity to automate the culpability scoring of STATS19 collisions, rather than having to individually map to an existing tool to provide a score.

The motor vehicle drivers were similar in age and gender distribution whether a Cambridgeshire resident or not (Table 4); however non-Cambridgeshire resident drivers were more likely to be involved in fatal collisions also more likely to be driving a goods vehicle compared to the Cambridgeshire residents. This might reflect the major road network around Cambridgeshire and the potential higher speeds driven on these roads.

The logistic regression identified the higher risk of motorcycle drivers in being culpable for the collision. The vulnerability of motorcyclists is commonly reported (DfT 2019a) and road safety campaigns have been aimed specifically at them e.g., Think Bike. Acorn Types tend not to be reported but seems they could provide useful information for targeting specific driver populations. Type 6 drivers tended to be at lesser risk and describes 'professional or managerial commuters living in modern estates or large houses'.

Geodemographic profiling of drivers using ACORN (CACI Ltd 2014) incorporates government and consumer marketing preferences to inform companies of marketing choices for the ACORN Types. At the category level for Acorn profiles higher frequencies of drivers were noted as being in the comfortable communities, affluent achievers or financially stretched suggesting some variation in the driver population of Cambridgeshire (Figure 4). Focussing on the Acorn groups there were four that had higher frequencies of culpability compared to the non-culpable drivers and were distributed between the ‘Executive wealth’ (B), ‘Mature money’ (C), ‘Countryside commuters’ (F) and ‘Striving families’ (M) (Figure 5). Type 23 was the most frequent Type for all drivers in the sample and are described as ‘owner occupiers in small villages’ in the Countryside commuter group (F) and tend to be older couples owning their homes. There is less reliance on the internet and smartphone use and shopping is at well know department stores although budget supermarkets like Aldi might be used. When considering the use of risk indexation there was some variation in the culpable fatal drivers compared to the MAIS3+ drivers. The wealthier Types appeared to be less likely to be represented in the Cambridgeshire population for fatal collisions (Types 5 and 10) and Type 5 for MAIS3+ collisions. Interestingly Type 29 (Established suburbs, older families) appear to be risky drivers in fatal collisions but not for MAIS3+ collisions where Type 27 (Suburban semis, conventional attitudes) was present but both Types are in the ‘Steady Neighbourhoods’ Group. The descriptors of the ACORN Types help to expand the understanding of how geodemographic profiling could be used to target specific populations through marketing choices. It also does show the difficulty in trying to target specific drivers in a culpable driver population because of the spread of Types and small sample used in this study. However, there were some interesting detail about the Types that could be used to stream road safety interventions at specific driver groups rather than relying on national campaigns, for example, delivering road safety interventions through supermarkets most likely visited by that Type.

Conclusion

This study aimed to identify the potential of using geodemographic profiling to provide targeted road safety interventions to those drivers causing MAIS3+ serious injury collisions in Cambridgeshire. The process involved several stages to identify the drivers involved in serious (MAIS3+) collisions and manually determine their culpability prior to geodemographic profiling. The study showed as proof of concept it is possible to use geodemographic profiling to identify those drivers culpable of causing serious and fatal injury collisions. However, there are challenges to address before it could be used in everyday practise to design targeted road safety interventions by local Council Road Safety Teams. The use of data linkage is not novel but is time consuming and reliant on the quality of the routinely collected data and the level of data governance and data sharing agreements required to undertake any linkage. Fortunately, Cambridgeshire and Peterborough Road Safety Partnership data sharing agreement enabled access to both TARN and STATS19 data for the linkage which enabled the identification of MAIS3+ injuries sustained in collisions and the collision details. Using MAIS3+ clinical definition ensured the most serious collisions were included in the data analysis but consequently they only represented a small proportion of collisions that occurred on the roads in Cambridgeshire. Analysis on a small set of drivers was valuable to determine if there were any geodemographic differences between drivers causing serious or fatal collisions but there is potential to widen the analysis using clinically defined injuries at the moderate (MAIS2) or minor level (MAIS1).

This analysis would be possible by extending the linkage to include STATS19, HES and TARN data. This would enable all collision severities to be analysed to fully understand the relationship between driver culpability and geodemographic profiling. HES is limited for identifying injury severity as it uses ICD and not AIS. A recent expert derived ICD_AIS map (Barnes et al., 2020, Loftis et al., 2016) has been developed to convert ICD to AIS that would enable clinically defined MAIS3+ injuries to be identified and provide an opportunity to analyse large datasets of all collision severities. To enhance the linkage process establishing unique identifiers would benefit the process and could be the NHS number.

Culpability was also determined for the driver sample and used STATS19 variables to calculate a culpability score after mapping onto the Robertson and Drummer responsibility tool. To date STATS19 has not been mapped to any culpability tools to determine driver culpability in MAIS3+ or fatal collisions. Culpability for this study was determined by manually reviewing each drivers' contribution to the collision using the STATS19 variables mapped to the existing Robertson and Drummer responsibility tool. This was a time-consuming process but what it did identify was that there were consistent variables, often contributory factors and vehicle manoeuvres in the collisions that rendered the driver culpable in the collision. This suggested that there was a possibility to automate the culpability scoring of STATS19 using specific variable combinations. The culpability scoring in this study specifically focussed on drivers of motor vehicles and not on other road users which will need to be considered. This work is being undertaken separately to establish whether a STATS19 derived 'tool' is a viable option and will work to consider its validity and reliability. The manually rated culpability cases from this study can be used to validate the results. If successful, this would provide an additional tool for Road Safety teams to use to identify those drivers causing collisions, and further explore geodemographic profiling to understand any patterns of behaviour. Additionally, this could be applied at the government level and released for research purposes alongside the linkage work with HES.

The use of geodemographic profiling of culpable drivers is a novel approach to target road safety interventions and this study identified that some Types appear to be overrepresented in being culpable for fatal and MAIS3+ collisions. However, the sample of MAIS3+ collisions in this study was too small to make inferences of real differences in comparison to large geodemographic studies (Quddus 2015). Furthermore, the focus was on MAIS3+ collisions and did not consider culpable drivers for slight collision severities. However, this study has shown there is potential to use the Acorn Type profiles to target specific populations. As an example, the older generation, known to be vulnerable (Clarke et al., 2010) could be focused on as they seem to be present in several Types across the financial Groups identified in this study. Exploring the methods applied in this study provides the proof of concept that culpable drivers can be identified, and the exploration of geodemographic profiles might have an influence on culpability and would benefit from targeted road safety interventions. Moreover, the potential to use this methodology on all collision severities would enhance the understanding of drivers involved and has the potential to match specific interventions to specific geodemographic types.

To enable the use of this methodology in local Councils would require data analysts to manipulate the data and run the culpability syntax. At the STATS19 level the injury severities of slight, serious, and fatal would be the injury definitions unless a linked injury (TARN or HES) and STATS19 was used. Appending the geodemographic profiles to the dataset would also be required providing a powerful dataset to identify culpable drivers living in a geodemographic area and mapping these to specific local road safety interventions.

If the methodology of assigning culpability to drivers causing collisions were successful at a national level, there would be options to append this information to the existing reporting style for specific ACORN Types. For example, Figure 7 below illustrates Type 23 which suggests that people in this type have slightly higher than average BMI on the risk index and adding crash culpability risk in a similar way might provide useful information and opportunities to target road safety interventions at these Types.

Figure 7 Type 23 infographic reproduced from The Acorn User Guide (CACI 2014 page 44)

Recommendations

Opportunities for Practitioners

For road safety practitioners this method will allow profiles of all drivers to be compiled to enable targeting interventions at relevant sections of the population. The ability to undertake this type of work will rely on having routinely collected data linked together and then manipulated to identify culpable drivers. Varying geodemographic tools are used at local Councils and could be appended to a linked collision dataset. This project specifically focussed on serious injury collisions and the potential to reduce these on the roads as part of a road safety target strategy. However, it can be used to identify all culpable drivers in all collision severities once the STATS19 tool becomes automated.

To achieve this would require the following.

- Routinely link STATS19 with TARN/HES data to identify drivers involved in serious MAIS3+ collisions at the local level (This could be nationally with data sharing agreements)
- Recommend the collection of unique identifiers to enhance the linkage process e.g., using the NHS number in STATS19.
- Apply a validated STATS19 culpable score to the linked dataset to identify the drivers causing collisions. This would rely on having a syntax available in the local analysis software format to enable routine running of the culpability scoring.
- Use geodemographic profiling of the culpable drivers to evaluate whether there are differences in injury outcome collisions to drive road safety intervention targeting.

Next Steps

- Develop an automated culpability tool derived from STATS19 variables and contributory factors to enable application to mass datasets rather than hand scoring cases.
- To further explore STATS19 variables to assign culpability for all collision severities and explore other factors that could influence the culpability outcome.

- Evaluate the reliability of using the MAIS3+ map from linked HES-STATS19 data
- Evaluate STATS19 and HES linked data to explore the potential for using national data to identify culpable drivers and geodemographic profiling for all collision severities.
- Apply the methods to other geographical areas to determine the generalisability of the findings and explore trends in culpable driver populations.
- Devise a methodology for local Councils to utilise routine data collected for targeted road safety interventions.
- Hold focus groups with specific Acorn Types identified in the analysis to establish how they currently receive road safety messages and the potential of receiving local targeted messages.

References

Acock, A. C. (2016) *A gentle introduction to Stata*. 5th edn. College Station, Texas: Stata Press.
af Wåhlberg, A. E. (2002) 'Characteristics of low speed accidents with buses in public transport', *Accident Analysis and Prevention*, 34(1), pp. 637–647.
af Wåhlberg, A. E. (2009) *Driver Behaviour and Accident Research Methodology: Unresolved Problems*. Farnham, Surrey: Ashgate Publishing Limited.
Anderson, T. K. (2005) *Spatial Variations in Road Collision Propensities in London*. Available at: http://www.casa.ucl.ac.uk/working_papers/paper96.pdf (Accessed: 31 October 2018).
Anderson, T. K. (2010) 'Using geodemographics to measure and explain social and environment differences in road traffic accident risk', *Environment and Planning A*, 42(9), pp. 2186–2200.
Ashby, D. I. and Longley, P. A. (2005) 'Geocomputation, geodemographics and resource allocation for local policing', *Transactions in GIS*, 9(1), pp. 53–72.
Association for the Advancement of Automotive Medicine. *The Abbreviated Injury Scale-2005 Revision, Update 2008*. Des Plaines, Illinois: 2008.
Barnes J, Loftis KL, Jones L,PriceJP, Gillich PJ, Cook K, Brammer A, et al (2020) Development of an expert derived ICD-AIS map for serious AIS3+ injury identification, Traffic Injury Prevention, 21:3, 181-187, DOI: 10.1080/15389588.2020.1725494
Bernhoft, I. M. (2011) *Results from epidemiological research - prevalence, risk and characteristics of impaired drivers*. DRUID project deliverable. Available at: http://www.forskningsdatabasen.dk/en/catalog/2389475090 (Accessed: 31 October 2018).
Brault, M. and Dussault, C. (2002) 'Comparison of two responsibility analysis methods regarding fatal crashes', *Alcohol, Drugs and Traffic Safety*, (5), pp. 423–430.
Braunholtz D. and Elliott D. (2019) Estimating and adjusting for changes in the method of severity reporting for road accidents and casualty data. *ONS*
Breen J (2012) High-level Group on Road Safety Consultation on the development of the injuries strategy draft: 1st October 2012 2nd Working Document: Next steps in the development of the injuries strategy.
https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/ser_inj/ser_inj_breen.pdf
Broughton, J., Amoros, E., Bos, N. M., Evgenikos, P., Hoeglinger, S., Holló, P., Pérez, C. and Tecl, J. (2008) *Estimation real number of road accident casualties.* SafetyNet project deliverable. Available at: https://www.narcis.nl/publication/RecordID/oai:library.swov.nl:328572 (Accessed: 1 November 2018).
Broughton, J., Keigan, M., Yannis, G., Evgenikos, P., Chaziris, A., Papadimitriou, E., Bos, N. M., Hoeglinger, S., Pérez, K., Amoros, E., Holló, P. and Tecl, J. (2010) 'Estimation of the real number of road casualties in Europe', *Safety Science*, 48(3), pp. 365–371.
Brubacher, J., Chan, H. and Asbridge, M. (2012) 'Development and Validation of a Crash Culpability Scoring Tool', *Traffic Injury Prevention*, 13(3), pp. 219–229.
Burns, L., See, L., Heppenstall, A. and Birkin, M. (2018) 'Developing an Individual-level Geodemographic Classification', *Applied Spatial Analysis and Policy*. Springer Netherlands, 11(3), pp. 417–437.
CACI Limited (2014) *The Acorn User Guide: The consumer classification*. London. Available at: https://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf (Accessed: 9 January 2019).
Cambridgeshire Insight (2019) Cambridgeshire Road Traffic Collision Counts. Available at: https://data.cambridgeshireinsight.org.uk/dataset/cambridgeshire-road-traffic-collision-counts (Accessed: 9 December 2019).
Clark DE, Osler TM and Hahn DR. *ICDPIC: Stata module to provide methods for translating International Classification of Diseases (Ninth Revision) diagnosis codes into standard injury categories and/or scores.* Statistical Software Components S457028, Boston College Department of Economics. 2009.
Clarke, D. D., Ward, P., Bartle, C. and Truman, W. (2010) 'Older drivers' road traffic crashes in the UK', *Accident Analysis and Prevention*, 42(4), pp. 1018–1024.

Cercarelli, L. R., Arnold, P. K., Rosman, D. L., Sleet, D. and Thornett, M. L. (1992) 'Travel exposure and choice of comparison crashes for examining motorcycle conspicuity by analysis of crash data', *Accident Analysis and Prevention*, 24(4), pp. 363–368.
Craig, A., Tran, Y., Guest, R., Gopinath, B., Jagnoor, J., Bryant, R. A., Collie, A., Tate, R., Kenardy, J., Middleton, J. W. and Cameron, I. (2016) 'Psychological impact of injuries sustained in motor vehicle crashes: Systematic review and meta-analysis' BMJ Open 2016;6:e011993. doi: 10.1136/bmjopen-2016-011993
Cryer, C. P., Westrup, S., Cook, A. C., Ashwell, V., Bridger, P. and Clarke, C. (2001) 'Investigation of bias after data linkage of hospital admissions data to police road traffic crash reports', *Injury Prevention*. BMJ Publishing Group Ltd, 7(3), pp. 234–241.
Department for Transport (2019a) *Reported road casualties in Great Britain: 2018 annual report*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/834585/reported-road-casualties-annual-report-2018.pdf (Accessed: 23 April 2020)
Department for Transport (2019a) *Reported road casualties in Great Britain: 2018 annual report*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/834585/reported-road-casualties-annual-report-2018.pdf (Accessed: 23 April 2020)
Department for Transport (2019b) *Contributory factors for reported road accidents (RAS50)*. Available at: https://www.gov.uk/government/statistical-data-sets/ras50-contributory-factors (Accessed: 27 March 2019).
Department for Transport (2012) *Linking Police and Hospital data on Road Accidents in England: 1999 to 2009 results*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/230598/hes-linkage.pdf (Accessed: 19 November 2018).
Department for Transport (2015) *Estimating clinically seriously injured (MAIS3+) road casualties in the UK*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/556648/rrcgb2015-03.pdf.
Department for Transport (2015) THINK! Campaign marketing plan 2015-17. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/462802/think-marketing-plan-2015-2016-and-2016-2017.pdf
Dipnall, J. F., Berk, M., Jacka, F. N., Williams, L. J., Dodd, S. and Pasco, J. A. (2014) 'Data Integration Protocol In Ten-steps (DIPIT): A new standard for medical researchers', *Methods*, 69(3), pp. 237–246.
Edwards, P., Green, J., Roberts, I., Grundy, C. and Lachowycz, K. (2006) Deprivation and Road Safety in London A report to the London Road Safety Unit Part A: Relationships and Risks Section 1: Is there a link between deprivation and road traffic injury in London? London: Transport for London. Available at: http://content.tfl.gov.uk/deprivation-and-road-safety.pdf (Accessed: 1 November 2018).
Elvik, R., Høye, A., Vaa, T. and Sørensen, M. (2009) *The Handbook of Road Safety Measures*. 2nd edn. Bingley: Emerald Publishing.
European Commission (2018) European Road Safety Observatory (ERSO) Serious Injuries https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/ersosynthesis2018-seriousinjuries.pdf
European Center for Injury Prevention (ECIP). *Algorithm to transform ICD-10 codes into AIS 90 (98 update) and ISS*. Pamplona, Spain. 2006.
Farr, M., Wardlaw, J. and Jones, C. (2008) 'Tackling health inequalities using geodemographics: a social marketing approach', *International Journal of Market Research*, 50(4), pp. 449–467.
Fell, J. C. (1976) 'A Motor Vehicle Accident Causal System: The Human Element', *Human Factors*, 18(1), pp. 85–94.

Graham, D., Glaister, S. and Anderson, R. (2005) ‘The effects of area deprivation on the incidence of child and adult pedestrian casualties in England’, *Accident Analysis and Prevention*, 37(1), pp. 125–135.
Groeger, J. A. (2011) ‘How Many E’s in Road Safety?’, in Porter, B. E. (ed.) Loo, B. P. Y. and Anderson, T. K. (2016) *Spatial Analysis Methods of Road Traffic Collisions*. London: CRC Press.
Guest, R., Tran, Y., Gopinath, B., Cameron, I. D. and Craig, A. (2016) ‘Psychological distress following a motor vehicle crash: A systematic review of preventative interventions’, *Injury, International Journal of the Care of the Injured*, 47, pp. 2415–2423.
Hagger-Johnson, G., Harron, K., Goldstein, H., Aldridge, R. and Gilbert, R. (2017) ‘Probabilistic linking to enhance deterministic algorithms and reduce linkage errors in hospital administrative data’, *Journal of Innovation in Health Informatics*, 24(2), pp. 234–246.
High Level Group on Road Safety consultation on the development of the injuries strategy. *2nd Working Document: next steps in the development of the injuries strategy final.* November 2012., Available at: https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/ser_inj/ser_inj_breen.pdf.
Imprialou, M. and Quddus, M. A. (2017) ‘Crash data quality for road safety research: Current state and future directions’, *Accident Analysis and Prevention*, In Press.
International Transport Forum (2011) Reporting on Serious Road Traffic Casualties Combining and using different data sources to improve understanding of non-fatal road traffic crashes. OECD/ITF Paris. Available at: https://www.itf-oecd.org/sites/default/files/docs/road-casualties-web.pdf (Accessed: 1 November 2018).
International Transport Forum (2018) *Road Safety Annual Report 2018*. Available at: www.itf-oecd.org/road-safety-annual-report-2018 (Accessed: 6 December 2019).
International Transport Forum and Organisation for Economic Co-operation and Development (2008) TOWARDS ZERO Ambitious Road Safety Targets and the Safe System Approach. OECD
Jarvis, S. N., Lowe, P. J., Avery, A., Levene, S. and Cormack, R. M. (2000) ‘Children are not goldfish-mark/recapture techniques and their application to injury data’, *Injury Prevention*, 6, pp. 46–50.
Jones, S. J., Begg, D. J. and Palmer, S. R. (2013) ‘Reducing young driver crash casualties in Great Britain-Use of routine police crash data to estimate the potential benefits of graduated driver licensing’, *International Journal of Injury Control and Safety Promotion*, 20(4), pp. 321–330.
Köhnken, G. and Brockmann, C. (1987) ‘Unspecific postevent information, attribution of responsibility, and eyewitness performance’, *Applied Cognitive Psychology*, 1(3), pp. 197–207.
Loftis KL, Price JP, Gillich PJ, Cookman KJ, Brammer AL, St Germain T, et al. Development of an expert based ICD-9-CM and ICD-10-CM map to AIS 2005 update 2008. *Traffic Inj Prev*. 2016;17(Suppl. 1):1–5.
Mathijssen, R. and Houwing, S. (2005) The prevalence and relative risk of drink and drug driving in the Netherlands: a case-control study in the Tilburg police district. SWOV Institute for Road Safety Research, The Netherlands. Available at: https://www.swov.nl/sites/default/files/publicaties/rapport/r-2005-09.pdf (Accessed: 1 November 2018).
Morris, A., Mackay, M., Wodzin, E. and Barnes, J. (2006) ‘Some injury scaling issues in UK crash research’, in *Proceedings of of the 2003 International IRCOBI Conference on the Biomechanics of Impact*. Lisbon, Portugal 24-26 September, pp. 283–291.
Office for National Statistics (2011) *2011 residential-based area classifications*. Available at: https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications (Accessed: 14 March 2019).
Parliamentary Advisory Council for Transport Safety (2016) *Safe System*. Available at: http://www.pacts.org.uk/safe-system/.
Pérez, K., Weijermars, W., Bos, N., Filtness, A. J., Bauer, R., Johannsen, H., Nuyttens, N., Pascal, L., Thomas, P. and Olabarria, M. (2019) ‘Implications of estimating road traffic serious injuries from hospital data’, *Accident Analysis and Prevention*. Elsevier Ltd, 130, pp. 125–135.

Plant, K. L., McIlroy, R. C. and Stanton, N. A. (2018) 'Taking a "7 E's" approach to road safety in the UK and beyond', in Charles, R. and Wilkinson, J. (eds) *Ergonomics and Human Factors*. Birmingham, 23-25 April.
Quddus, M. A. (2015) 'Effects of geodemographic profiles of drivers on their injury severity from traffic crashes using multilevel mixed-effects ordered logit model', *Transportation Research Record: Journal of the Transportation Research Board*, 2514, pp. 149–157.

Roberts, S. E., Vingilis, E., Wilk, P. and Seeley, J. (2008) 'A comparison of self-reported motor vehicle collision injuries compared with official collision data: An analysis of age and sex trends using the Canadian National Population Health Survey and Transport Canada data', *Accident Analysis and Prevention*, 40(2), pp. 559–566.
Robertson, M. D. and Drummer, O. H. (1994) 'Responsibility analysis: A methodology to study the effects of drugs in driving', *Accident Analysis and Prevention*, 26(2), pp. 243–247.
Steinbach, R., Edwards, P. and Grundy, C. (2013) 'The road most travelled: The geographic distribution of road traffic injuries in England', *International Journal of Health Geographics*, 12, pp. 1–7.
Terhune, K. W. (1983) 'An evaluation of responsibility analysis for assessing alcohol
Thomas, P., Filtness, A., Yannis, G., Papadimitriou, E., Theofilatos, A., Martensen, H. and Diapendaele, K. (2016) 'Developing the European road safety decision support system', in *7th Expert Symposium on Accident Research (ESAR)*. Medical School of Hannover, Hannover, Germany, 9- 10 June
Tromp, M., Ravelli, A. C., Bonsel, G. J., Hasman, A. and Reitsma, J. B. (2011) 'Results from simulated data sets: Probabilistic record linkage outperforms deterministic record linkage', *Journal of Clinical Epidemiology*, 64(5), pp. 565–572
University of Leicester (2005) The Cambridgeshire Trauma Audit and Research Project. Unpublished.
Wagenaar, W. A. and Reason, J. T. (1990) 'Types and tokens in road accident causation', *Ergonomics*, 33(10–11), pp. 1365–1375.
Wang, C., Quddus, M. A. and Ison, S. G. (2011) 'Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model', *Accident Analysis and Prevention*, 43, pp. 1979–1990.
Ward, H., Lyons, R., Gabbe, B., Thoreau, R., Pinder, L., and Macey, S., (2010) *Road Safety Research Report No. 119 Review of Police Road Casualty Injury Severity Classification* – A Feasibility Study. Dft. London
Wasi, N. and Flaaen, A. (2013) 'Record Linkage using STATA : Pre-processing , Linking and Reviewing Utilities', *The Stata Journal*, 15(3), pp. 672–697.
Webber, R. and Burrows, R. (2018) The predictive postcode : the geodemographic classification of British society. London: Sage Publishing.
West, R. and Hall, J. (1997) 'The role of personality and attitudes in traffic accident risk', *Applied Psychology: An International Review*, 46(3), pp. 253–264.
West, R. (1997) *Accident Script Analysis*. Transport Research Laboratory report 274, Prepared for Road Safety Division, Department of the Environment, Transport and the Regions. Available at: https://trl.co.uk/sites/default/files/TRL274.pdf.
World Health Organization (2016) *International Classification of Diseases, 11th Revision (ICD-11)*. World Health Organization. Available at: http://www.who.int/classifications/icd/en/ (Accessed: 2 November 2018).
World Health Organization (2018) *Global Status Report on Road Safety 2018*. Available at: https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/.
Yannis, G., Papadimitriou, E., Chaziris, A. and Broughton, J. (2014) 'Modeling road accident injury under-reporting in Europe', *European Transport Research Review*, 6, pp. 425–438.

Zonfrillo MR, Weaver AA, Gillich PJ, price JP, Stitzel JD. New methodology for an expert-designed map from International Classification of Diseases (ICD) to Abbreviated Injury Scale (AIS) 3+ severity injury. *Traffic Inj Prev.* 2015; 16: Sup2. S197-S200.

Appendices

Appendix 1: Data Integration* Protocol in Ten-steps (DIPIT)

DIPIT Step	Action	Strategy	Standard
1	Define the data requirements	• _ Define research hypotheses • _ Establish files to integrate • _ Assess data quality	Documentation of research hypotheses, files needed to integrate and data quality issues
2	Establish ethical, legal and privacy issues	Establish ethical, legal and privacy issues for each data file to integrate	Documentation of standards met
3	Order the files to integrate	Set up a flowchart for all files to be integrated, incorporating all file names	Flowchart of file hierarchy
4	Establish the file formats	Amend the flowchart in step 3 to document the file format for each file integrated and the final master file	Inclusion of all file formats in flowchart
5	Define the variables of interest	Create a table containing the variable of interest for research containing as a minimum: • _ Final variable name • _ Original variable name • _ Source file of variable • _ Preliminary file(s) for variable • _ Description of variable	Table of variables of interest for research incorporating a standard naming format, structured order and identification of file source
6	Table of variables of interest for research incorporating a standard naming format, structured order and identification of file source	Create a table containing the variable(s) links and linkage method(s) used containing as a minimum: • _ Link variable(s) • _ Method of linkage • _ Automation used (if applicable)	Table of data file links, variables used and linkage method
7	Document the integration* path	Document the structure of the path taken for integration* to include as a minimum: • _ The integration* of the primary files • _ The saving of the Master file format in a standard file naming structure • _ The variables of interest to be retained • _ The variables standard naming format • _ The merging of all files into the Master file • _ A log of statistics of the key variables, and missing data analysis	Documentation of path of data file integration* hierarchy incorporating primary and secondary files, logs and naming convention
8	Flowchart the type of integration*	Document on flowchart type of integration*: • _ one-to-one • _ many-to-one • _ one-to many • _ many-to-many	Method of integration* included in flowchart and linkages used
9	Document the integration* outcome	Define linkage quality measure. Table of mismatches of records by variable to contain as a minimum: • _ Variable name • _ Source of mismatch • _ Reason for mismatch	Documentation of degree of variable mismatches (e.g., log): which variables, percentage matched/mismatched. Document linkage quality measure (e.g., F-measure graphs)
10	Check variables and missing data	Initial data inspection to include as a minimum: • _ Analysis of key variable(s) • _ Missing data analysis	Document initial investigation of variables. Define minimum percentage of missing data acceptable for research based on industry convention and document future handling of missing data

(Dipnall et al., 2014, p. 239)

Appendix 2: STATS19 mapped variables to the Robertson and Drummer responsibility tool to determine culpability.

Robertson and Drummer tool	STATS 19 mapped variables	
1. Condition of Road		Score
Sealed road*	Assumed to be sealed unless stated in description	
Two or more lanes and smooth	Road type – single carriageway (6) or slip road (7) if more than one lane or one-way street (2) if more than one lane	1
Divided road	Road type – dual carriageway (3) or roundabout (1)	1
Two or more lanes and rough	Road type – single carriageway (6) or slip road (7) if more than one lane or one-way street (2) if more than one lane combined with contributory factor – poor of defective road surface (101) or Special Conditions at site – Road surface defective (5)	2
Unmarked, thin and smooth	Road type – single carriageway (6), slip road (7) or one-way street (2) if either does not have separate lanes	2
Unmarked, thin and rough	Road type – single carriageway (6), slip road (7) or one-way street (2) if either does not have separate lanes combined with contributory factor – poor of defective road surface (101) or Special Conditions at site – Road surface defective (5)	3
Unsealed road	Assumed to be sealed unless stated in description	
Smooth	Assumed to be sealed unless stated in description	2
Rough and/or corrugated	Assumed to be sealed unless stated in description combined with contributory factor – poor of defective road surface (101) or Special Conditions at site – Road surface defective (5)	3
2. Condition of Vehicle		
Roadworthy	No vehicle defect contributory factors	1
Unroadworthy (contribution to accident unclear)	Contributory factors 201-206 or 999 present but no indication in the description of their influence	2
Unroadworthy (contributing to accident)	Contributory factors 201-206 or 999 present with indication in the description of their influence	4
3. Driving Conditions		
Day	Light conditions variable – daylight (1)	
Clear and/or cloudy	Light conditions variable – daylight (1) combined with Weather conditions variable – Fine without high winds (1)	1
*Fog and/or mist, clear and windy (>40 kph)	Light conditions variable – daylight (1) combined with Weather conditions variable – Fine with high winds (4) or Fog or mist – if hazard (7)	2
*Visibility good and road wet	Light conditions variable – daylight (1) combined with Weather condition variable – Fine without high winds (1) and Contributory factor – Wet road (103)	2
Showers and/or rain	Light conditions variable – daylight (1) combined with Weather conditions variable – Rain without high winds (2) or Rain with high winds (5)	3
Night	Lighting conditions variable -Darkness: street lights present and lit (4) or Darkness: street lights present but unlit (5) or Darkness: no street lighting (6) or Darkness: street lighting unknown (7)	
†‡Clear	Lighting conditions variable -Darkness: street lights present and lit (4) or Darkness: street lights present but unlit (5) or Darkness: no street lighting (6) or Darkness: street lighting unknown (7) combined with Weather conditions variable – Fine without high winds (1)	1
‡Cloudy	No map	2
Fog/mist/showers/rain/ice/wind	Lighting conditions variable -Darkness: street lights present and lit (4) or Darkness: street lights present but unlit (5) or Darkness: no street lighting (6) or Darkness: street lighting unknown (7) combined with Weather conditions variable – Rain	3

	without high winds (2) or Snowing without high winds (3) or Fine with high winds (4) or Rain with high winds (5) or Snowing with high winds (6) or Fog or mist – if hazard (7)	
4. Type of Accident		
Single-vehicle		
No influence from other vehicles	Number of vehicles variable indicates one vehicle or if the number of vehicles variable indicates more than one vehicle but in examining the vehicle type variable only one of the vehicles is a motor vehicle	1
Influence from other vehicles	No map	3
Multi-vehicle		
Striking vehicle attempting to avoid	Number of vehicles variable indicates more than one and examining the vehicle type variable indicates more than one motor vehicle, combined with the first point of impact variable, the manoeuvres variable and content of the description.	2
Striking vehicle not attempting to avoid	Number of vehicles variable indicates more than one and examining the vehicle type variable indicates more than one motor vehicle, combined with the first point of impact variable, the manoeuvres variable and content of the description.	1
Struck vehicle in the wrong	Number of vehicles variable indicates more than one and examining the vehicle type variable indicates more than one motor vehicle, combined with the first point of impact variable, the manoeuvres variable and content of the description.	1
Struck vehicle in the right	Number of vehicles variable indicates more than one and examining the vehicle type variable indicates more than one motor vehicle, combined with the first point of impact variable, the manoeuvres variable and content of the description.	3
5. Witness Observations		
No apparent reason	No map	1
Reckless		
Swerving	Contributory factor 'swerved' (409)	1
Irregular driving	No map	1
Negligent		
Witnessed road infringement	See section six	1
Lack of road sense	Failing to take account of factors presented in the contributory factors presented in table 5.24	1
Vehicle fault	See section two or contributory factor codes 201-206 and 999	3
Driver not to blame	No variables, contributory factors or material in the description indicating the driver was at fault for the collision	4
6. Road Law Obedience		
Was driver obeying road laws?		
Yes	No offences indicated by contributory factors or variable codes	3
No	Breath test variable code one (positive), any of the contributory factor codes indicated in table 5.28, any defects indicated in section two, any combination of factors indicated in section five which may combine to indicate a standards of driving offence	1
7. Difficulty of Task Involved		
Straight road or sweeping bend	Contributory factors 108 or 703 not present	1
§Across lanes in	Not indicated directly by STATS19, see below	

Heavy traffic	Manoeuvre variable, left (07) or right (09) turn combined with the description indicating heavy traffic	2
Light traffic	Manoeuvre variable, left (07) or right (09) turn combined with the description indicating light traffic	1
Winding road/sharp bend/U-turn	Contributory factors 108 or 703 present	2
Overtaking	Manoeuvre variable, overtaking (13-15)	2
Avoiding unexpected traffic	No map	3
8. Level of Fatigue		
Only if mentioned in police reports	Contributory factor ‘fatigue’ (503) present	2
* Add 1 if road has been newly surfaced. † If in heavy traffic, add 1 point. ‡ If not listed, add 1 point. § Scores 1, if under the guidance of traffic signals.		

Appendix 3: ACORN categories and frequency of Cambridgeshire drivers for each Acorn Type

Category	Group		Types		Cambridgeshire drivers (frequency of Types)
1 Affluent Achievers	A	Lavish Lifestyles	1	Exclusive enclaves	0
			2	Metropolitan money	
			3	Large house luxury	
	B	Executive Wealth	4	Asset rich families	40
			5	Wealthy countryside commuters	
			6	Financially comfortable families	
			7	Affluent professionals	
			8	Prosperous suburban families	
			9	Well-off edge of towners	
	C	Mature Money	10	Better-off villagers	44
			11	Settled suburbia, older people	
			12	Retired and empty nesters	
			13	Upmarket downsizers	
2 Rising Prosperity	D	City Sophisticates	14	Townhouse cosmopolitans	1
			15	Younger professionals in smaller flats	
			16	Metropolitan professionals	
			17	Socialising young renters	
	E	Career Climbers	18	Career driven young families	21
			19	First time buyers in small, modern homes	
			20	Mixed metropolitan areas	
3 Comfortable Communities	F	Countryside Communities	21	Farms and cottages	63
			22	Larger families in rural areas	
			23	Owner occupiers in small towns and villages	
	G	Successful Suburbs	24	Comfortably-off families in modern housing	30
			25	Larger family homes, multi-ethnic areas	
			26	Semi-professional families, owner occupied neighbourhoods	
	H	Steady Neighbourhoods	27	Suburban semis, conventional attitudes	23
			28	Owner occupied terraces, average income	
			29	Established suburbs, older families	
	I	Comfortable Seniors	30	Older people, neat and tidy neighbourhoods	8
			31	Elderly singles in purpose-built accommodation	
	J	Starting Out	32	Educated families in terraces, young children	11
			33	Smaller houses and starter homes	
4 Financially Stretched	K	Student Life	34	Student flats and halls of residence	2
			35	Term-time terraces	
			36	Educated young people in flats and tenements	
	L	Modest Means	37	Low cost flats in suburban areas	
			38	Semi-skilled workers in traditional neighbourhoods	

			39	Fading owner occupied terraces	
			40	High occupancy terraces, many Asian families	29
	M	Striving Families	41	Labouring semi-rural estates	
			42	Struggling young families in post-war terraces	
			43	Families in right-to-buy estates	
			44	Post-war estates, limited means	49
	N	Poorer Pensioners	45	Pensioners in social housing, semis and terraces	
			46	Elderly people in social rented flats	
			47	Low income older people in smaller semis	
			48	Pensioners and singles in social rented flats	18
5 Urban Adversity	O	Young Hardship	49	Young families in low cost private flats	
			50	Struggling younger people in mixed tenure	
			51	Young people in small, low cost terraces	9
	P	Struggling Estates	52	Poorer families, many children, terraced housing	
			53	Low income terraces	
			54	Multi-ethnic, purpose-built estates	
			55	Deprived and ethnically diverse in flats	
			56	Low income large families in social rented semis	10
	Q	Difficult Circumstances	57	Social rented flats, families and single parents	
			58	Singles and young families, some receiving benefits	
			59	Deprived areas and high-rise flats	7
6 Not Private Households	R	Not Private Households	60	Active communal population	
			61	Inactive communal population	
			62	Business addresses without resident population	5

CACI Ltd 2014

Appendix 4: Calculation equations for the risk index

Step 1: calculate the proportion of the Acorn type being examined in the total of Acorn types within the Cambridgeshire population

$$Acorn\ type\ population\ proportion\ = \frac{Acorn\ type\ frequency\ in\ the\ population}{Total\ Acorn\ types\ present\ in\ the\ population}$$

Step 2: Calculate the expected frequency for each Acorn type within a sub-population

$$Expected\ frequency\ in\ the\ sub-population = Acorn\ type\ population\ proportion\ \times Sub-population\ size$$

Step 3: Calculate the risk

$$Risk\ index = \frac{Actual\ Acorn\ type\ frequency\ in\ the\ sub-population}{Expected\ Acorn\ type\ frequency\ in\ the\ sub-population} \times 100$$

Scores over 100 show over-representation in the sub-population compared to the whole population. The score works as a ratio whereby scores of 200 indicate there are twice as many Acorn Types in the sub-population as the distribution in the population would predict.