Investigating the Transition from Normal Driving to Safety Critical Scenarios

by

Evita Papazikou

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of

Doctor of Philosophy of Loughborough University

July, 2019

Abstract

Investigation of the correlation between factors associated with crash development has enabled the implementation of methods aiming to avert and control crash causation at various points within the crash sequence (Evans, 2006). Partitioning the crash sequence is important because intricated crash causation sequences can be deconstructed and effective prevention strategies can be suggested (Wu & Thor, 2015). Towards this purpose, Tingvall et al. (2009) documented the so-called integrated safety chain which described the change of crash risk on the basis of a developing sequence of events that led to a collision. This thesis examines the crash sequence development and thus, the transition from normal driving to safety critical scenarios.

The current research utilises Naturalistic Driving Studies (NDS) and more specifically Strategic Highway Research Program 2 NDS (SHRP2 NDS) data to investigate the crash sequence. Trip-based time series data covering 2.5 minutes prior to the events (crashes and near-crashes) and the corresponding driver and event data were extracted from the SHRP 2 NDS dataset by Virginia Tech Transportation Institute (VTTI). After the data cleaning, matching and transformation process, 773 events with 553 drivers were available for analysis. With the data sampled at 10 Hz, over 1 million data points were included to the final dataset. The analysis conducted in three stages regarding the time sequence in crash development. Firstly, the time period during normal driving stage was investigated, followed by the whole crash sequence and finally, the last time period towards safety critical scenarios was examined.

Safety indicators during normal driving were characterised and functional relationships, providing dynamic thresholds in relation to speed, for departure from normal driving were derived. Longitudinal and lateral acceleration, yaw rate and TTC presented different distributions across gender and age groups. Moreover, relevant safety indicators generated with an empirical process, were employed to examine the whole crash sequence development and recognise deviations from normal driving. The descriptive analysis revealed that yaw rate, longitudinal and lateral accelerations may be feasible determinant of crash risk in earlier stages. Therefore, in the last 30 seconds prior to events, the driver braking, and steering behaviour was explored by extracting events of relevant interest. Examining the events mean values and their duration, thresholds for emerging situations were proposed.

Lastly, TTC values were further investigated and their evolution during crash sequence was analysed by using multilevel mixed effects modelling. According to the random slope model that was estimated, TTC values are affected by vehicle type, longitudinal acceleration, speed, and time within the crash sequence expressed by the timestamp variable.

The outputs of this thesis can be adopted by insurance companies to formulate normal driving profiles for different driver groups, and also, by the automation industry to evaluate or design new collision avoidance or warning systems.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisors Professor Pete Thomas and Professor Mohammed Quddus for their valuable guidance and support throughout my doctoral studies. They dedicated time and effort in helping me to structure my ideas and overcome the difficulties this PhD entailed, while simultaneously, motivated my broader development as an independent researcher.

I would also like to thank Professor Andrew Morris for his input as an internal assessor to my research. He was always available to discuss part of my work and provide advice when I needed it. Moreover, I thank Steve Reeds for helping me with the pilot study and in particular with the preparation of the instrumented vehicle, and the processing of the produced data.

Special thanks to IIHS and especially to Dr David Kidd who believed in my ideas and made possible the access to SHRP2 NDS data by providing me with a sponsorship. I also thank Virginia Tech Transportation Institute (VTTI) for their technical support concerning the access and initial comprehension of SHRP2 NDS data.

I am very thankful to my fellow researchers and colleagues in Design School and in school of Architecture, Building and Civil engineering for their feedback and their assistance in several aspects of this work. Moreover, I would like to thank Professor Rob Thomson and Dr Pinar Boyraz Baykas for their valuable comments on my project during my one-week researcher visit in SAFER, Chalmers University of Technology, Gothenburg, Sweden.

I am also very grateful to all my friends for their continuous support and understanding during these challenging years. Special thanks to my beloved friends Anna-Maria, Anna, Alkistis, Stella, Sasa, Nathan who were always there to listen my concerns and made this journey much more enjoyable.

Last but not least, my greatest Thank you is addressed to my family and especially to my grandfather and my mother who taught me the value of constantly learning, always supported my dreams and believed in me and my perseverance. This thesis is dedicated to them.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

The perception of society with respect to transport systems has significantly changed in the recent years. Vehicles were supposed to be a mean of convenience and social status that offers industry a large degree of freedom. Nevertheless, nowadays they constitute a rising concern regarding the high number of crashes, the environmental constraints and the high fuel costs involved. Government, industry and society in general, in order to address the above issues, are directed towards to what is known as sustainable means of transport (Eskandarian, 2012).

More specifically, as human error contributes to 90% of the recorded collisions (Treat et al., 1979, Staubach, 2009), industry moves towards the reduction of human involvement in the task of driving by introducing automated systems and, furthermore, several companies have commenced the process of developing semi-autonomous or fully autonomous vehicles. The development and widespread use of self-driving vehicles could cause a significant decrease in crashes; it could therefore save many lives (Hevelke and Nida-Rumelin, 2014) and simultaneously improve congestion problems, and reduce emissions and fuel consumption. However, in addition to the introduction of the promising self-driving vehicles, there are currently many different ways of supporting the driver in the task of driving without being involved in a crash.

Advanced driver assistance systems (ADAS) have been developed to improve comfort and safety and intervene before the crash occurs. These technologies are designed to monitor and understand vehicle or driver performance and provide assistance by i) taking over particular driving tasks (Richardson et al., 1997; Young et al., 2011) ii) providing warnings of imminent critical situations, and iii) mitigating the consequences of inadequate driver behaviour (Markkula, 2015). Technologies such as adaptive cruise control and lane-departure warning are significantly improving safety today and they are also setting the stage for autonomous vehicles.

Therefore, the development of new technologies that could be applied to all stages of vehicle automation and be able to detect not only human error, but also self-driving cars' (i.e. machines) errors, would be beneficial to road safety. Nevertheless, in order for the collision avoidance systems to be successful, the driver needs should be considered (Lee, 2004). More customisable systems, adjustable to individual driver style and characteristics can prevent drivers from deactivating them and can be substantially useful particularly for elderly or very young drivers.

1.2 Research problems and motivation

Road crashes are considered a major problem in human progress as they constitute a prevalent cause of injury which can result in disability or death. More than 1.3 million people die annually on roads in crashes rendering them as the eighth leading cause of death worldwide (WHO, 2018). 54% of these fatalities concern vulnerable road users. According to NHTSA, 58% of police-reported fatal road crashes involved only one vehicle with road departure crashes to provoke 52% of the total road fatalities. Rear-end crashes, although they constitute the most common type of road crash in United States, usually do not exhibit the same severity.

The main risk factors for road crashes are speeding, driving under the influence of alcohol or other substances, distracted driving, unsafe road infrastructure or vehicles and inadequate law enforcement of traffic laws. Despite the fact that the situation seems to be significantly improving in Europe, comparing to other regions (21% decrease from 2010), traffic crashes and their outcomes constitute an extremely serious social problem costing governments about 3% of GDP, with low-middle income countries to reach 5% (Aeron-Thomas and Jacobs, 2011).

There is a massive development of new in-vehicle systems to overcome the inherent problems associated with safety, efficiency and the economy and to provide more comfortable and environmentally acceptable solutions for future road traffic. With the motivation, to enhance road safety, the automotive industry is trying to develop more “intelligent” vehicles. Intelligent safety systems need to be able to detect hazardous situations and critical events and intervene before they result in crashes (Bishop, 2005).

Traditionally, road traffic safety analysis has relied mostly on crash statistics as the main data source. Over the years, however, numerous problems associated with crash data have been discussed. The most important aspects are that not all crashes are reported and information for the pre-crash traffic conditions, as well as, information on the behavioural aspects of road users is rarely available.

Therefore, there is a need to use some kind of surrogate measures to complement crashes, i.e., traffic safety indicators, to increase the possibility of a better understanding of the processes characterising the normal traffic and driving, as well as, critical situations including crashes (Laureshyn et.al., 2010). In this way, safety critical events, that are very useful as valid surrogates for crash analyses, can be more easily identified (Guo et al., 2010). Detecting safety critical events and investigating them, in an effort for understanding the evolution of the crash process, would hopefully lead to the reduction of crashes. The potential of systems that could detect the deviation from normal driving and stop it in the first stage of the crash development process, before it culminates in a crash, could be fundamentally beneficial to road safety.

Advancements in technology facilitate large driving data collection efforts, such as Naturalistic Driving Studies (NDS), which aid in the comprehension of driver behaviour and pre-crash conditions. Understanding driver behaviour could lead to better behaviour prediction and early detection of driver manoeuvres, therefore to the development of more efficient advanced driver assistance and warning systems.

By employing suitable indicators and setting appropriate thresholds, a collision can be avoided. In order to select these thresholds that detect unsafe driving, a better understanding of normal driving is required. A model of the crash sequence has been developed by Tingvall et al. (2009), that indicates the transition from a normal driving situation to an inevitable crash. Regarding the last stages of the crash sequence, the interventions in order to avoid the crash are well known, but there is not so much investigation about the early stages and the factors affecting the deviation from normal driving and developing crash scenarios. In other words, there is a lack of research regarding the transition from normal driving to deviation of normal driving and about the way this can lead to a near-crash or crash scenario. How could we recognise that there is such deviation and whether there is adequate time to regain normal driving conditions?

Detecting deviation from normal driving by investigating vehicle kinematics and driver behaviour that are mostly responsible for near-crash and crash events could enhance road safety by informing either in-car warning systems or systems that automatically regain normal driving. ADAS have already used emergency braking or automotive steering when the situation is crucial, but what about the simple inconsistency and latent hazards in driving that could lead to a crash? This research could contribute to the recommendations for thresholds of normal driving and more trusted, adjustable ADAS, indicators to detect deviations from it and will provide new insights into the understanding of the crash sequence development.

1.3 Research aim, objectives and research questions

The aim of this research is to investigate the transition from normal driving conditions to safety critical driving scenarios and the development of crash risks. The main objectives, hence, are:

1. To formulate a comprehensive theoretical framework of factors quantifying the transition from normal driving to a safety critical scenario and influencing crash risk.
2. To examine the potential contribution of Naturalistic Driving Studies to an understanding of vehicle kinematics
3. To characterise safety indicators during normal driving
4. To develop safety indicators for detecting deviations from normal driving
5. To formulate safety indicators during safety critical driving scenarios
6. To model the evolution of Time-to-collision (TTC) values during event sequence development
7. To recommend potential thresholds for safety indicators for designing safe and trusted ADAS

In particular, this thesis will address the following research questions:

1. How can we characterise normal driving?
2. What are the parameters that play an important role in characterising driving and how do they evolve during the stages of crash sequence?
3. How could we detect the deviation from normal driving that can lead to a crash?

1.4 Thesis structure

This thesis consists of 6 main chapters which are summarised below:

Chapter 1 introduces the research background, problems and motivation, the research aim and objectives, the research questions and outlines the structure of the thesis.

Chapter 2 conducts an in-depth literature review on driver behaviour and safety, including safe and normal driving, crash sequence development, safety critical events, safety measures and indicators and naturalistic driving studies.

Chapter 3 discusses the research methodology to be undertaken for the thesis. The chapter begins with the choice of the research approach and continues with the choice of the data and the steps required towards obtaining it. This is followed by the description of an empirical methodology to quantify normal driving and then, by a methodology to create indicators to detect deviations from normal driving. The algorithms for data of interest extraction are presented and a descriptive analysis on driver pre-event behaviour is outlined. Finally, the multilevel mixed effects model is described, which will be applied to explore Time to Collision and crash sequence.

Chapter 4 includes the preliminary analysis of the data, describing also the extensive data manipulation and pre-processing.

Chapter 5 shows the results of the statistical modelling, the analysis for pre-event driver behaviour and also, the results of the methodologies for quantifying normal driving and creating indicators to detect abnormalities in driving.

Chapter 6 constitutes of the discussion regarding the issues raised by this research, the contribution to knowledge and the limitations and finally, brings the thesis to an end with the overall conclusions and recommendations for future work.

The flowchart in Figure 1.1 below outlines the content of the thesis.

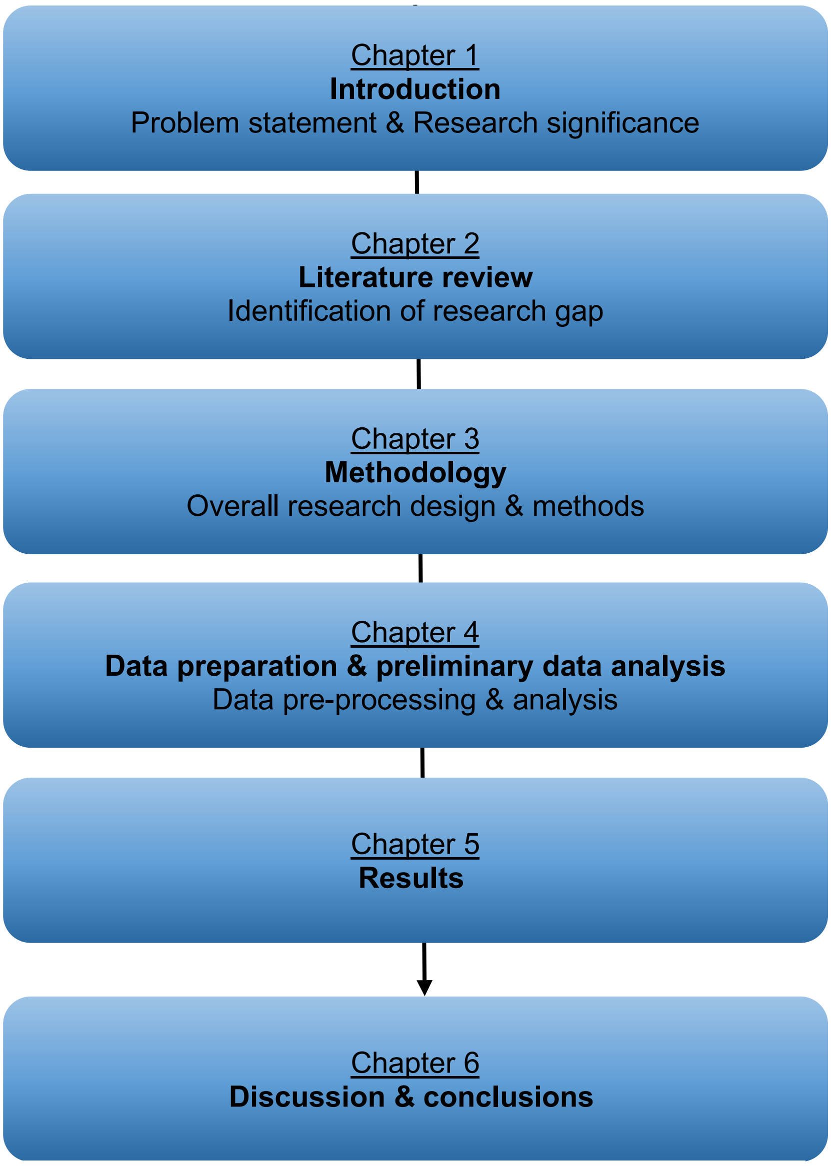


Figure 1.1 Thesis content flowchart

2 Literature review

2.1 Introduction

This literature review is focused on road safety. Initially, it introduces the theoretical background consisting of approaches to improving road safety as over the years road safety research has moved from a reactive to a more proactive approach. In this section driver behaviour models are reviewed as they play a crucial role in the success of Advanced Driver Assistance System (ADAS). Since they are employed to predict driver state, manoeuvres and intention, it would be useful to look into their main principles and their evolution. Next, the integrated safety chain model is described as a significant system approach to reduce injury and crash risk by introducing the time element and dividing the crash sequence in stages in order to be able to correct any deviations timely or mitigate the consequences of a crash.

The second section of the literature concerns road safety research metrics that they have been utilised to assess driver behaviour and detect hazardous situations during the driving process. To acquire a better understanding of the behavioural and proximal indicators used in road safety, the comprehension of the concept of normal and safe driving is necessary. Thus, the section starts with what is considered safe driving, the quality of driving dimensions, normal driving and then continues with the safety critical events definition and the other indicators.

The third and last part of this chapter focuses on the tools – methods- that have been utilised in road safety research to investigate crash development and driver behaviour, including simulators and in-depth crash investigation and elaborating on Naturalistic driving studies.

In order for an appropriate methodology to be developed for this research, it is essential to understand how previous research has investigated driver behavior and safety. The review of the relevant theory, metrics and tools will aid to the comprehension of current knowledge with the theoretical and methodological contributions discussed and will set the scene for the identification of the research gap. The literature review can be summarised in Table 2.1.

Table 2.1 Literature review plan

Theory	**2.1 Introduction** **2.2 Approaches to Road safety research** 2.2.1 Driver behaviour models 2.2.2 Integrated safety chain model
Metrics	**2.3 Behavioural and proximal Indicators of road safety** 2.3.1 Driving scenarios 2.3.2 Safety critical events 2.3.3 Other parameters, safety measures and indicators
Tools	**2.4 Methods of investigating driver behaviour and crash development in road safety** 2.4.1 Simulators 2.4.2 In-depth crash investigation 2.4.3 NDS

2.2 Approaches to road safety research

Over the years, road safety research has focused on analysing data to understand crash occurrence, to determine influential risk factors and to assess crash severity. Based on this understanding, significant conclusions can be drawn to prepare more effective and efficient crash prevention policies (data-driven approach) (Wegman et al., 2017).

Historically, road safety research has relied on a reactive approach where actions were taken after crashes have occurred, investigating peaks in distributions and diagrams while recently road safety has moved towards a more pro-active approach where there is no need to wait for the crashes to happen before acting. The knowledge acquired before can be used and adjusted to different conditions. Furthermore, in the past, the road user approach was adopted where the focus was on human error as the exclusive cause of crashes and therefore the road users were entirely responsible for crash occurrence (WHO, 2004). Nowadays, a systems thinking approach is followed that steers away from the traditional “human error” blame focussed approach and considers all the components of a system, i.e. road users, vehicles, roads, that contribute to crash risk (Filtness et al, 2016). The Safe system theory incorporates the principle that zero fatalities should be tolerated from road crashes and aims to implementation of practices and tools to lead the design, operation and use of a safe road system (OECD/ITF, 2016). This proactive approach entails risk comprehension and assessment in a road network for priority interventions to be identified.

Well-known interventions refer to improving human behaviour (speed, alcohol, seat belts, and helmets) through legislation, enforcement, and campaigns; safer infrastructure through planning and design; and safer vehicles through better crashworthiness, active vehicle safety, and vehicle inspections (Wegman, 2017).

2.2.1 Driver behaviour as predominant factor in road safety - A review of driver behaviour models

Driver behaviour models aimed at the comprehension of the human as driver and they have firstly developed to explore driver intentions, state and manoeuvres, vehicle condition and environmental factors in order to benefit road safety and driving experience (AbuAli and Abou-zeid, 2016). More specifically, these models can substantially contribute to the development and the progress of ADAS by providing for instance, information of the optimum timing for warnings to avoid imminent danger, benefit the validation of simulation models, and facilitate the development of autonomous vehicles in providing more naturalistic lateral and longitudinal control behaviour (e.g. obstacle avoidance control strategies).

Driver behaviour has an effect in the behaviour of the vehicle (Macadam, 2003; Rashevsky, 1966). In other words, vehicle responses depict the behaviour of the driver and vehicle kinematics could be considered measurements of driver behaviour. Most of the active safety technologies utilise these measurements to assess the driving state and provide warnings or interventions aiming to advance driver behaviour or alleviate the consequences of deficient driver behaviour in case of an imminent risk.

Driving incorporates several behavioural aspects and activities, including driver inattention, distractions or impairments (human factors) or vehicle control behaviour e.g. braking or steering. This thesis will focus on drivers longitudinal and lateral control of the vehicle during the whole crash sequence development. Therefore, it would be *useful to examine how driver behaviour modelling evolved, how the dynamics between the driver, the vehicle and the environment have previously described and what is the knowledge gained that can be exploited in future studies.*

Engineering factors regarding improvements in roadway and automotive trends are promising but it should be taken into account that sometimes instead of enhancing safety, they undermine it. This is based on the Wilde's theory of risk homeostasis which claims that "drivers have a target level of risk per unit time, so that physical changes to the traffic system stimulate user reactions that reset safety to its prior level" (Evans, 2004; Wilde, 1982). For example, if a driver keeps in mind that he has a very effective braking system, he may not choose to brake early relying on the system or when the pavement is improved, he chooses higher speeds. Therefore, the driver will adjust the perceived risk in the new circumstances and the improvements in vehicle or roadway will not have a lasting effect in safety. Factors that influence traffic safety are engineering and road user. Figure 2.1 is a "non-qualitative judgmental estimate" where the areas represent the significance attached to various factors according to Evans, 2004.

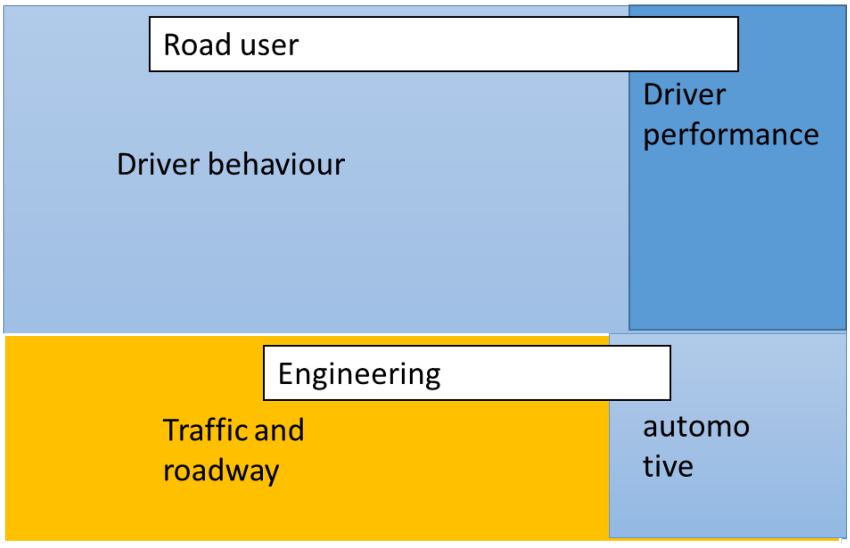


Figure 2.1 Non-quantitative judgmental estimates of relative significance of various factors (adapted from Evans, 2004)

The road user or human factors seem to play the largest role and especially driver behaviour. Driver behaviour is observed as the most unforeseeable factor in the driver, roadway and vehicle system as drivers may change their behaviour according to various physical or mental circumstances and distractions caused by other passengers or in-vehicle technologies (Zhou et al., 2008; Suzdaleva and Nagy, 2018; Eboli et al., 2017a; Eboli et al., 2017b). Changes in driver behaviour could lead to substantial changes in traffic safety. Driver behaviour is influenced by family, personality, beliefs, socio-economic status, etc. According to Evans, 2004, four are the areas of influence: fear of adverse consequences, social norms, mass media and legislative interventions.

It is essential though to separate driver performance from driver behaviour. The former represents what the driver can do while the latter what the driver does do. In other words, driver performance reflects the skills and capabilities of the driver to maintain control of the vehicle but his choices in handling situations and his accepted risks constitute the driver behaviour (Lee, 2006).

According to Rasmussen, 1987, three levels of human performance emerge: the skill-based, the rule-based and the knowledge-based performance. The lowest level, the skill-based behaviour refers to patterns of behaviour with no conscious control. The next level, the rule-based behaviour represents “a sequence of subroutines” or procedure that has been empirically acquired previously while the knowledge-based level involves problem solving during unfamiliar situations where no existing rules are applicable. In 1991, Lehto recommended a fourth level, this of judgement based behaviour that reflects the significance of value judgements and emotional reactions in ruling behaviour.

Relative to cognitive control of driving, a hierarchy of 3 levels has been recommended consisted of the strategic, tactical or manoeuvring and the operational or vehicle control level (Michon 1985; Molen and Botticher, 1987). The strategic level refers to the general stage of trip planning including setting goals for the trip, choice of route and mode and assessment of risks and costs involved. The decision making in this level is memory-driven and not restricted by real time (Ranney, 1994), (Norman and Bobrow, 1975). The tactical level involves manoeuvre control for negotiating driving situations such as turning, overtaking, gap acceptance and obstacle avoidance. Finally, the operational level includes automatic action patterns e.g. braking, shifting. Manoeuvring and vehicle control decisions are mainly data-driven as they refer to the immediate driving environment (Norman and Bobrow, 1975). Moreover, tactical level decisions require seconds to take place while control decisions demand only milliseconds (Ranney, 1994). This hierarchy (hierarchical control model) supports the dynamic relationship among the parallel activities at the three levels. The table below presents the relationship of control hierarchy of driving with Rasmussen’s taxonomy (Hale et al., 1990; Molen and Botticher 1987).

Table 2.2: Classification of selected driving tasks by Michon's control hierarchy and Rasmussen's skill-rule-knowledge framework (adapted from Hale et al. 1990, Figure 1, p. 1383).

	Strategic	Tactical/ Manoeuvring	Operational/ Control
Knowledge	Navigating in unfamiliar area	Controlling skid	Novice on first lesson
Rule	Choice between familiar routes	Passing other vehicles	Driving unfamiliar vehicle
Skill	Route used for daily commute	Negotiating familiar intersection	Vehicle handling on curves

2.2.1.1 Types of driver behaviour models

Driver behaviour constitutes the prevalent cause of crashes, contributing to over 90% of them (Lee, 2005). Over the years, researches have tried to conceptualise driver behaviour developing various models to enhance road safety by predicting driver intentions, vehicle and driver state, driving manoeuvres and environmental factors (Abuali and Abou-zeid, 2016). These models could be separated into two large categories according to Michon, 1985; those that are input-output models and those that are motivation oriented. Then, they are distinguished between taxonomic and functional models that are regarded as systems whose parts respectively do and do not dynamically interact (Michon, 1985). Taxonomic models (task analyses and trait models) express no dynamic relations between components in contrast with functional ones that offer greater potential for understanding the complex task of driving (Ranney, 1994). Table 2.3 presents a summary of driver behaviour model types according to Michon, 1985 who defines four basic categories: (1) task analyses, (2) trait models, (3) mechanistic/adaptive control models and (4) motivational/cognitive models.

Motivational models assume that driving is self-paced and that drivers choose the amount of risk they are willing to accept (Ranney, 1994). In Vaa, 2007, an overview of motivational models has been conducted chronologically, with the starting point to be the field of safe travel of Gibson and Crooks (1938) and the end the model of task difficulty of Fuller, 2005. As criticised in Winter and Happee, (2012), “motivational models tend to be unspecific, suffering from a vagueness which makes them unfalsifiable”.

Risk compensation models (Wilde 1982), risk threshold models (Naatanen and Summala 1976) and risk avoidance models (Fuller 1984) constitute examples of motivational models. One of the most known motivational risk compensation model is Risk Homeostasis Theory (RHT; Wilde 1988) which is based on the assumption that the level of accepted subjective risk is a relatively stable personal parameter. Therefore, changes to the road, vehicle or driving competency will not necessarily improve traffic safety as almost any improvement will be compensated by a less cautious or faster driver (Michon,1985).

RHT has introduced the behavioural adaptation and the driver incentives (Trimpop, 1996). Nevertheless, it has aroused extensive debates (Wilde et al. 2002; Cacciabue, 2007) and criticism (e.g. McKenna, 1982; Huguenin, 1982; Hoyos, 1984; Wilde, 1984; Wilde and Kunkel, 1984; O’Neill and Williams 2004). One of the main issues is that the theory is not testable, and it has been accused to be circular as it cannot be falsified. Moreover, it has been stated that the target level of risk should not be greater than zero as most drivers will not experience any crash injury in their lifetime (Vaa, 2007). Nevertheless, RHT has been acknowledged for one of its key components, risk compensation, and still has a serious role in driver behaviour and traffic safety measures discussions.

Risk-threshold models recommend that drivers attempt to maintain a stable balance between subjective, perceived risk and objective risk, therefore they introduce the existence of a control process. A typical model of this type is the motivational model of Naatanen and Summala (1976), later renamed the zero-risk model (Summala, 1985; 1988). According to their model, the perceived risk (R) in traffic is the product of the level of subjective probability of a hazardous event and the subjective importance of the consequences of the event.

Behaviour is assumed to be directly related to the level of R. In most circumstances, R is perceived to equal zero, meaning that drivers generally feel and act as if there is no real risk at all. Risk-compensation mechanisms are activated if a threshold is exceeded in order for the risk level to be reduced. In other words, this type of model differs from the previous in the existence of a threshold and the operation of safety margins (Summala, 1988). While for risk compensation models, driver performance is continuously adapted, for risk-threshold models the compensation initiates in case the perceived risk exceeds a threshold. Safety margins, defined in terms of the temporal or spatial distance between a hazard and the driver's vehicle, are suggested as alternatives to the stable-risk parameter proposed by Wilde (Summala, 1988).

Table 2.3: Summary of driver behaviour model types (adapted from Michon, 1985)

	Taxonomic	Functional
Input-Output (Behavioural oriented)	**Task analyses** McKnight & Adams (1970a, 1970b) McKnight & Hundt, 1971) Perchonok (1972) Fell (1976)	**Mechanistic models** Greenberg (1959), Edie & Foot (1960) Herman *et at.,* 1959 Alberti & Belli, 1978 **Adaptive control models** <u>Servo-control</u> e.g. McRuer & Weir, 1969; Weir & McRuer, 1968) Mc-Ruer *et al.,* 1977 Preyss, 1968; Wierwille & Gagné, 1966; Young, 1969 <u>Information flow control</u> Kidd & Laughery (1964) Wolf & Barrett (l978a, 1978b)
Internal state (Psychological)	**Trait models** (Conger et al., 1959), (Fleishman 1967, 1975) (Shaw & Sichel, 1971)	**Motivational models** **Cognitive (process) models** <u>Compensation Models</u> Taylor, 1964 Wilde's Risk Homeostasis Theory (Wilde, 1978; 1982; Wilde & Murdoch, 1982) <u>Risk threshold theory</u> Klebelsberg (1971; 1977) Näätänen & Summala (1974; 1976) <u>The threat avoidance model</u> Fuller (1984)

The conceptual basis for the risk-avoidance model (Fuller 1984) is the conflict between the two prevalent driver motivations: moving towards a destination and avoiding hazards. As it is not possible for drivers to move in a straight line unobstructed, they need to repeatedly avoid obstacles and potential hazards along the way. Repeated exposure to obstacles is the basis for learning how to detect risks on the road.

Although motivational models presented great potential to assist in comprehension of the complex task of driving, they have been criticized for not allowing validation due to the lack of specificity regarding their internal mechanisms (Michon 1985; Van der Molen and Botticher 1988).

To fill the gap of motivational models, hierarchical models have been introduced. Ranney (1994) characteristically mentioned them as the 'second-generation motivational models' (p. 747). As mentioned in the previous section, there are three (Rasmussen, 1983; Michon, 1985; Van der Molen and Botticher, 1988) or four levels (Hatakka et al. 2002; Hollnagel et al. 2003; Panou et al. 2007) of driving performance and researchers have also suggested the various hierarchies and taxonomies combination into a two-dimensional matrix (Hale et al. 1990) or a three-dimensional cube (Summala, 1996; Theeuwes, 2001). Although these hierarchies work well as qualitative concepts, there is no quantitative support in the literature.

In contrast to motivational models, adaptive control models provide precise quantitative results. Nevertheless, some behavioural researchers have not been supportive towards the engineering-oriented character of adaptive control models (Boer, 1999; Hancock, 1999; Ranney, 1999; Van Winsum, 1999). According to Michon (1985), adaptive control models include classic control models (involving signals that are continuous in time) and information-flow control models (involving discrete decisions). Adaptive control models have been used for driver assessment through parameter identification techniques (Boer et al., 2005) and they are provided by Bekey et al. (1977), Brackstone and McDonald (1999), Guo and Guan (1993), Jagacinski and Flach (2003), MacAdam (2003), Plöchl and Edelmann (2007) and Reid (1983). Adaptive control models involve microscopic engineering models (Brackstone and McDonald 1999) and manual control models in the frequency or time domain (McRuer and Jex 1967; McRuer et al. 1977) but also, complex computational simulations of traffic situations (Cacciabue, 2007).

The lack of ability to make good predictions despite being able to accurately fit the measured data is typical for this type of models. Nevertheless, Sheridan (2004) supported that using a simple adaptive control model can lead to a cognitive comprehension of driver distraction and advance modelling and driver performance prediction based on motorway and vehicle design. Successful applications of adaptive control models have been also for the simplest of tasks, such as curve negotiation, car-following, and regulation against wind gusts (Allen et al. 2005; Boer et al. 2005).

According to Michon (1985), a model that represents a sufficient range of realistic driving behaviours will be naturally complex and will include at least between 5000 and 10,000 elements. As driving involves plentiful design factors and random influences, this appears to be a pointless or impossible endeavour (Carsten, 2007).

In a more recent categorization of driver behaviour models, Markkula, (2015), mentioned three model types: conceptual, statistical, and process models. Conceptual models are those that depict the way drivers interact with the world (e.g., hierarchical models (Michon, 1985) and information processing (Wickens et al., 2016) and Statistical models reflect driver behaviour in statistical terms—for instance, as a reaction-time distribution (Green, 2000). Finally, process models of driver behaviour constitute mathematical descriptions of the driver in the world (e.g., how they act on information under specific circumstances) suitable for computer simulations. Usually, process models produce an output e.g., an action such as braking, utilising discrete (moment-by-moment) current or historical data. Conceptual models, given mathematical descriptions of their components, can be implemented as process models and furthermore, statistical models can be components of process models (Markkula, 2015).

2.2.2 The integrated safety chain model

Although it is known that many factors are responsible for crash risk, their exact role in the crash development is most of the times undetermined. Therefore, steps towards the comprehension of crash development would be beneficial in order to identify the causative chain of crash events and, eventually, define appropriate countermeasures effective in mitigating crash risk by limiting the influence of these factors (Wu and Thor, 2015). Regarding traffic safety, researchers have tried to investigate the correlation between factors associated with the development of a crash sequence and, finally, implementing this awareness for the methods aiming to avert and control crashes at various points within the crash development (Chapman, 1954; Evans, 2004).

Partitioning the crash sequence is important not only because intricate crash causation relationships can be untangled, but also because effective prevention strategies can be suggested. As the Tri-Level Study of Accident Causes identified, for the majority of traffic crashes more than one factor was responsible (Treat et al., 1979). This recommends that if anyone of the factors was absent, the crash would have been prevented or the severity of the crash would have been affected (Shinar, 2007; Treat et al., 1979). This definition is in line with epidemiological approaches of causation (Rothman, 2012).

Tingvall et al., (2009) introduced the time element and documented the so-called integrated safety chain where the system is designed backwards from a possible event. The challenge in taking a potential crash event as a starting point and working backwards to avoid such an outcome is to prevent the hazardous situation in the earliest possible stage of the chain. However, this is not always possible, and actions should be taken then, to reduce the kinetic energy in case of crash and therefore injury risk (OECD/ITF, 2016).

The chain includes four stages (after normal driving) starting from the deviation from normal driving, continuing with the emerging situation, critical situation and concluding to the stage where crash event is unavoidable. More specifically, the first phase of integrated safety is normal driving. Although there is not a definition for normal driving, it is usually considered to be a common way of driving that may though include risky behaviour (Sunnevang, 2016). It is influenced by the drivers' education, cognition, motivation, economic incentives and finally enforcement. When drivers deviate, they pass to the next stage, deviation from normal driving. This usually happens due to inattention, unawareness, stress, fatigue or a violation such as speeding, and the situation can be reverted by infrastructure interventions (speed bump) or a driver warning system. In some occasions, the situation can escalate to vehicle drifting or too small headways (emerging situation). Similarly, a warning or other intervention in driving could aid the driver and the vehicle to return at the normal driving stage, but a failure here could lead to a critical situation where the driver loses vehicle control or engages in an error, rendering a crash inevitable. In this stage, vehicle crash protection systems should keep impact below injury levels. Therefore, in each stage, an action or a reaction could have taken place and the crash could have been avoided or the consequences could have been mitigated (Lie, 2012). In other words, in every step, a systems approach should be employed to comprehend and integrate driver behaviour, vehicle systems, road and traffic environment, speed control and post-crash reaction with the purpose to incorporate them in an effective way.

The integrated safety chain that has been introduced by Tingvall et al. (2009), was further developed by Lie (2012) and Strandroth et al. (2012) that used it as a method to show a combined effect of coexisting enhancements of vehicle and road safety technologies. Figure 2.2 shows a visual representation of the crash development stages.

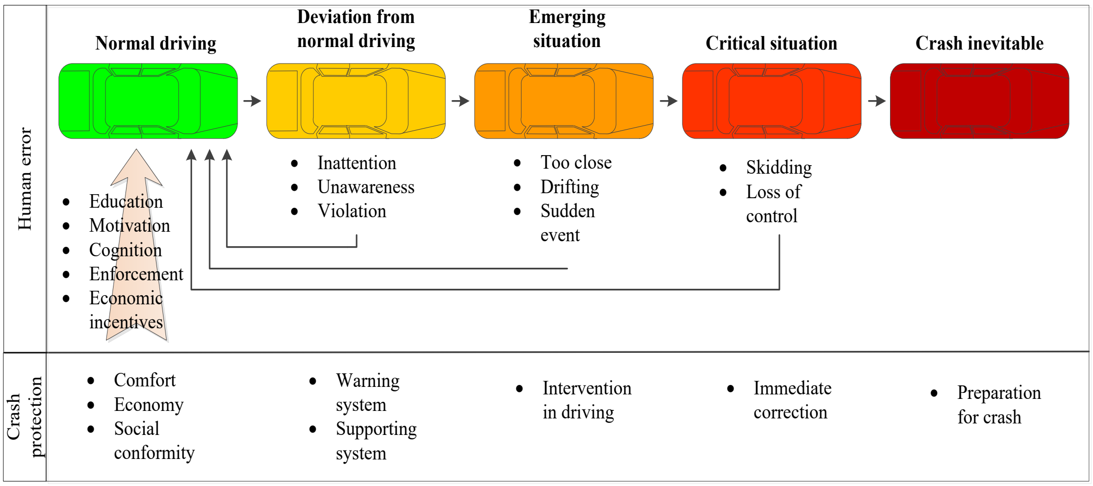


Figure 2.2 Crash sequence (adapted by Tingvall et al., 2009)

A significant approach for mitigation of crashes is for systems to correct deviations as early in the process as possible (Eugensson, 2011). Even if a crash protection system is effective, the risk of injury is reduced by every crash avoided. According to Lie, 2012, systems that are active closer to the crash, in order to be beneficial, need to act autonomously while technologies acting close to normal driving can “use the driver as an important agent”. Hence, the integrated safety chain model can potentially clarify the driver role regarding the safety technologies and the different stages before a crash.

The deficiency of this model is that does not provide itself the safe levels of performance. The conditions around normal driving affect the absolute levels of safety. Defining normal driving in relation to speed, driver capacity and potential crash configurations is crucial for the design of safety technologies (Lie, 2012).

Subdividing the crash sequence after the event seems to be the key for severity reduction and crash prevention strategies. Nevertheless, understanding of the connection between the crash contributing factors at different stages is restricted by the lack of complete and accurate pre-crash information.

2.3 Measuring driver behaviour - Behavioural and proximal indicators of road safety

In this second subsection of the literature the metrics that are commonly used to determine compromises in driver safety are reviewed. Traffic conflict technique holds a significant role between other behavioural and proximal indicators of road *safety that are cited here. According to TCT, traffic can be depicted as a pyramid where there exists a transition from normal vehicle interactions to safety relevant events that can conclude to fatal crashes (Hyden, 1987). The latter are rare events and are situated in the peak of the pyramid. This continuum of traffic events imposes a connection between the number of serious conflicts and collisions (Yang, 2012). This thesis explores the transition from a normal driving situation to safety critical events, therefore it would be fruitful to review literature about all phases of traffic, including safe driving and quality of driving dimensions, normal, uneventful driving and, safety critical events or conflicts. Furthermore, safety critical events are a proximal indicator of road safety that will be explored in the framework of this research; hence it is important to understand what safe and normal driving entails, in order to distinguish and detect risks in driving process.*

Driving can entail a number of microscopic events, not only crashes. The interaction between the road users can be seen as a continuum of safety related events that constitute different levels of a pyramid (Hyden, 1987). "Undisturbed passages" are situated at the bottom while crashes are found at the very top rendering the pyramid as a severity scale (Figure 2.3).

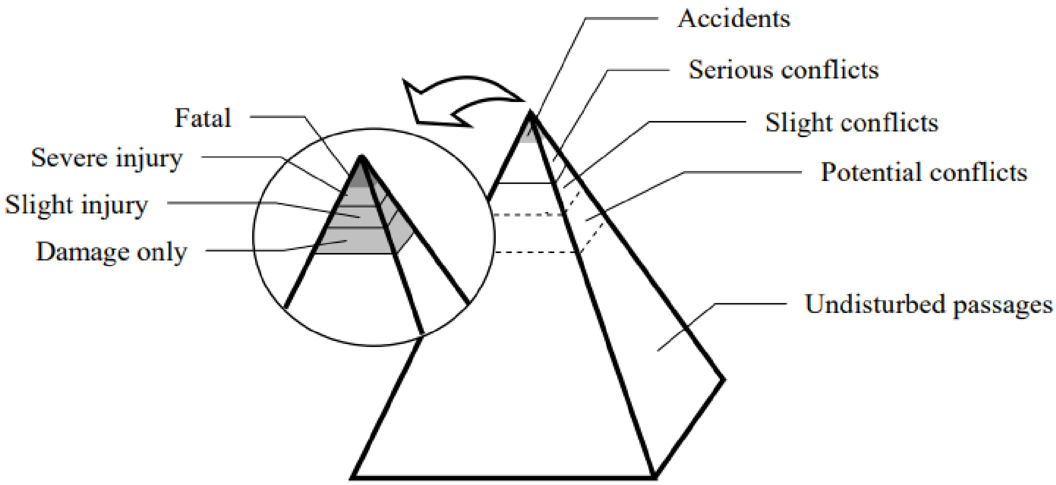


Figure 2.3 "Safety pyramid", Hyden (1987)

One dimension of severity is the proximity to a crash and the other refers to the potential consequences in case of a crash had occurred. The Swedish Traffic Conflict Technique assumes that near-crashes or traffic conflicts share the same underlying causes and therefore, they can act as a tool to evaluate and reduce hazardous situations. By observing near-crashes that take place in a greater frequency than crashes, one can make improved predictions about the latter with significantly less time and effort. However, according to Shinar (1984) the usefulness of TCT as a surrogate measure has been validated through studies that provide mixed results. For instance, Williams (1981) concluded in his review that TCT is not as useful as a predictor of crashes as the relationship between the conflicts and the crashes is not established and the way “conflicts reflect hazard in the road system” has not been demonstrated. On the other hand, there are researchers that argue the validity of crash data and claim that some safety indicators can better describe unsafe situations and predict crashes ((Migletz et al., 1985; Svensson, 1992; Archer, 2005). Miscellaneous measures and indicators have been defined and used in previous studies and they are associated with different aspects of driving.

2.3.1 Driving scenarios

Fitts and Posner (1967) make frequent reference to driving as an example of complex feedback dependent skilled behaviour. This is the way that they express the requirements of driving: ‘‘In driving an automobile, for example, one does not randomly respond to stimuli on the road. Instead, one makes responses in accordance with some internal model which involves reaching a destination at a certain time while obeying various traffic regulations, accommodating oneself to the other traffic on the road, and adapting one’s driving in numerous other ways to the immediate environmental situation’’. While being one of the most responsible human activities, driving is more likely to cause death compared to any other typical citizen activity (Evans, 2004). Gibson and Crooks, (1938), noted that ‘‘of all the skills demanded by contemporary civilization, the one of driving an automobile is certainly the most important to the individual, in the sense at least that a defect in it is the greatest threat to his life’’.

Several studies have prevalent aim to recognise safety critical or hazardous situations in driving. However, a key point, to better understand the complex mechanism of crashes and the safety critical scenarios theoretic base, would be to clarify what safe driving is.

2.3.1.1 Safe driving

Gibson and Crooks, (1938) described driving in terms of a field of safe travel where drivers in order to move themselves towards their destination, they modify their speed and direction to avoid hazards. The field of safe travel is defined by the roadway and other vehicles and includes the possible unobstructed paths the vehicle may take. A step further, Emberger (1993), defined safe driving as “optimal control of the vehicle with respect to its environment and its driver’s capabilities”. Tardy perception of obstacles, misinterpretation of safety boundaries in driving, inappropriate behaviour or reaction in unpredictable traffic situations- the cause of many crashes-are normally categorised as human error, such as unawareness or unwatchfulness (Parkes and Franzen, 1993).

Olson et al. (2009) argue that, the demand of driver task and the driver state (including impairment and distraction) play a critical role in the driver safety system. Oftentimes, crashes happen when an unforeseeable event occurs, demanding a novel (non-automatic) response while the driver is in lack of vigilance. One assumption is that the attention is needed for safe driving is corresponding to the driving demand or unpredictability of the environment.

In more general terms, Zaidel and Noy (1997) in the book “Ergonomics and safety of Intelligent driver interfaces” mention the term Quality of driving and they give definitions of the eight dimensions that are supposed to determine quality of driving as an indicator of safe driving. Characteristically, it is argued that “Quality of driving is a multidimensional construct, comprising performance variables as well as higher order strategic behaviours”. Below, there are the definitions of the quality of driving dimensions. Many of them have also been mentioned in other studies (e.g., Quimby, 1998; Risser 1993).

2.3.1.2 Quality of driving dimensions

The first three dimensions relate to the driver’s position control of the vehicle in space and time (Zaidel and Noy, 1997).

1.Speed maintenance. The degree to which speed is maintained within safe margins, is adjusting to traffic circumstances and differentiates in a timely, rational and gentle mode.

2.Headway maintenance. The degree in that the longitudinal distance of the vehicle from other road users is within safe margins, is adaptive and permits good visibility.

3.Lane position. The degree to which the lateral position of the vehicle is within a lane and that the choice of lane is appropriate on a consistent basis, with no uncontrolled drifting and without encroaching on the path of other road users.

The next three (Dimensions 4-6) are regarding to interactions between the driver and the actual or possible presence of other road users, where conflicts and priorities have to be assumed and resolved.

4.Turning and crossing. The degree to which drivers make the appropriate visual checks, adjust spatial placement and speed before and during turning or crossing and keep a safety bound in gap acceptance and other conflict points resolutions.

5.Traffic control devices. The degree to which drivers properly understand the meaning of markings, signs, signals, and other control devices; comprehend the intentions and positions of other road users; and follow a course corresponding to the opportunities afforded by the controls.

6.Vehicle handling. The degree to which the driver uses the vehicle's controls in a proper way, with little overt attention, and accomplishes a consistently gentle ride.

The last two dimensions (7-8) are associated with the drivers' management of the vehicle, their attention and the environment.

7.Dynamic space management. The degree to which drivers handle their time and space in a way that advances the amount of space surrounding their vehicle without drawing away from the safety of others; pick up the safer course in complex conditions and preserve manoeuvring flexibility.

8.Dynamic time management. The degree to which drivers look ahead in time and space, pay attention to secondary indications in order to predict likely future traffic participants' positions and refrain situations demanding extra manoeuvring.

In a previous study (Zaidel, 1992), a factor analysis demonstrated that QOD measures had good psychometric properties, too.

2.3.1.3 Normal driving

Driving is a complicated activity that demands eye, hand and foot coordination and simultaneously awareness of the road environment (Masuri et al., 2012). However, from a safety aspect, most of the times is uneventful. In normal driving conditions, drivers act in a mode where their effort does not extend their comfort zone boundaries (Engström, 2011). They adjust to the vehicle kinematics, other road users and the environment, so they can travel safely and in comfort (Bärgman, 2016).

There are many studies relevant to safety critical events or studies about unsafe and risky driving, but there is a lack of definition regarding normal driving. While all studies refer to criteria in order to detect safety critical events and hazardous situations during normal driving, or distracting driving, there is inadequate research about the normal driving itself, the starting point of the deviation from this, the manner that it can be measured and the way that all the variables, characterising normal driving, alter within the stages of the crash or near-crash development.

Olson et al (2009) refer to the normal driving as baseline, routine and uneventful driving. Klauer et al. (2006), at their report, calculate the risk due to distractions in safety critical events and in "normal, baseline driving", as well. Fundamental questions remain: what do we really mean with the term normal driving? How can we define and measure (in quantitative terms) normal driving and the deviation from that?

A sufficient definition of normal driving, that is written in the frame of this research, is the following: Normal driving is this combination of factors-both human and vehicle kinematics-that produce a typical mode or pattern of driving, relying on the principle of road safety (safety of the driver and people on the road) and ensuring compliance to the driving regulations. The outcome (normal driving) is supposed to enable the driver and the passengers feel safe inside the vehicle. In other words, there should be a feeling that there is adequate time for a safe reaction on any event or any activity of other road users.

2.3.2 Safety critical events

2.3.2.1 <u>A definition</u>

Even though algorithms continuously improve their ability to automatically recognise and anticipate safety critical events, today there is a lack of a common established definition of what a safety critical event exactly is. For example, the 100-Car Naturalistic Driving Study defines safety critical events as follows (Dingus et al., 2006a):

- **Crash**: situations in which there is physical contact between the subject vehicle and another vehicle, pedestrian, cyclist, animal, tree or fixed objects (e.g. pavement, electricity columns, bars, railings)
- **Near-crash**: situations requiring a rapid, severe, evasive manoeuvre to avoid a crash. In other words, a crash is about to happen, but it is avoided due to an intervention.
- **Incident**: situations requiring an evasive manoeuvre occurring at less magnitude than a near crash.

Another more integrated definition of safety critical events is: *"Situations (including crashes) that require a sudden, evasive manoeuvre to avoid a crash or to* *correct for unsafe acts performed by the driver himself/herself or by other road users"* (Bagdadi, 2012).

Furthermore, according to Swedish TCT, **Conflicts** are "undesired phenomena". "*Serious conflicts are in the same way as traffic accidents, the result of a breakdown in the interaction between the road user, environment and vehicle. A serious conflict is characterised by the fact that no one voluntarily gets involved in such a situation. The necessary evasive action is usually braking, but may also be swerving or acceleration, or a combination of these".* Various TCTs have been employed in different European and North American countries and were based on the detection and count of near-crashes occurring in traffic. In 1977 an international co-operation begun, to compare and combine TCTs, and a standardised definition of a traffic conflict was agreed: *"A conflict consists in an interaction between two road-users (or between one road-user and the road environment) that would shortly lead to a collision unless one at least of the road-users involved performed an evasive action"* (Muhlrad, 1993).

Each time that a driver performs a critical driving manoeuvre can be considered as a failure in the traffic safety system that decreases the overall traffic safety. A critical manoeuvre is identified by its abruptness caused usually by the very short reaction time available. That is an indication that the driver did not plan for the performed action. In hazardous conditions, drivers perform differently than they do normally and therefore, investigating their reaction during these moments could benefit the development of driver behaviour models (Bagdadi, 2013). Near crash scenarios can also enhance knowledge of a driver's response to a real crash situation. This kind of crash surrogate measure (or measures of crash proximity) has been proved to be beneficial in traffic safety analysis. As these scenarios follow a similar chain of events with real crashes, understanding them may lead to a deeper comprehension of crashes and the external circumstances surrounding them. Below in Figure 2.4 the crash generating process according to Wu and Jovanis (2012) is depicted.

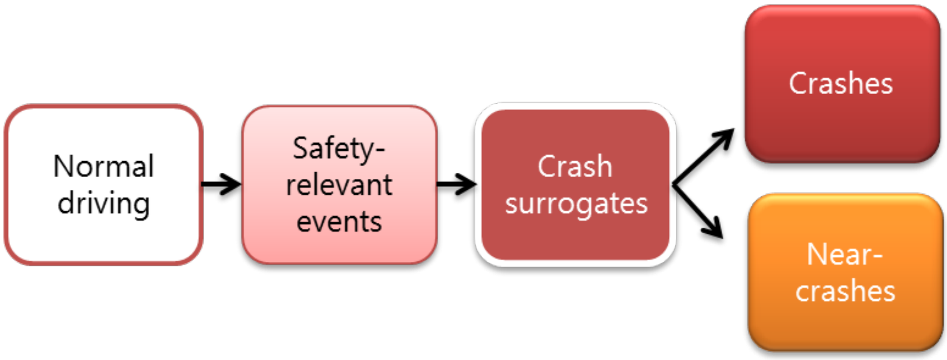


Figure 2.4 Chain of events-crash generating process (Wu and Jovanis, 2012)

2.3.2.2 Crash surrogates

In terms of the development of crash surrogates for traffic safety analysis, remarkable research has been conducted over the last 40 years (e.g. Perkins and Harris, 1967; Datta, 1979; Hauer, 1982; Hydén, 1987; Chin and Quek, 1997; Shankar et al., 2008; Tarko et al., 2009; McGehee et al., 2010; Guo et al., 2010). There is a necessity for quicker (before the number of crashes becomes larger) safety analysis and richer in data than this available from crash records. This goal can be achieved by surrogate research which gives the chance of improvement or evaluation of safety countermeasure effectiveness (Datta, 1979; Grayson and Hakkert, 1987; Archer, 2004).

In order to investigate crash occurrence and recommend proper countermeasures towards road safety, researchers have the potential to use near-crash events and specifically these with similar etiology to crashes.

In road safety analysis, especially with surrogates, the challenge is to develop valid reliable diagnostic procedures that would be useful in estimating safety issues for locations in the network or drivers in the population. The standardisation of diagnoses would play an important role so findings may be implemented across studies through the acquisition of a stable knowledge base.

According to a recent study, the use of near-crash events as crash observations, can contribute in the reduction of standard errors for the estimation of the effects of factors contributing to crash occurrence. Apparently, this happens because of the sample size extension (Guo et al., 2010). Moreover, there are additional potential advantages (e.g. Williams, 1981; Hauer, 1982, 1999; Grayson and Hakkert, 1987):

- Once the surrogate events are well-defined, the models should be able to assess the factors connected with the conditional probability of a crash and afterwards, develop countermeasures for reducing crash probability.
- Considering the difference between crash and near-crash events, a key point would be to investigate the factors that stop a near-crash from becoming a crash given that both events share similar generating processes.
- Taking into account the crash development process, it would be beneficial to examine the factors that may lead to near-crash events during normal driving. Therefore, it could be achievable to decrease the probability of near-crash event occurrence, and thus, crash occurrence.

2.3.2.3 The desirable criteria for crash surrogate

The concentration of the historical crash surrogate research was mainly on traffic conflicts. Nevertheless, in recent years, the literature approaches wider issues regarding crash surrogates including which are the suitable features of surrogates; some of the criteria suggested for surrogates are (Wu and Jovanis, 2012):

1. The surrogate ought to have a brief period of data set (Tarko and Songchitruksa, 2005) and be more repeated than crashes (Svensson, 1998). This constitutes a cardinal criterion to the earliest traffic conflict studies.

2. A surrogate should be tied in with an objectively significant outcome (Tarko et al., 2009). A surrogate is an event with traits as those of crashes (Davis and Swenson, 2006; Davis et al., 2008; Shankar et al., 2008; McGehee et al., 2010; Guo et al., 2010) and beneficial as a complement to crashes, particularly in the comprehension of crash frequency and severity (Hauer, 1999; Tarko et al., 2009). This criterion reinforces the idea that both crashes and surrogates are events characterized by several aspects such as conditions, driver behaviour, vehicle features and event attributes (Shankar et al., 2008).

3. A surrogate should have a causative and statistical relationship to crashes. Closely related to this idea is the notion that surrogates should have the characteristics of near-crashes in a hierarchical continuum; crashes are at the highest level, while passes with a minimum of interaction are at the lowest level (Svensson, 1998; Guo et al., 2010).

4. Surrogates should conceive the outcome of management in a manner similar to the manner this would influence crashes (Hauer, 1999; Shankar et al., 2008; Tarko

et al., 2009). For this criterion, the factors that lead to the surrogate would have to be similar to these that contribute to a crash.

5. Surrogate are “markers” associated with a crash, with a time scale base (e.g. the crash event is considered as a time endpoint) (Shankar et al., 2008). This endpoint perception is easily detected for crashes but may be harder for surrogates. In spite of this difficulty, this criterion constitutes a potent representation of time-dependencies within the analysis framework of the crashes and surrogates.

2.3.2.4 The multiple dimensions of crash surrogate measures

The majority of the crash surrogates presented in literature are surrogate measures with only one metric. Examples include: time-to-collision (e.g. Hydén, 1987; Chin and Quek, 1997); deceleration rate (e.g. Hydén, 1987); post-encroachment time (e.g. Hydén, 1987, 1996; Topp, 1998); deceleration-to-safety time (e.g. Topp, 1998); gap time, encroachment time, time-to-zebra (e.g. Várhelyi, 1996); proportion of stopping distance (FHWA, 2003); shock-wave frequency (e.g. VanArem and DeVos, 1997); “Jerks” (composite g-force and speed) (e.g. Gully et al., 1995); standard deviation of lateral position (e.g. Vogel, 2003); design consistency; time-line crossing (e.g. Vogel, 2003; Gordon et al., 2009); right-lane departure waning (e.g. Gordon et al., 2009); and time-to-right-edge crossing (e.g. Gordon et al., 2009). Although all the above metrics refer to vehicle kinematics, there are undoubtedly events in which no kinematic trigger is involved. Driver’s fatigue or several distractions may be responsible for such events.

An appropriate context might make these discrete metrics very useful crash surrogates. This context is the key for the desirably positive traits depicted in surrogate criteria two to five. There are few that have noticed the lucid relation between surrogate metric and context, such as Davis et al. (2008), Shankar et al. (2008) and McGehee et al. (2010).

2.3.3 Other parameters, safety measures and indicators

There are many parameters that play a determinant role in characterising driving and have an impact on driving performance. Some of the most important ones are presented below. Most of them in the literature are referred as safety indicators.

More specifically, proximal indicators express the proximity to the crash regarding the time and space. In other words, they denote closeness of the other road users or vehicles in terms of the possible point of a crash (Niezgoda et al., 2012).

Examples of such indicators are the Traffic Conflict Technique (TCT), Time to Collision (TTC), Time to Accident (TA) and Post Encroachment Time (PET). There are also other kind of indicators though, e.g. frequency of red light violations, mean speed, major lane violations that are not proximal. Nevertheless, all the aforementioned are considered behavioural indicators of road safety.

2.3.3.1 Time-to-collision (TTC)

The original definition of TTC was devised by Hayward in 1972 and it is one of the most widely-used indicators of traffic safety (Laureshyn et al., 2010). TTC is the time that remains before a collision takes part in case that both vehicles maintain the same course and speed.

For the case of rear-end collision (Minderhoud and Bovy, 2001):

$\boxed{TTC= x_1-x_2-l_2/u_2-u_1, \quad u_2>u_1}$

With x_1 denoting the position of the leading vehicle, u_1 the speed of the vehicle, x_2 the position of the following vehicle, l_2 the length of the following vehicle and u_2 the speed of the vehicle.

For the case of a head-on collision:

$\boxed{TTC=x_2-x_1 /u_2+u_1}$

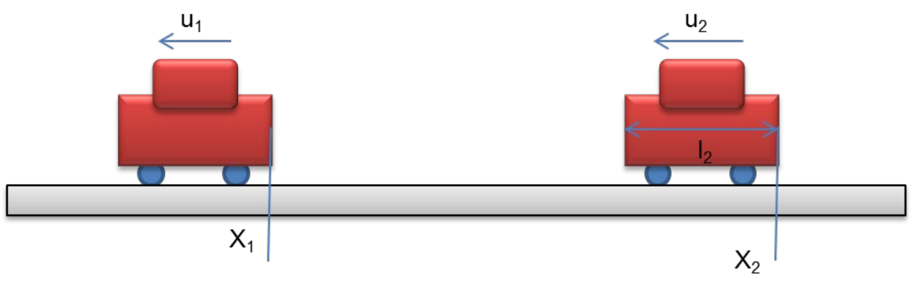


Figure 2.5 Calculation of TTC for rear-end collision type

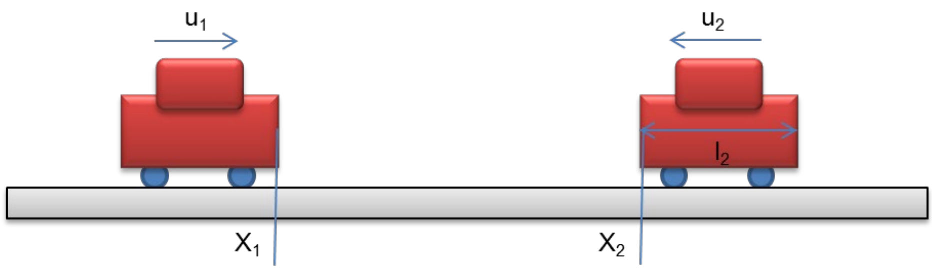


Figure 2.6 Calculation of TTC for head-on collision type

When there is a collision course the TTC-value becomes finite and declines with time, therefore the critical measurement in order to assess conflict severity is the minimum TTC during the conflict. It can be assumed that TTC occurs (value is finite) only in the case of lower speed of the lead vehicle comparing to the speed of the following vehicle. This assumption implies occurrence of collision course. TTC is a continuous variable, it can be calculated for any moment as long as the road users are on the collision course and its distributions have been applied in several studies (e.g. Fancher et.al.,1997; Van Arem and De Vos, 1997) to pinpoint traffic safety impacts (Archer, 2001). Moreover, Ward et al. (2015) calculated TTC for two dimensions extending the applicability of the safety indicator to more traffic scenarios.

Studies use recorded minimum value of TTC of relevant events, value when evasive action began or duration of the defined TTC event. There is still no general agreement between researchers regarding the critical value of TTC; nevertheless, for safety studies this value is proposed to be greater than 1.5 second and less than 5 seconds (Martens and Brouwer, 2011). Hayward (1972) suggested that 4 seconds is a critical value of TTC below which, there is a crash prone situation, Brown et al. (2001) identified a TTC threshold of 3 seconds, while Archer (2005) suggested an upper threshold of 1.5 second as indicative of hazardous situations. Moreover, according to Sayed and Zein (1999), TTC values lower than 1 second entail high crash risk, between 1 and 1.5 moderate, and between 1.6 and 2 low risk of collision.

Raw TTC measure cannot be considered as valid in all critical situations. TTC value is calculated from distance and speed and it can give the same value for 20 km/h and 160 km/h. To discern the severity of these unequal conditions TTC can be used in combination with other measures e.g. absolute value of speed at defined TTC event or braking rate (Niezgoda et al., 2012).

To enhance road safety, Lee (1976) suggested the use of TTC over the speed, distance or acceleration/deceleration for a control strategy of following vehicles. Farah et al. (2017) used TTC as a surrogate safety measure to investigate head-on collisions with the opposite vehicles during a takeover manoeuvre on two-lane rural highways. Similarly, Hegeman (2008) and Shariat-Mohaymany (2011) have also used TTC as a risk measure of overtaking behaviour.

Furthermore, TTC is supported to be a significant factor for judging the moment to start braking, and the role of TTC information in the control of braking has been investigated (Winsum and Heino, 1996). Recently, Lee et al. (2016) introduced an advanced braking algorithm using new longitudinal safety index based on TTC and warning index while Li et al. (2016) proposed a collision risk algorithm for connected environment based on real-time TTC values.

Finally, several studies examined the factors affecting TTC. Farah et al. (2009) with the employment of a Tobit regression model investigated minimum TTC and proved that not only geometric design of road, but also traffic and driver characteristics explanatory variables present a significant effect on minimum TTC. Similarly, Winsum et al. (1997), (1999) considered both kinematic and human factors characteristics for car following modelling based on TTC and the states of the following vehicles. Furthermore, with driving experiments, Hamdar et al. (2016) confirmed the real-world observations that roadway (lane width, shoulder width, median existence, median type, horizontal and vertical curves) and weather associated factors (foggy weather, icy and wet road surface conditions) influence TTC.

2.3.3.2 Post-encroachment time (PET)

The PET value, initially introduced by Allen et al. (1977), is defined as the temporal difference between the moment that the first road-user departs from a possible area of collision and the moment that the second reaches the same spot.

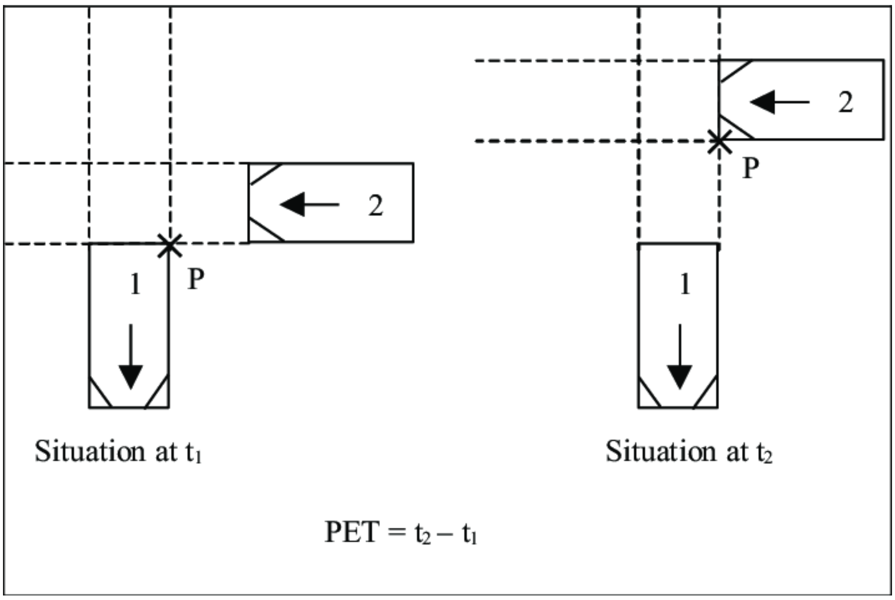


Figure 2.7 PET

The PET value indicates the magnitude to which they missed each other. In urban areas, when PET is equal to or lower than 1 second denotes a possibly critical situation (Van der Horst et al., 2014). PET is more easily acquired than TTC as it does not require collision course nor relative speed or distance data (Archer, 2005). Nonetheless, PET applicability is restricted to conflicting vehicle trajectories (angle crashes) as the collision areas should not be changing dynamically with vehicle kinematics as they do, for instance, in rear-end collisions (Cunto, 2008).

2.3.3.3 Time headway

Time headway (**H**) is the time between the front of two vehicles passing the same point (Vogel, 2003).

$$H= t_2-t_1,$$

With t_2 denoting the time at which the following vehicle passes a specific point and t_1 the time at which the leading vehicle passes the same point.

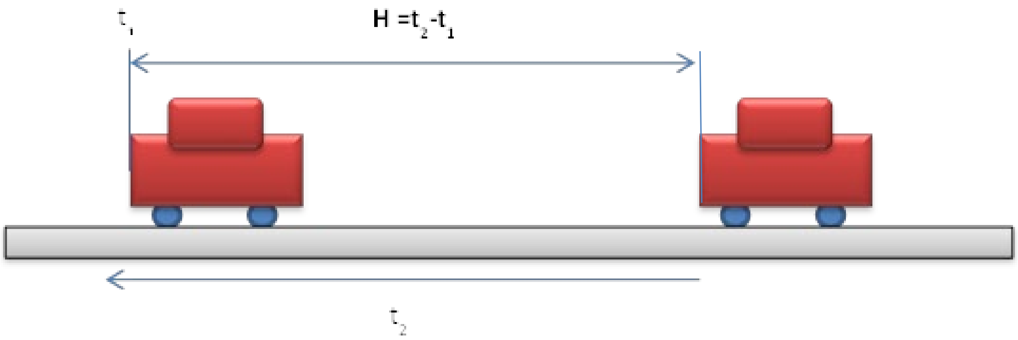


Figure 2.8 Calculation of Headway

Headway is a traffic safety indicator which is used mostly in measures and research regarding continuous flow i.e. at the highway. Apart from that, it is also a useful variable in traffic streams measures in the transportation engineering sector. Safe headway varies among countries. Reliable correlation between road safety and headway value is with regard to critical headway, which is the threshold between safe and unsafe driving. However, the value still differs in various approaches (the most often cited in literature values are between 1.0–1.5 seconds).

As far as the relation between Time Headway and Time-To-Collision concerns, high value of H implies high value of TTC, while short Time Headway does not mean short Time-to-Collision. Short TTC is not possible to sustain without provoking a collision where short H can be maintained for a long period of time. This assertion influences traffic safety measurement as lays TTC more connected with the occurrence of possible near-crash events (Niezdoga et al., 2012).

2.3.3.4 Time gap

Time Gap (**TG**) is a parameter that depicts the actual distance between road users expressed in time units. In its conventional definition it is applied to vehicles following in a flow. TG is the time between the rear-end of the leading vehicle passing a certain point and the front of the following vehicle reaching this point (Vogel, 2003).

$$TG = H - l_1/u_2$$

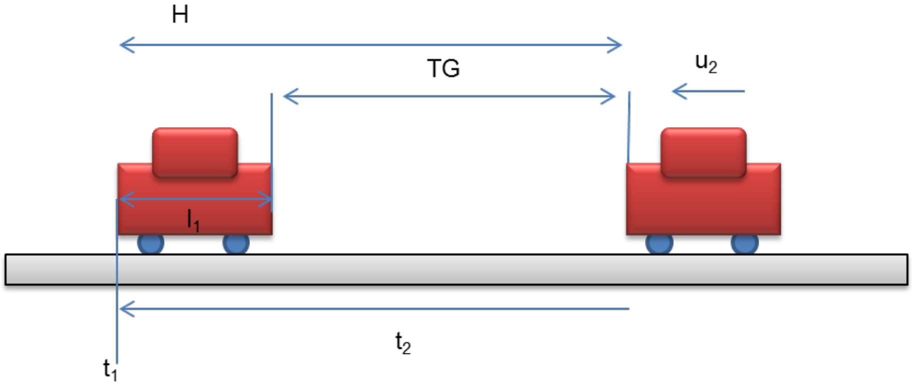


Figure 2.9 Calculation of Time Gap

As it is clear from the definition TG is a single value measured directly at a specific place. In order for it to be continuous and more closely aligned with above indicators, a “predicted” Time Gap can be used, i.e., the Time Gap that will be measured if the road users maintain their course and speed (Laureshyn et al., 2010).

Time Gap is not so strongly associated with collision risk compared to TTC, since it only takes under consideration the spatial proximity between the road users (in time units), but not their relative speeds. Nevertheless, it is helpful for detection of potential risks at earlier stages of a conflict. This is illustrated by an example of two cars following each other at the same speed and on the same course (i.e., no collision course). The size of the time interval between the vehicles (i.e., TG) affects significantly the pace of the TTC decrease in case that the first one starts braking and the vehicles suddenly are on a collision course. Therefore, TG depicts the probability of TTC quickly declining if the road users are on a collision course (Niezdoga et al., 2012).

2.3.3.5 Speed

Driving speed is a widely used parameter associated with crash risk and crash severity (Elvik et al., 2004). Road safety is affected by speed in two main ways. Firstly, the higher the speed the less the time available for the driver in order to react properly in possible risky situations on the road. Moreover, it provides less time to the other road users to react (e.g. vulnerable road users). Secondly, the speed is strongly related to the kinetic energy; increasing speed produce high kinetic energy that is mainly responsible for fatal or severe crashes. According to National Highway Traffic Safety Administration, the violation of the speed limit or fast driving regarding the road conditions (speeding) is a leading factor contributing to traffic crashes, and also a contributing factor in almost 3/10 fatal crashes.

However, the role of speed in safety critical events is controversial. Although it is often connected with crash occurrence, studies have found that, much lower speeds than the mean speed on the road, have also an increased crash rate (Solomon, 1964; Cirillo, 1968; RTI, 1970).

Several human factors and their intricate interactions affect completely the choice of speed, e.g. attitudes and risk of apprehension. Therefore, despite the fact that time-based indicators take under consideration the speed of the road users and the spatial proximity, the driver's speed choice during a passage (road user's speed profile) is significant as well, due to important behavioural information and the description of the encounter as a process (Laureshyn et al., 2009a,b).

Federal Highway Administration proposes two other indicators (conflict measures) which are the proportion of **stopping distance** and the **deceleration rate**.

2.3.3.6 Stopping distance (SD)

Stopping Distance is the total distance travelled before the actual reaction and after this, until the vehicle completely stops.

Reaction Distance + Braking Distance = Stopping Distance

2.3.3.7 Reaction distance (RD)

Reaction distance is the result of the speed and the perception-reaction time of the driver/rider (Taoka, 1989). “Total reaction time related to the use of evasive tactics in response to driving hazards thus involves the elements of initial and full awareness (perception) of specific objects in the driving visual field, required mental evaluations of such perception, the search for alternative evasive tactics, decision (or indecision) concerning appropriate reaction, and the time required for physical (bodily) reaction to operate vehicle controls followed by the time necessary for vehicle response to the controls used” (Baker and Fricke, 1986). Therefore, the total reaction time is consisted of three parts: (1) mental processing time (2) movement time and (3) device response time. In spite of the fact that “reaction time” is always a part of the “response time”, frequently the two terms are used interchangeably (Archer, 2001).

2.3.3.8 Braking distance (BR)

The braking distance is the distance needed for a vehicle that has a speed υ to come to complete stop after the brake is pushed. It is primarily affected by the original speed of the vehicle and the coefficient of friction between the tires and the road surface (Fricke, 1990).

$$d = \frac{\upsilon^2}{2\mu g} \quad \text{(m)}$$

Where μ is the coefficient of friction between the road surface and the tires, g (m/s^2) is the gravity of Earth, u (m/s) the speed of the vehicle on the moment of braking and d (m) is the distance travelled.

2.3.3.9 Deceleration rate (DR)

It is the rate at which a vehicle must decelerate to avoid collision.

2.3.3.10 Acceleration

Wahlberg introduces a measure of driving behaviour, g-force (acceleration-force), which has been predicted to be positively correlated with traffic crash frequency. Longitudinal acceleration is commonly used to measure certain traffic conditions such as crashes and near-crashes in NDS or studies where kinematic vehicle data is gathered (McLaughlin et al., 2008; Nishimoto et al., 2001; Van Winsum and Brouwer, 1997; Van Winsum and Heino, 1996; Yan et al., 2008) and it is thus considered to be a valid risk indicator.

Moreover, longitudinal acceleration has been widely used for the investigation of braking behaviour or as indicator of driving performance. More specifically, Bagdadi and Várhelyi (2013) analysed acceleration profiles and investigated the acceleration rate of change to distinguish critical and potential critical situations while af Wåhlberg (2006) investigated the driver acceleration and deceleration behaviour and revealed that combined celeration (overall mean speed change) measure constitutes a better predictor for road crashes. Furthermore, it is found that drivers with crash history indicate more abrupt braking, thus, the harsh braking responses could be an indication of crash proneness (Bagdadi and Várhelyi, 2011). Acceleration is also an important variable in risk assessment physics-based motion models in the context of intelligent vehicles, allowing prediction of how a situation will evolve in the future (Lefèvre et al., 2014).

Finally, lateral acceleration is a well-known indicator utilised for lane change detection (e.g. Farid et al., 2006) and lateral driver behaviour in general. It is mainly affected by vehicle speed and the horizontal curve radius and it is considered a reliable predictor for skidding and rollover crashes (Wang et al., 2015).

2.3.3.11 Yaw rate

A yaw rotation is a movement around the yaw axis of a rigid body that alters the direction it is heading, to the left or right of its direction of motion. The yaw rate or yaw velocity of a car, is the angular velocity of this rotation, or rate of change of the heading angle. Therefore, it is an important lateral stability indicator as it depicts the deviation of the vehicle's straight course. It is usually measured in degrees per second or radians per second.

Many studies have utilised it for control algorithms and lane change models, (e.g. Park et al, 2015) or to examine the steering behaviours in collision avoidance and lateral movements, (e.g. Wu et al, 2017) while Sudweeks (2015) with a functional yaw rate classifier reduced the number of false triggers 42% relatively to the 100-Car NDS (Dingus et al., 2006) and maintain the majority of safety relevant events (92% of crashes, 81% of near-crashes).

2.3.3.12 <u>Lane keeping parameters</u>

There are also some other traffic safety indicators concerning lateral behaviour of a driver. The most popular of the rest indicators of this kind are **mean lane position**, **standard deviation of lane position** (**SDLP**), **lane exceed** and **Time-To-Line-Crossing** (**TLC**) (Martens and Brouwer, 2011).

Lateral behaviour, also known as 'lane keeping', reflects driving performance. Nonetheless, it is strenuous to define where the limits between safety and unsafe behaviours are. As far as the previous statement concern, it is possible that the experienced drivers do not maintain constantly precise lane position but try to keep the course of the vehicle within some satisfactory range. Simultaneously, raised lane swerving and lane exceed reflects low vehicle control and thus, a higher crash risk (Martens and Brouwer, 2011).

Despite these limitations SDLP is one of the most prevalent performance metrics and describes the degree of driver's vehicular control in any driving conditions. A typical standard deviation of lane position for baseline driving is just under 0.2 m, approximately .18 m for driving on the road and approximately 0.23 m for simulators (Olson et al., 2009).

A similar measure is TLC, which is the time until the lane marking, with a constant speed and fixed steering angle. It is not deniable that TLC < 1s entails an increased safety risk. TLC points out the likelihood of a lane exceed to occur within a short time frame and thus, detects a potential hazard before the lane exceed actually takes place.

2.3.3.13 <u>Human error</u>

Human error includes unsafe acts that are an outcome of several psychological precursors and preconditions. An unsafe act is a violation, or an error occurred while there is a potential hazard and they can be divided into unintended (slips, lapses, mistakes) and intended actions (mistakes and violations) as illustrated in Figure 2.10.

Slips are attentional failures like intrusion, omission, reversal, misordering and mistiming while lapses are memory failures such as omitting planned items, placelosing, forgetting intentions. Furthermore, there are rule based mistakes including misapplication of a good rule or application of bad one, as well as, knowledge based mistakes that can include many variable forms. In terms of violations, there are routine violations, exceptional violations and acts of sabotage (Reason, 1990).

Human error constitutes the main causing factor of the 57% of road crashes and contributes in over 90% of them (Treat et al., 1979). That is the reason that human factor within road safety has received rising attention the last decades (Flumeri et al., 2018). Psychological disciplines have acquired more significance in terms of road safety and transformed to a useful tool for comprehension and interpretation of driver behaviour (Bucchi et al., 2012).

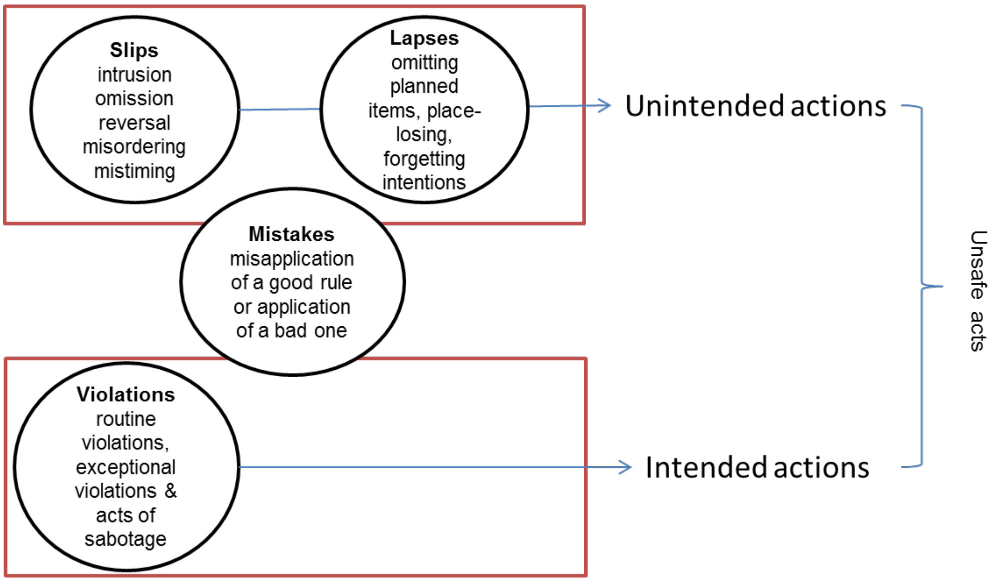


Figure 2.10 Unsafe acts

2.3.3.14 Distractions

Driving distraction, generally, is defined as “the deviation of driver’s attention away from operating safe driving toward a competing activity” (Young et al., 2008). There are several definitions of driver distraction in the literature, though. Ranney et al. (2000) stated that “driver distraction may be characterized as any activity that takes a driver’s attention away from the task of driving”. They placed driver distraction in four categories: visual, auditory, biomechanical, and cognitive distraction. Smiley (2005) defined it as “misallocated attention” while Stutts et al. (2005) as “an object or event that draws one's attention from the task of driving." Last but not least, Olson et al. (2009) indicated that "driver distraction occurs when inattention leads to a delay in recognition of information necessary to accomplish the driving task". Last but not least, according to Liang the significance lying on where the driver is looking, long glances off the road can increase considerably the risk (Liang et al., 2012).

Any cognitive process such as daydreaming, mind wondering, decision making issues, as well as, using in-vehicle information systems (IVI's) such as navigation systems and cell phones can influence driver's attention on driving and therefore, can cause driving distraction (Almahasneh et al., 2014). The two prevalent types of distraction are visual distraction and cognitive distraction. The latter one can be defined as ''mind-off-road'', while the first one as ''eyes-off-road'' (Victor, 2005). Both types of distraction have a potential impact on driving performance such as steering control, lane variation, response to risky situations, and visual perception efficiency.

It has been estimated that driver distraction is a prevalent contributing factor in 25-30% of crashes (Wang et al., 1996) while more recent statistics showing that distracted drivers constitute 9.2% of road fatalities (NHTSA, 2016). There are a large number of ways to measure distraction (e.g., Green, 1995; Tijerina et al., 2003). One can assess driving performance, task performance, ratings of difficulty, spare capacity, etc. (Green et al., 2004).

2.3.3.15 <u>Fatigue</u>

An important safety parameter regarding the human performance is fatigue as it is believed that it is responsible for 2% - 25% of car crashes (Lee, 2008). Karrer et al. (2004) stated that there are not yet satisfactory scientific definitions of fatigue; actually, due to the vagueness of the term, it is possible to be used for various phenomena caused by different factors.

Fatigue has three aspects (Shinar, 2007):

"(1) Bodily changes, such as reductions in physiological potentials and neuron-muscular capabilities, (2) Performance changes, such as output and reaction time and (3) Subjective sensations, such as feelings of tiredness and sleepiness".

Saxby et al. (2007) support the theory of active and passive fatigue (Desmond and Hancock, 2001). The first one is a consequence of physical workload on the drivers from actions such as acceleration changes and steering, while the second one is an outcome of monotonous and unchallenging driving tasks. The characteristics of active fatigue are symptoms of distress, but those of passive one are connected with reduced mental clarity or focus and distracted behaviour.

Lane-keeping performance is reduced by drivers with lack of sleep the previous night. This is similar to decreased performance of drivers with a blood alcohol content of 0.07% (Fairclough and Graham, 1999). Moreover, driving at normally sleeping hours or driving for extended periods are highly associated with raised crash risk.

According to several studies, reaction times are considerably increased due to fatigue especially in emergency situations as it affects driver performance and behaviour. When drivers are lacking adequate sleep, their performance is reduced on attention-based tasks e.g. are slower in risks perception (ERSO, 2006, Strahan, et al., 2008). Therefore, fatigue has a remarkable contribution to crash risk.

2.3.3.16 Environment

There is no doubt that environmental factors play important role regarding traffic crashes or near-crash events. According to empirical findings of a study by Wu and Jovanis, (2011), the conditional probability of a crash is reduced e.g. with the presence of a road-way median, a dry pavement or daylight.

More specifically, environmental factors can be divided into:

Road Characteristics: The road characteristics include roadway alignment, intersection design, roadway profile and pavement type and they affect driving in various ways. Statistically, most fatal crashes happened on straight roadways, while one third of single-vehicle crashes occurred on a curve. A 44.5% of this kind of crashes took place on roadways that were straight, level, and paved with bituminous, while a 16.5% occurred on curved, level, bituminous roadways (United States Department of Transportation, 2010).

Road Classification: Most kinds of crashes (except for the multi-vehicle crashes) occurred in a larger proportion in rural than in urban surroundings.

Weather: The weather factor is highly associated with visibility and roadway surface conditions. While the majority of fatal crashes occurred on dry roads under normal conditions, diverse weather conditions such as snowing and fog, have a serious impact on driving performance.

2.3.4 A Theoretical framework of Factors influencing or identifying crash risk

Factors and variables that affect normal driving and can describe the transition from normal driving to a safety critical scenario have been reviewed. Figure 2.11 categorises the most important parameters and indicators according to the literature review; some of the categories constitute safety indicators that can quantify the transition from normal driving to a safety critical scenario while others include factors that affect this transition, influencing crash risk. More specifically, the environmental factors, the traffic characteristics and the geo-demographical driver profile in combination with human factors parameters affect the driving behaviour that is depicted in turn in the behavioural indicators. Utilising the most suitable from these indicators per case, the transition from normal driving conditions to a deviation can be monitored and, hence, crash risk could be identified.

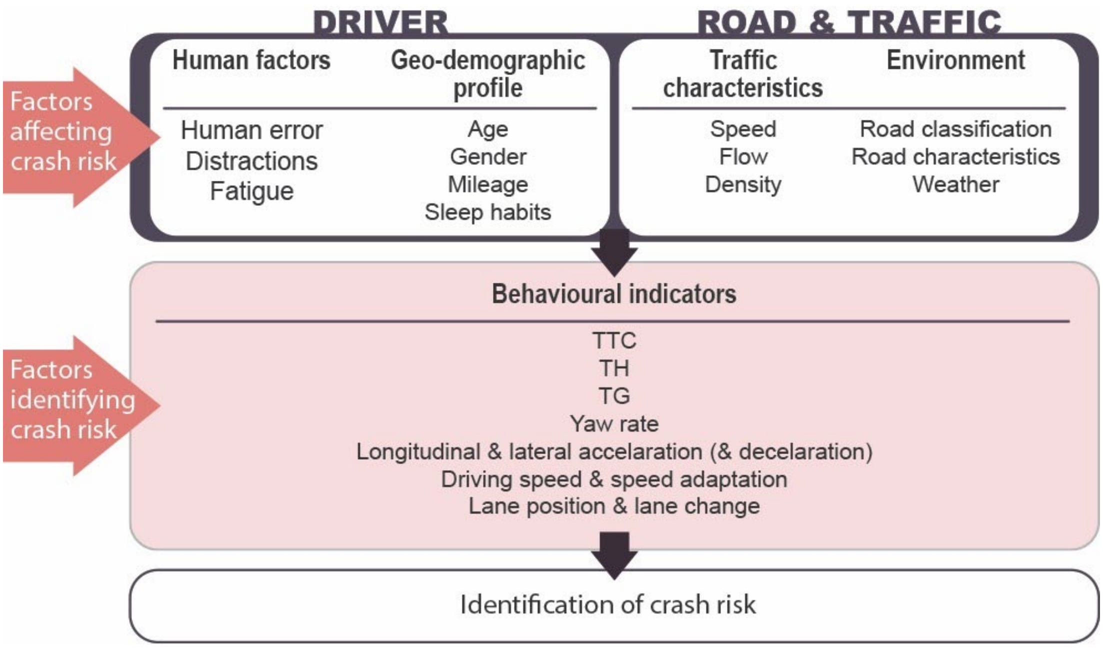


Figure 2.11 Theoretical framework of Factors that affect or identify crash risk

2.4 Methods of investigating crash development and driver behaviour in road safety

Over the years the necessity for approaches to comprehend and assess driver behaviour related to safety, has concluded to 3 substantially different, but complementary methods that enable the exploration of driver behaviour during the crash development sequence. Each of them exhibits advantages and disadvantages that will be cited in the following subsection. Nevertheless, all can benefit road safety providing insights for crash causation and especially the role of driver in it.

2.4.1 Driving simulators

This kind of data derives from experimental studies where the specification of the scenario is defined a priori, similarly with laboratory settings (e.g., Caird and Hancock, 1994; Hancock et al., 1991), or test-track experiments (e.g., Bärgman et al., 2015; Kiefer et al., 2003; Summala et al., 2012). Driving simulation has been used widely for training and research and the main reasons are efficiency (cost) and safety (Nilsson, 1993; Kaptein et al., 1996; Godley et al., 2002; Bella, 2008; Moroney and Lilienthal, 2009). Simulators can produce hazardous driving conditions without the risk of physical injury (J.C.F. de Winter et al., 2012) and without the expenses of a real-world field test (Rudin-Brown et al., 2009, Bella, 2008). Except for these, simulators provide full experimental control (Nilsson, 1993; Kaptein et al., 1996; Godley et al., 2002; Bella, 2008; Moroney and Lilienthal, 2009), ease of data collection (Nilsson, 1993; Godley et al., 2002; Bella, 2008, De Winter et al, 2012), the ability to reproduce events for the participants that are rear in real word (Kaptein et al., 1996), accurate measurements of vehicle proximity to other objects or vehicles (Rudin-Brown et al., 2009) and environmental benefits (lack of fuel consumption and road damage) (Kaptein et al., 1996; Moroney and Lilienthal, 2009).

The use of simulations, however, comes with disadvantages. Simulation sickness is very common, especially in the elderly or under demanding driving conditions. This simulator discomfort that can affect the validity of research results, is believed to have association with system design, technological deficiencies and people's tendency to motion sickness (De Winter et al., 2012, Rudin-Brown et al., 2009). Another disadvantage is the false sense of safety that simulators give as there is not real risk or real consequences, therefore no real responsibility (Käppler, 1993).

2.4.2 In-depth crash investigation data

Data collected from traditional in-depth crash investigations provide answers mostly related to injuries (Fagerlind et al., 2010; Otte et al., 2003; Seeck et al., 2009), and can be very instructive as it may include more than 500 crash-related variables per case (Usame et al., 2015). Studies have examined injury outcomes for different road users and collision types (e.g. Morris et al., 1995; Welsh et al., 2006; Yao et al., 2007), offering a deeper understanding of injury mechanisms. Moreover, researchers have used in-depth data to identify scenarios for automated driving systems and ADAS testing (e.g. Lenard et al., 2014; Nitsche et al., 2017; Nilsson et al., 2018; Sander and Lubbe, 2018).

In depth-crash investigation is based on expert assessments (e.g., Dunn et al., 2014; Habibovic et al., 2013; Sandin and Ljung, 2007; Van Elslande and Fouques, 2007) or other epidemiological methods investigating crash occurrence (e.g., Carney et al., 2015; Hickman et al., 2010; Toth et al., 2003; Victor et al., 2015). Notable in-depth studies in Europe are "Road Accident In Depth Studies" (RAIDS) (Cuerden and McCarthy, 2016), in the UK, which incorporated the legacy studies "On The Spot" (OTS)(2000-2010) , "Co-Operative Crash Injury Study" (1998-2010) (CCIS), Truck Crash Injury Study (TCIS) (1995-2010) and Heavy Vehicle Crash Injury Study (HVCIS) (1995-2010), and the German In-Depth Accident Study (GIDAS)(1999-).

The investigations can either be on-scene (i.e. taking place while the emergency services are attending the scene of the accident) or retrospective (commonly on the next day of the accident, involving the examination of the damaged vehicles) (Cuerden and McCarthy, 2016). It should be noted, however, that retrospective cases, even though can offer a significant insight in the collision's impact on the involved vehicles, usually do not consider crash site evidence (Cuerden et al., 2008). Pre-crash kinematics are usually reconstructed from post-event vehicle position, vehicle disfigurement and tire tracks (Niehoff and Gabler, 2006) while environmental factors such as weather and road conditions are documented along with the driver and witness written thorough descriptions of the event (Paulsson, 2005; Sandin and Ljung, 2007; Seeck et al., 2009). These accounts are collected by questionnaires and interviews and can provide details about the driver state e.g. sleepiness, distractions.

In-depth studies offer highly detailed information on accident causation factors and injury mechanisms (Hill et al., 2001; Reed et al., 2018), however, their representativeness is to be considered. Investigators often give priority to fatal and serious collisions over the ones concluding to slight injuries (Reed et al., 2018). Nevertheless, the approach of in-depth crash investigation data cannot reach the stage of in-depth comprehension of actual driver behaviour for the seconds before the event (Bärgman, 2016). This data could be exploited in epidemiological studies for crashes and injuries (Kullgren, 2008; Lefler and Gabler, 2004), as a basis for vehicle kinematics crash simulations and as a way of investigating crash causation factors (Bärgman, 2016).

2.4.3 Field Operational Tests (FOT)

FOT focus on the evaluation of the use and safety effects of in-vehicle information or warning systems (Bao et al., 2012; Benmimoun et al., 2011; Bezzina and Sayer, 2015; Carsten et al., 2008; Dozza et al., 2010; Fancher et al., 1998; LeBlanc et al., 2006; Ljung Aust et al., 2011; Mononen et al., 2012; Sayer et al., 2011; Sayer et al., 2010; Viti et al., 2008). These studies apply naturalistic driving-like methods and often enable statistical inference as they provide treatment and baseline (control) phases. Nevertheless, the term baseline in NFOTS is also utilised to describe normal, everyday driving with no comparison to a treatment (control) phase (Othman, Thomson, and Lannér, 2014; Sayer et al., 2007; Tivesten and Dozza, 2015). The term Field Operational Test has been used by some authors for studies conducted in actual roads in traffic (Festag et al., 2011).

A significant advantage of these studies is the highest ecological validity (Carsten et al., 2008; Ljung Aust et al., 2011; Sayer et al., 2011) that they provide by evaluating the actual system through the study of natural behaviour in the real world, on real roads, in everyday driving (Schmuckler, 2001). A disadvantage is that usually the systems are not available to drivers as individual products and therefore, it is difficult to distinguish the safety effects of the different systems (Ljung Aust et al., 2011). Moreover, the use of surrogate safety measures (e.g near-crashes) and the selection of the participants (Dozza et al., 2010) set generalisability limitations.

EuroFOT and TeleFOT are characteristic examples of large-scale European FOT studies focusing on active safety systems and effects of nomadic devices respectively (Schagen and Sagberg, 2012).

2.4.4 Naturalistic driving studies

One of the major challenges for reduction or even elimination of crashes is the detection of critical and dangerous situations and the appropriate reaction to them (Lefèvre et al., 2014). New trends in research on crashes involve Naturalistic Driving Studies (NDS). NDS are based on a wide range collection of data regarding the driver, the vehicle, and the environment information in real world and these data sets have proven to be extremely valuable for the analysis of safety critical events such as near crashes and crashes.

Dingus et al. (2006) give a definition of naturalistic as “Unobtrusive observation; observation of behaviour taking place in its natural setting”. Actually, naturalistic driving studies include the collection of data from instrumented vehicles used by their drivers in their routine. A variety of data sources collect the data. These data sources can vary from relatively simple accelerometers, gyroscopic sensors and GPS to different sources as vehicle tracking radar, lane tracking cameras, as well as driver-state sensing such as eye-tracking systems. More specifically, lateral and longitudinal acceleration are measured by vehicle-based accelerometers and these measures help separately or combined with TTC estimated by radar, to primarily identify possible events. Yaw rate is also useful in the identification of large heading changes within a short period of time.

The events are typically detected through unusual vehicle kinematics that are being recorded electronically through accelerometers and gyroscopic sensors. As soon as they are identified due to kinematic measures, the events are screened by the use of forward-facing and interior video; they are kept if are valid safety-related events and discarded if not. Data for the period shortly before, during and shortly after the event are then retained. The outcome is a set of possibly plenteous data that provides insights to near-crashes and crashes that have been formerly unavailable.

In an NDS, data collection lasts per driver from some weeks (Fancher et al., 1998; Leblanc et al., 2006; Najm et al., 2006; Reagan et al., 2006; Sayer et al., 2008) to a few months or years (Hjälmdahl, 2004; Neale et al., 2005; Reagan et al., 2006; Carsten et al., 2008; euroFOT-Consortium, 2010). NDS present important potential for the comprehension and better assessing of crash causation and play a significant role regarding transport safety. This is the main reason for which an important public funding has been provided in Europe, US and Japan in order to perform large-scale NDS (Dozza et al., 2013; Jonasson and Rootzén, 2014).

The 100-Car Naturalistic Driving Study in Virginia is the first study aiming to collect large-scale naturalistic driving data involving 78 out of 100 drivers using their own vehicle (Dingus et al., 2006). The main objective was pre-crash incident relating data for crash causation understanding. Later, as part of the Second Strategic Highway Research Program (SHRP2: www.shrp2nds.us; Antin, 2011; Boyle, 2009), the SHRP2 ND study was conducted, the largest study of its kind involving 3000 volunteer drivers in 6 states of America. One purpose of the Program was to reduce the highway crashes and injuries (more details about the study in section 3.3). The corresponding large-scale ND study in Europe - U-Drive- was fulfilled in 2016 under the funding of European Commission with the aim of studying road user behaviour on both safety and environment (Eenink, 2014).

There were also smaller scale ND studies and different projects in Europe and worldwide adopting ND methodologies (including FOTs):

❖ In Europe:

- **PROLOGUE**: aimed to evaluate the feasibility and usefulness of a large-scale ND study in Europe and to generate relevant recommendations (www.prologue-eu.eu; Sagberg et al., 2011);
- **INTERACTION**: aimed at a higher comprehension of driver interactions with in-vehicle technologies (interaction-fp7.eu);
- **2BeSafe**: focusing on the behaviour and safety of powered two-wheelers (www.2besafe.eu; Laporte and Espié, 2012);
- **DaCoTA**: evaluating the usefulness of the ND method for collecting large-scale, representative information about safety performance indicators and exposure in the different EU Member States (www.dacota-project.eu; Thomas et al, 2013);
- **SeMiFOT**: aimed at implementing and developing the Naturalistic Field Operational Test (N-FOT) method as a method to investigate crash causation and the effect of new safety systems (Victor et al, 2010).
- Large Field Operations Tests (**FOTs**) such as:
  - **EuroFOT**: focusing on the use of Advanced Driver Support Systems, (www.eurofot-ip.eu)
  - **TeleFOT**: studying the use of nomadic devices, (www.telefot.eu).

Field Operational Tests aim attention at evaluation of systems or functions whereas NDS mainly focus on crash-explanatory factors. Data in both cases can be exploited for the analysis of Mobility Efficiency and Environment impacts (Eenink, 2014).

❖ Worldwide:

- **FMT study** (Dinges et al., 2005)
  A study conducted in the US and Canada to explore the effectiveness of fatigue management technologies using NDS on truck drivers.
- **Australian 400 Car-NDS** (Regan et al., 2013)
  Study conducted in Australia with the aim to understand what drivers do when they drive in normal and safety critical situations.
- **The Michigan TRI study** (Sayer et al., 2007)
  Study of University of Michigan Transportation Research Institute aiming to assess the frequency and conditions of drivers' secondary tasks and to explore the relationship of them with driving performance.
- **EMS Study** (Levick and Swanson, 2005)
  Study conducted in Arkansas involving Maimonides Medical Centre in New York evaluating the use of on-board computer monitoring devices, with real time auditing feedback on emergency vehicle drivers' risky behaviour.
- **IT and Lorry Driver Study** (de Croon et al., 2004)
  Study conducted in Amsterdam to explore the effect of on-board computer systems on Dutch truck drivers.

Through naturalistic driving datasets, there is the ability of investigation of all the stages of the crash sequence that include the pre-crash phase, the during-crash phase, and the post-crash phase. According that, recent research has tried using the data to look into crash surrogate analysis; crash risk analysis using crash surrogates; and crash sequence analysis (Wu and Thor, 2015) .

The crash surrogates are frequently forerunners of crashes. Therefore, it would be beneficial to use them so as to not only define crash causal factors but also, recommend potent ways or interventions to halt a crash in time. This is the leading goal of the aforementioned studies (Wu and Thor, 2015) . More specifically, Wu and Thor (2015) argue that it will be beneficial to use NSD data for crash sequence analysis by "quantifying and characterizing a crash surrogate during the progression of a crash". Therefore, there is the opportunity of measuring and predicting the probability of a scenario concluding to a crash. Nevertheless, in order to decisively estimate crash risk, better comprehension of the association between multiple crash contributing factors is needed.

2.4.4.1 NDS as a way of investigation of crash development

In the field of traffic safety, naturalistic driving studies data enables the investigation regarding the development of a crash. NDS collect driving information related to vehicle kinematics, driver behaviour and roadway data (Wu and Thor, 2015). Thus, except for the kinematic variables, there is often the collection of other sets of data. These may include (Wu and Jovanis, 2013):

1. Context variables – they describe the physical attributes (the road and the environment) during the event including the geometric alignment and environmental factors (e.g. day or night, rain or snow). With connecting on-board GPS to existing geographical information systems, there is an opportunity to obtain some geometric characteristics (e.g. most state highway departments hold roadway inventory systems).
2. Event features – pre-event conditions, as well as conditions during event occurrence, for instance, the presence of fatigue and the occurrence of driver distraction (often identified by type of distraction).
3. Driver features – information received during the introduction of the "subject" to the study and typically include age, stated prior driving record, tendency to risky driving and physiological factors such as vision and reaction time.

Although a few parts of events are not able to be investigated (such as the other vehicles drivers' actions and scenes out of the range of cameras and sensors), the observational data from individual drivers over a long time period, including near-crashes and crash events, is an indisputable boon. However, assessing the near crashes and searching for a clearer relationship between them and crashes remains a challenge.

2.4.4.2 Identification of safety critical events and crash risk in NDS

Identifying safety critical events in NDS may be demanding. To date, safety critical events are identified by searching utmost values of vehicle dynamics, e.g., high lateral or longitudinal accelerations, with kinematic triggers (Dingus et al., 2006a; Batelle, 2007; Lee et al., 2011).

While reviewing video sequences to recognise safety critical events, analysts usually make an effort to establish some empathic link with the drivers to understand whether the driver experienced the event as being safety critical. Actually, various drivers, such as sensation seekers might display high decelerations even under normal driving conditions (Jonah et al., 2001). In addition, video information from outside the vehicle is generally not adequate to discriminate between normal driving and safety critical events. On the contrary, rapid reactions and surprised or frightened expressions on the drivers' face – referred as oops reaction in Victor et al. (2010) – are more reliable indicators of safety critical events. Previous work carried out by Kobayashi (2007) and Molinero et al. (2009) reinforces "the underlying assumption that driver reaction is a specific indicator of a safety critical event". Moreover, as claimed by Dozza and Gonzalez (2013) driver reaction "may be the key to pinpointing safety critical events from NDS" for understanding of crash causation and evaluation of intelligent vehicle systems. To date, identification of safety critical events from NDS demands a subjective estimation of crash risk due to the fact that objective definitions based on kinematic triggers exhibit very poor performances (Dingus et al., 2006a; Faber et al., 2012).

Previously in the chapter, indicators that can determine compromises in driver safety have been reviewed. Some of these have been used in different studies (NDS and FOTs) in order to trigger safety critical events from naturalistic driving data. The kinematic search criteria used in several studies for the identification of events are presented in Table 2.4. It should be noted that the vehicle sensors for collecting the data, traffic, road and weather conditions may differ between the following studies and that could have an effect on the triggering thresholds.

Table 2.4 Kinematic search criteria for Safety Critical Events (SCE)

(adapted from Nitsche et al., 2013)

Study	Longitudinal acceleration (g)	Lateral acceleration (g)	Yaw rate (°/s) or Swerve (°/s^2)	Time-to-collision (s)
100-car study (Dingus et al., 2006)	< - 0.6 & > 0.6	> 0.7	≥±4° (within 3 s time window)	$TTC_{front} \leq 4$ $TTC_{rear} \leq 2$
Dacota NDS (Pilgerstorfer et al., 2012)	< - 0.25	> 0.25	n/a	n/a

Study	Longitudinal acceleration (g)	Lateral acceleration (g)	Yaw rate (°/s) or Swerve (°/s²)	Time-to-collision (s)
DDWS FOT (Hanowski et al., 2008)	≤ - 0.35 when u*≥24km/h ≤ -0.5 when u<24 km/h	n/a	≥171.9 °/s² when u≥24km/h	TTC_{front} ≤ 1.8 when u≥8km/h, yaw rate ≤│4°/s│, & azimuth ≤│0.8°│
Driver Distraction in Commercial Vehicle Operations (Olson et al., 2009)	≤ - 0.2 when u≥1.6 km/h	n/a	≥114.6 °/s² when u≥8km/h	TTC_{front}≤2 when u≥8km/h, yaw rate ≤│6°/s│, & azimuth ≤│0.12°│
EuroFOT (Benmimoun et al., 2011)	≤ - 0.6 when u < 50 km/h ≤(-0.4-0.6)*((u-50) /100) – 6 when 50 ≤u≤150km/h ≤ -0.4 when u>150km/h	(0.7-0.25)*(u/40) +0.25 when u<40km/h > 0.7 when 40≤u≤50km/h (0.4-0.7)*((u-50)/50) + 0.7 when 50<u≤100km/h >0.4, u>100km/h	>50 when u< 40km/h (25-50)*((u-40) /10)+50 when 40≤u≤50km/h (15-25)*((u-50) /35) + 25 when 50<u≤ 85km/h >15 when u>85km/h	TTC<1.75
Naturalistic Teen Driving Study (Lerner et al., 2010)	≤ -0.65	≥ 0.75	≥ ±4°/s (within 3 s time window)	TTC_{front} ≤4

Study	Longitudinal acceleration (g)	Lateral acceleration (g)	Yaw rate (°/s) or Swerve (°/s²)	Time-to-collision (s)
Naturalistic Truck Driving Study (Blanco et al., 2008)	≤ - 0.2 when u≥1.6 km/h	n/a	≥114.6 °/s² when u≥8km/h	TTC_{front}≤2 when u≥8km/h, yaw rate ≤│6°/s│, & azimuth ≤│0.12°│
Teen driver study (McGehee et al., 2007)	≤ - 0.5	≥ 0.54	n/a	n/a
U DRIVE NDS	≤ - 0.2	n/a	n/a	n/a
SHRP2 NDS (Hankey et al., 2016)	≤-0.65 & ≥0.5	≤-0.75 & ≥0.75	±8° /s (within 0.75s time window)	n/a

Drivers execute plenty of braking events during normal daily driving, varying from gentle braking for slight speed reduction to forceful braking so as to stop the vehicle as fast as possible, but in a controlled way. Safety critical driving manoeuvres are defined as "unplanned braking or swerving due to preceding erroneous or risky driving by the driver him/herself or other road-users" (Bagdadi and Várhelyi, 2013). Nevertheless, fallible or risky driving is unpredictable regarding the place and the time it may occur and is therefore hard to be recognised using conventional traffic conflict estimation methods such as, e.g. the traffic conflict technique (Hydén, 1987), which is based on observations on a site.

In any case, drivers who perform precarious driving manoeuvres curtail the safety margins not only for themselves but also for other road users. A smaller safety margin to counterbalance their own errors or those of others raises the hazard of safety critical events such as crashes or traffic conflicts (Risser, 1985).

2.4.4.3 <u>Methods employed in naturalistic driving studies data for road safety analysis</u>

Naturalistic driving studies and crash progression analysis are relatively new fields on road safety analysis. Nevertheless, there are several methods that have been developed and involve processing naturalistic driving data.

Data from NDS have been successfully applied to analyse road safety issues using different statistical or other methods. Dozza et al., (2013) presented a method for NDS data analysis called chunking where the data is divided into equivalent segments in order to provide robust parameter calculation and increase statistical sensitivity. Dozza and González, (2013) stated that automatic video processing can significantly help in recognizing safety critical events from big video data and the key for this is the detection of drivers' sudden motion. In line with that, Vlahogianni et al., (2014) reported that there are cases of events that cannot be identified without integrating video analytics. On the contrary, Wu and Jovanis (2013) claimed that with a statistical multi-stage modelling framework it is possible to screen and define SCE without any video screening. However, Tontsch et al., (2013) showed that detecting critical incidents in NDS using only kinematic data is a very difficult process with several issues (many false alarms and missed incidents), especially in urban areas where the employment of threshold-based selection of events could be more problematic due to the irregularity of traffic and the constant braking and accelerating. Furthermore, Scanlon et al., (2015) analysing driver evasive manoeuvring prior to intersection crashes, found that almost 20% of the drivers did not perform any evasive manoeuvre.

On an event-based analysis, a positive correlation between crashes, near crashes and safety-relevant events has been found (Wu et al., 2014). Moreover, the relationship between crashes and near crashes has been conceptualised and a crash-to-surrogate ratio can be estimated (Wu and Jovanis, 2012). On the other hand, Jonasson and Rootzén, (2014) proved, based on statistical extreme value theory, that there is a severe discrepancy between the rear-striking near-crashes and the rear-striking crashes. Kluger et al., (2016) employed a discrete Fourier Transform with K-means clustering to determine patterns in vehicles accelerations and flag parts of time series data that are likely to be near-crashes and crashes. Additionally, Talebpour et al. (2014), identified specifically near-crashes in connected vehicle environments using an algorithm based on drivers' accelerations and behavior during car-following situations and pinpointed the differences between drivers.

On driver level analysis, Wu et al., (2014) report that young drivers have more possibilities to get involved in crashes or crash-related events while drivers' sleeping habits seem to affect the number of events. A method for assessing high risk drivers is the critical jerks (hard brakings) as Bagdadi (2013a) refers to an association between the frequency of the critical braking events and crash involvement. Furthermore, Wu and Lin, (2019) using SHRP2 data and mixed-effects model found that driver behavior, critical driving situations, and environment influence the variation of perception time among drivers.

Taking under consideration the vehicle mass and the relative speed of the road users, event severity can be also estimated (Bagdadi, 2013b). Besides, Zheng et al. (2014) using cluster analysis found that the velocity when braking has strong relationship with the driving risk level involved in near-crash cases, therefore, plays important role in the progression of an event. Similarly, Wu and Thor, 2015 developed an approach (Safety frontier concept) for comparing and dissecting the differences in the crash sequence that lead to different outcomes. Chong et al. (2013) tried to model car following and evasive driver behaviour using fuzzy rule based neural network machine learning technique while Jovanis and Wu, (2015) presented a flexible exposure-based analysis structure that can include driver, event and environment characteristics and detect baseline hazards. Ghasemzadeh and Ahmed, (2018) with logistic regression and multivariate adaptive regression splines (MARS) explored traffic conditions, driver age, experience and posted speed limits and found that they play a significant role on the driver lane keeping ability, while Naji et al., (2017) used K-means clustering and ordered logit regression to find that congested road condition, time of the day, weekend, age and driving years are the important factors related to environment and driver. Moreover, Bärgman et al., (2017) applying counterfactual simulations reported that the choice of the model in order to evaluate intelligent safety systems is significant when the driver is in the loop. Guo et al., (2017) using a case-crossover approach with semiparametric Bayesian models, underlined that it is important to take into account the variability in risk introduced by environment, road geometries and drivers' promptness to be involved in secondary tasks. Lastly, Muronga and Ruxwana (2013) employing the Theory of Planned behavior (TPB), concluded that through the use of NDS, it is feasible to detect hazardous driver behavior which could facilitate the design and implementation of projects related to driver improvement.

2.4.4.4 Naturalistic driving studies approach vs others

Despite the fact that the crash rate decreases, the number of crashes annually is still high. Mechanistic safety features such as air bags, seat belts, and collapsible steering wheels have been significantly developed and they have importantly lowered the crash-relevant deaths. However, systems that support drivers in avoiding crashes are a key requirement for further decline of fatalities on the roads. These driver assistance systems demand a deeper comprehension of driver behaviour before a safety critical event takes place. Historically, data collection for studying driver behaviour was conducted by epidemiological, simulator, and test track studies. Despite these are useful techniques regarding driver behaviour, they are weak in investigating the combination of the factors that result in a safety critical event (Dingus et al., 2006).

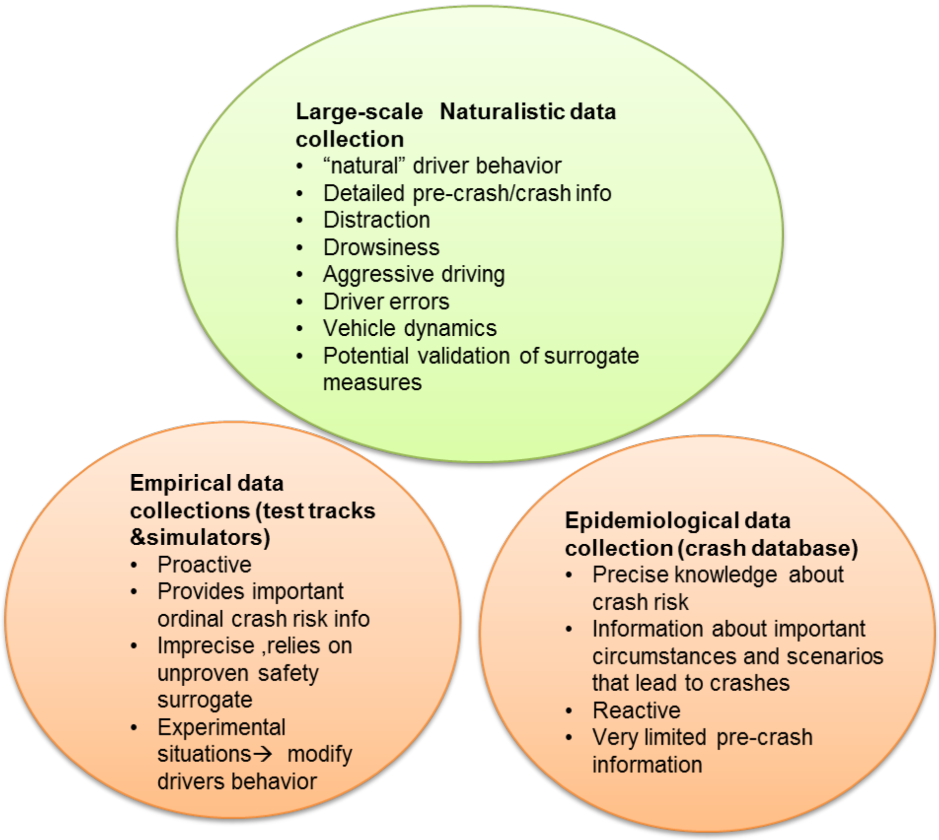


Figure 2.12 The empirical, naturalistic and epidemiological methods in driving safety research (adapted by Dingus et al., 2006).

An illustrating example might be the police's crash report of a rear-end collision as "following too close" while factors that have contributed to the crash could be any kind of distraction, fatigue, traffic backed up from the intersection, etc. In other words, in this example, there might be additional driver and infrastructure related causes. Similarly, test track studies and simulator are not able to replicate the complex driving environment and the concurrent driver behaviours that conclude to several events.

On the contrary, naturalistic driving studies approach fills the gap in existing driving safety research methods as it includes data from a number of vehicle sensors and unobtrusively placed video cameras. Moreover, there are not instructions to the drivers and no experimenter. One limitation of NDS though, is that although they provide a large amount of data, it is mostly representative of baseline, uneventful driving while the safety critical events and especially the crashes are rare.

A key point towards effective crash countermeasures is the deep comprehension of pre-crash causal and contributing factors. Therefore, naturalistic driving data could be beneficial as it provides more details about the near-crash and crash events than is now available, even with a thorough crash investigation. In addition, this data offers richer external validity in terms of the larger context of driving than empirical methods do, such as simulators or test tracks (Dingus et al., 2006). In Figure 2.12, the characteristics of the empirical, naturalistic and epidemiological methods are summarised.

2.5 Summary and Identification of research gap

Over the years, road safety has moved from a reactive to a more pro-active approach that entails acting before crashes occur. Furthermore, a systems thinking approach is adopted which incorporates all the components of a system, i.e. road users, vehicles, roads, that contribute to crash risk.

Human factors are observed as the most unforeseeable in the driver, roadway and vehicle system. Changes in driver behaviour could lead to significant changes in traffic safety. Hence, different types of models have been employed to understand and predict driver behaviour in order to enhance road safety as it constitutes the prevalent cause of crashes, contributing to over 90% of them.

In line with the systems thinking approach, the integrated safety chain model by Tingvall documented the stages of crash development working backwards from a possible event until normal driving. The conditions around normal driving affect the absolute levels of safety. Defining normal driving constitutes a very important step for comprehension and more effective detection of abnormalities in driving and this has not gained systematically the attention of the researchers. More specifically, a relatively overlooked subject is the identification of SCE in NDS. This task considers the last stages of the crash sequence. Quantifying normal driving and detecting early deviations from it, that might lead to a compromise in driver safety, concern the very first stages of crash development and has not been extensively researched. Successful detection of deviations early in crash sequence could halt them before they culminate into crashes and this requires the investigation of the transition from normal driving conditions to safety critical driving scenarios.

This literature review revealed factors that influence crash risk as determined from miscellaneous tools, measures and indicators. There is a necessity to identify the suitable ones in order to determine key abnormalities and crash risk and establish the way in which these can be earlier detected within crash sequence.

The four main methods of investigating crash development and driver behaviour are driving simulations, in-depth crash investigations, field operational tests and Naturalistic driving studies. The literature showed that NDS, due to their nature, provide advantages that can be best exploited in driver behaviour analysis and pre-crash conditions exploration. Therefore, NDS appear to be the most suitable method to utilise in order to approach this thesis' research goals and as relatively new method is yet to be fully explored and utilised. Driver behaviour measured in NDS and depicted in vehicle kinematics is still to be investigated to give insights about the pre-crash driving conditions and driving style that can affect crash risk.

The literature review indicated a dearth of research regarding the early stages of crash development including normal driving and first deviation from it. It also revealed the advantages of NDS comparing with other methods of driver behaviour investigation and the usefulness of several indicators that can be employed to identify compromises in driver safety. Therefore, this research will attempt to fill in the gaps concerning the concept of normal driving and the ways to detect the first deviation from it by analysing safety critical scenarios in NDS, using behaviour indicators.

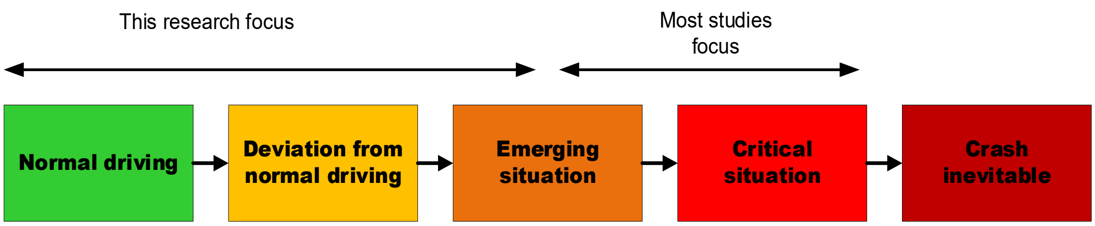


Figure 2.13 The focus of this PhD research regarding the crash sequence

3 Methodology

3.1 Introduction

This chapter includes the theoretical basis of the analytic methods applied, the general research approach together with the type of data and data analyses that were undertaken to fulfil the aim of this thesis. It describes all the philosophical underpinning to the chosen methods and demonstrates the pathway from the research gap to the results.

The choice of the data was an important decision for the fulfilment of the research objectives and several methodological decisions were made for the data request proposal to be submitted. Thus, a brief description of the specific data and the steps towards obtaining it are described here, in the methodology chapter. The data manipulation process with the preliminary analysis will be thoroughly described in Chapter 4 as it comprises significant part of the analysis. After the more generic section of the choice of the research approach and the data, the aim and objectives of the thesis are repeated in the form of a research design where the methods used per objective are also mentioned. Finally, the methodology applied for every objective is analysed in the relevant next sections of the chapter.

Naturalistic driving studies data was employed and specifically SHRP2 NDS data was obtained to investigate the transition from normal driving to safety critical driving scenarios. This was mainly conducted in 3 stages:

1) Firstly, safety indicators were quantified during normal driving conditions that were identified in the data.

1 minute of time series data per trip representing normal driving conditions was employed. Understanding normal driving and providing dynamic thresholds to quantify it will set the basis to detect deviations.

2) Secondly, safety indicators were developed and employed for the whole crash and near-crash sequence development to detect patterns of deviations from normal driving.

2.5 minutes of time series data per trip from normal driving conditions until the safety critical events were utilised. Exploring the whole event development will provide information regarding the onset of deviation and driving patterns. TTC indicator showed no pattern, rendering further analysis necessary. A model was developed to further investigate the evolution of TTC during event sequence development. The modelling methodology though is thoroughly presented in the last section of this Chapter. Moreover, the analysis of the other indicators in this stage raised the interest for the last seconds prior to the events when drivers perform evasive manoeuvres.

3) Safety indicators were analysed during safety critical driving scenarios to determine “evasive” braking and steering manoeuvres.

The last 30 seconds of time series data from every trip prior to safety critical events were used. Examining the last seconds before the event will give insights in pre-event driver behaviour and provide thresholds for emerging situations.

Figure 3.1 illustrates the time period concerning every stage of analysis in the event sequence development, from uneventful driving to safety critical event.

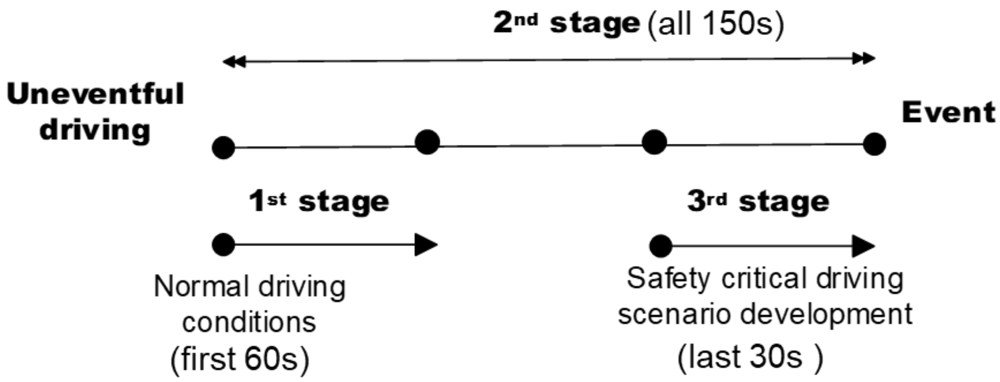


Figure 3.1 Stages of analysis regarding the event sequence development

3.2 The choice of the research approach (Naturalistic driving studies approach)

The literature review revealed factors that affect normal driving and variables that are able to explain deviation from normal driving. In order to detect this deviation and explore these factors and the way they influence driving, there are some methods available to investigate the whole concept of driving. Given all the advantages of the naturalistic driving studies (reviewed in Chapter 2), in the frame of this thesis research, naturalistic driving data was decided to be used in order for the research aim to be fulfilled. More specifically, data from SHRP2 NDS, the largest study of its kind, has been acquired and used in the framework of this research. Moreover, the researcher conducted a small scale driving study to become familiar with the data collection processes and safety critical events detection methods.

3.3 The choice of the data

Once the aim and the objectives of the research project were set, it was essential as the next step to find the suitable data for analysis. Several datasets were considered such us data from TELEFOT Project or NDS data from Austria managed from KFV. However, SHRP2 NDS Data, was identified as the most appropriate for the purposes of this PhD. SHRP2 NDS is the largest and most comprehensive study of its kind and provides numerous safety critical events (crashes and near crashes) for investigation along with a series of variables regarding the drivers and the events per se that could add to the value of analysis. As this research focuses on the transition from a normal driving situation to a safety critical driving scenario, a heavily post-processed and well-structured dataset that provides plenty of safety critical events and relevant variables for analysis was considered the best choice. The other two datasets considered, were not of the same magnitude and the provided variables were limited as were the safety critical driving situations.

This chapter will justify and support the choice of the specific dataset through a brief description of the SHRP2 NDS Database. Furthermore, the decisions taken for the data request are presented as part of the methodological analysis approach.

3.3.1 Strategic Highway Research Program 2 Naturalistic Driving Study Data (SHRP2 NDS)

The SHRP 2 NDS, the largest study of naturalistic driving behaviour to date, monitored approximately 3,500 participant drivers and produced over 4,3 years of naturalistic driving data between 2010 and 2013. Data were collected from six sites around the United States. The largest collection sites were in Seattle, Washington; Tampa, Florida; and Buffalo, New York (McClafferty, 2015). Over 3,300 participant vehicles were instrumented with a data acquisition system (DAS) that collected four video views (driver's face, driver's hands, forward roadway, rear roadway), vehicle network information, e.g., speed, brake, accelerator position, and information from additional sensors included with the DAS, e.g., forward radar, accelerometers, alcohol sensor (Hankey et al., 2013). More specifically, the data acquisition system was developed by VTTI (Virginia Tech Transportation Institute), and consists of four cameras, GPS, forward radar, vehicle network information, on-board computer vision lane tracking system, other computer vision systems and capability of storing data. The collected NDS data provides vehicle kinematics and network information recorded at 10 Hz frequency (10 observations per second) such as speed, acceleration, and steering wheel position.

Finally, 5,512,900 trip files were collected, and a manual validation of the algorithms used, identified 1,549 crashes and 2,705 near-crashes. A longitudinal deceleration-based algorithm produced the highest percentage of valid crashes and near-crashes. Baselines were selected via a random sample stratified by participant and proportion of time driven. 20,000 baselines, including all drivers in the SHRP 2 NDS, were prepared and an additional 12,586 baselines are also available for a fully proportional representation of all drivers in the study. Moreover, in the SHRP2 NDS the trip files represent 3,353 vehicles and 3,546 unique participants. The study collected approximately two petabytes of data, which can be categorized as shown in Figure 3.2.

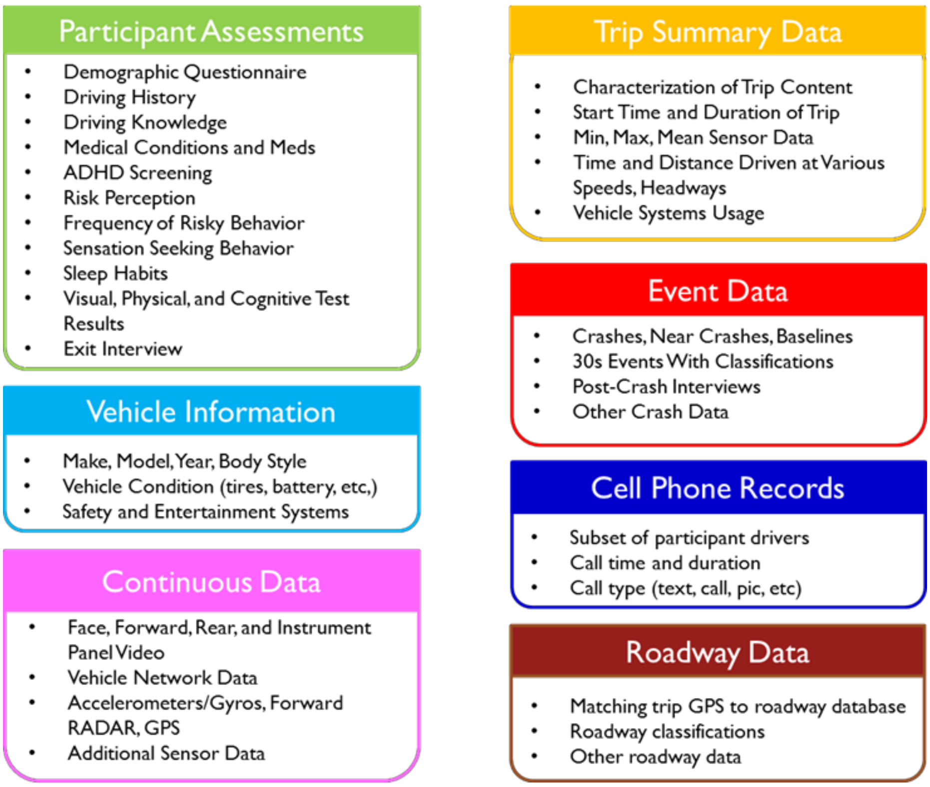


Figure 3.2 Representation of the data categories collected in the SHRP 2 project (Hankey et al., 2016)

Figure 3.3 depicts a screenshot of the data categories in Insight (the SHRP2 NDS data access site).

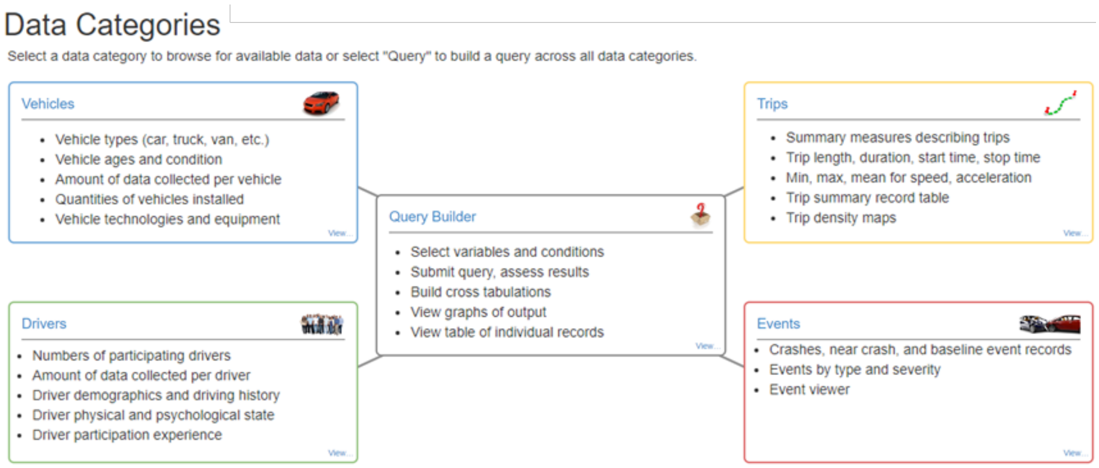


Figure 3.3 Screenshot of Insight Data categories

3.4 Steps towards obtaining the SHRP2 NDS data

3.4.1 Cost

The SHRP2 NDS dataset is very comprehensive and rich. In order to obtain part of it, there was a noticeable cost involved. After a successful research proposal to the Insurance Institute of Highway Safety (**IIHS**), a fruitful collaboration started, and financial support was received rendering the access of the data possible. The proposal can be found in the Appendix.

3.4.2 Online Data observation and methodological decisions

The Insight website provides visual access to all data categories; therefore, navigation through this enlightened the understanding of the data and proved to be useful for deciding what it is needed in order for the research questions to be met.

⇒ **1st decision to take: Videos or not?**

Initially it was not clear if videos will be used as part of the analysis for this project. After deep consideration of the research aim and objectives, it was decided that videos will not add significantly to the value of the analysis planned as the research will focus on vehicle kinematics to investigate the transition from normal driving to safety critical driving scenarios.

⇒ **2nd decision to take: What “control” data will be used regarding driving and drivers?**

All the crash, near-crash and baseline files contain records with all relevant data for about 20 seconds before and 10 seconds after the event. Baseline files of event-free driving are also produced and contain similar 30-second epochs either selected at random from all NDS trips or records matched to crashes and near-crashes. For instance, for a specific crash, epochs could be selected from the same driver’s trips at the same time and day of week on the same or similar roads.

Therefore, as the baseline events are producing from the same drivers (that have been involved in events) and trips (that usually concluded to events) , the decision was to ask for 2 more minutes of time series data before the event in order to use this as baseline, uneventful driving sample data. In other words, no baseline events were asked but it was decided to create them from the data it would be received.

Moreover, as the focus of this thesis is on the analysis of the 150 seconds period of pre-event kinematics, secondary events have not been taken into consideration due to the lack of confrontation with the initial data request criteria (secondary events occur a few seconds after the main ones).

As far as the drivers are concerned, there was no one who has no events at all. The only option that was given, it was data from drivers that completed the questionnaires but never participated in the actual driving process, therefore there was not time series data available for them. Except for the first part of the analysis that explores normal driving conditions, the rest of the research required such dataset (data from events) and this does not impose a bias existence problem.

3.4.3 Confidentiality assurance

Strict confidentiality protocols needed to be followed accompanying with procedures such us training in Human subjects Protection, signing ethical approval list and providing encryption of the data in the personal or other computer.

The data itself is encrypted, does not contain any participant name or any other identifying information and latitude or longitude are not provided for the crashes or the events that happened very close to the start of the trip.

3.4.4 Data request

After the confidentiality agreement was finalised and the methodological decisions were made, the data request proposal was submitted and an agreement was signed between Loughborough University, VTTI and IIHS (between the respectively responsible people of every organisation). The data was finally released in the end of June 2016. The data request proposal can be found in Appendix B.

3.5 Research Design

The aim of this thesis is divided into 7 objectives that will be achieved through methods that are listed below in the research design. As Yin (2009) described, research design is “a logical plan for getting from here to there, where “here” may be defined as the initial set of questions to be answered and “there” is some set of conclusions”. The research design, presented in Table 3.1, is actually a framework that shows the methods and data that will be used so as to achieve the aim and the research objectives. It depicts step by step the process from understanding the research problem to the interpretation of the findings.

Table 3.1 Research design

Aim	Objectives	Methods	Data	Chapter
To investigate the transition from normal driving to safety critical scenarios	1.To formulate a comprehensive theoretical framework of factors quantifying the transition from normal driving to a safety critical scenario and influencing crash risk.	*In-depth critical review of literature*	Research papers, journals, books	Chapter 2 (mainly §2.3, §2.3.4)
	2.To examine the potential contribution of Naturalistic Driving Studies to an understanding of vehicle kinematics	*Pilot study*	Data from Instrumented vehicle (driving study data)	Chapter 3 & 5 (§3.5.1, §5.2)

Aim	Objectives	Methods	Data	Chapter
	3.To characterise safety indicators during normal driving	*Empirical method, descriptive statistics, linear regression & statistical tests*	Pilot study data & SHRP2-NDS	Chapter 5 (§5.3)
	4.To develop safety indicators for detecting deviations from normal driving	*Empirical method, statistical tests*	SHRP2-NDS	Chapter 5 (§5.4)
	5.To formulate safety indicators during safety critical driving scenarios	*Algorithm for events extraction* *Algorithm for overlapping events, statistical tests*	SHRP2-NDS	Chapter 5 (§5.5)
	6.To model the evolution of TTC values during event sequence development	*Multilevel mixed effects model*	SHRP2-NDS	Chapter 5 (§5.6)
	7.To recommend potential thresholds for safety indicators for designing safe and trusted ADAS	*Research results*	SHRP2-NDS	Chapter 5 & 6 (§5.3, §5.5)

Objective 1 has been discussed in Chapter 2 where the relevant literature was reviewed while Objective 7 concerns the whole methodology. The sections that follow discuss the methods applied to approach the remaining objectives.

3.6 Pilot driving study design (Objective 2)

A short-scale driving study (pilot study) was conducted in order to gather driving data with the objective to explain the deviation from normal driving, find ways to detect it and investigate the thresholds where the deviation initiates. This pilot study would be an opportunity to validate the methods to collect and analyse naturalistic driving behaviour data. In other words, it constitutes a great chance to acquire familiarity with NDS process, vehicle equipment and the way that data is produced.

The vehicle that has been used was a fully instrumented Ford Mondeo owned by Loughborough Design School. The vehicle's equipment included four video cameras (forward road view, 2 for driver face and driver reaction from the passenger seat), GPS, speedometer and accelerometer (see Figure 3.4). Regarding the route, the initial thought was that inside the town centre is more possible to have deviations from normal driving or unusual incidents. However, in order for the final route to be chosen, a map of Loughborough was given to some drivers who were asked to note which points (junctions, roundabouts, etc.) they find difficult or problematic. Taking these under consideration, a route has been selected by the researcher so as to represent urban driving conditions in moderate traffic. It includes traffic lights, roundabouts and pedestrian crosswalks (Figure 3.5).

The study included 5 drivers (convenience sample as they were all LDS PhD students) who were asked to drive the test route using the fully instrumented university vehicle with the researcher as passenger. The test started at the Design School and ended at the Design School. With the sampling frequency at 100 Hz and the duration of every trip at 30 minutes, the data available for analysis resulted in a total of 1,343,027 observations.

The participants were asked questions about the driving task whilst driving and a questionnaire was filled in afterwards, in a form of a small interview. Prior to driving the test route participants had approximately 5 minutes to drive around campus with the researcher to familiarise themselves with the test car. The questionnaire and the participants consent form can be found in the Appendix A.

In order to process the data from the study, a software (with a built-in noise filter) developed by Race Technology was used that enabled the researcher to watch thoroughly the videos frame by frame and make diagrams of the speed, acceleration and other parameters in order to understand the data. SPSS statistics 19 was also used to provide the summary statistics. The results of the pilot study are presented in Chapter 5.

Figure 3.4 View of the in-vehicle cameras

Figure 3.5 Pilot study route

3.7 SHRP2 NDS Data processing framework

Although the data received from VTTI was heavily processed, a series of steps were taken to formulate the final dataset for analysis. Extensive data manipulation was undertaken to ensure the validity of the dataset, to provide the desirable variables and to develop the suitable for the specific analysis form of the dataset. The process included preliminary analysis for data understanding, data cleaning and transformation, TTC calculation, missing values imputation and other. The details about the process with the data processing framework and all the methodological decisions will be thoroughly described in Chapter 4.

3.8 Quantifying safety indicators during normal driving (Objective 3)

In the framework of this PhD, SHRP2 NDS data has been employed to explore normal driving and risk development. The main purpose, on a first stage, is to identify a range of indicators and validate their thresholds to detect a deviation from normal driving. It is also important to see whether the thresholds are consistent across different operational conditions (e.g. speed, traffic density) and driver groups (e.g. male vs female and young vs other drivers). For these purposes, a methodology consisting of four steps is developed:

Step-1: Identification of indicators that can be used to detect the deviation from normal driving

The first step is to identify variables that can characterise normal driving. Some of the variables, reviewed in Chapter 2, have been utilised in different studies and can be employed to trigger a SCE flag. In other words, they that can be used to identify and extract events of interest, i.e. safety critical events, in large naturalistic driving datasets. The kinematic search criteria used in several studies for the identification of SCE are summarised in Table 2.4. The criteria used in previous studies can provide a useful basis for determining the corresponding thresholds for normal driving. The variables that were chosen are: Longitudinal acceleration, lateral acceleration, yaw rate and TTC.

Step-2: Determination of Initial Thresholds by exploiting the values from existing studies.

These can be termed as 'Initial Thresholds'. It is noticeable that the identification of a safety critical event is mostly dependant on a combination of different indicators rather than only one. The use of multiple indicators would increase the overall correct detection rate of critical events by increasing the discrimination power. Likewise, the detection of normal driving would also require similar indicators; perhaps with different threshold values (see Figure 2.13, p.53). Based on arguments raised from Table 2.4, on relevant literature and the pilot study, initial indicator thresholds are proposed for normal driving.

Step-3: Validation of the Initial Thresholds

In order to validate the initial thresholds and examine whether they are applicable in detecting the deviation from normal driving, SHRP2 NDS data was analysed to determine the threshold values of the indicators. There were 3 sub-tasks involved:

⇒ *Step-3(i): Identify 'normal driving' (determined in the Chapter 2)*

Each set of time series data culminated in a crash or near-crash and the data for the prior two and a half minutes were requested. For the purposes of this study, only the first minute data of these two minutes of every event were selected and used as representative of the drivers' baseline driving (Figure 3.6, yellow arrow). Given that the crash development lasts for a few seconds, VTTI provides the last 30 seconds in case only the data from the events is requested and the results of a preliminary analysis, 90 seconds before the event (data from 2.5 minutes time prior to the event until 1.5 minute before the event) are safely considered to be uneventful, normal driving. However, some events happened on the very start of the trip before the end of two minutes period. These events were excluded. The details of the data cleaning process will be discussed in Chapter 4.

As most of the variables in time-series data had a sampling frequency of 10 Hz, 600 observations of every event were taken and merged in a final large dataset of 1,084,802 observations representing normal driving scenarios for 553 drivers involved. This dataset gathers data from 774 events consisting of 647 near-crashes and 127 crashes.

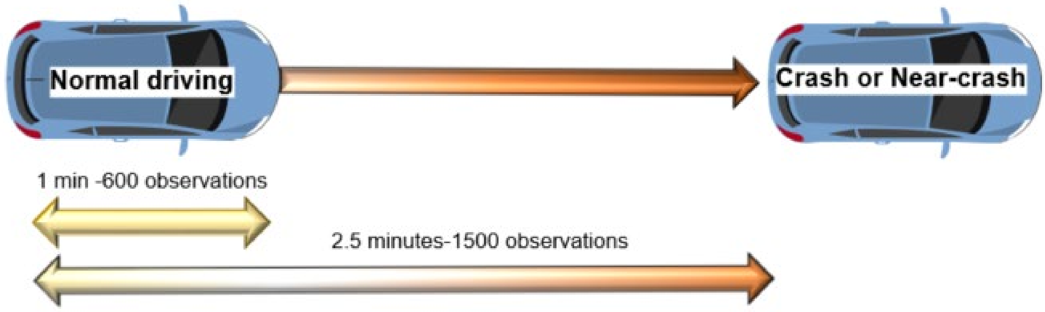


Figure 3.6 Selected Data for analysis

⇒ *Step-3(ii): Analyse the data to calculate real values of the indicators from normal driving*

After the normal driving dataset was extracted, descriptive statistics were employed to depict the values of safety indicators during baseline conditions. More specifically, the 1^{st} and 99^{th} percentile (in some cases the 99.9^{th} percentile as complementary) were utilised to validate the initial thresholds. The indicators that were examined were longitudinal and lateral acceleration, yaw rate and TTC and all of them, except for TTC, included positive and negative values. 1^{st} percentile value denotes that 99% of the values are higher than this; hence it could be useful only for indicators with negative values i.e. deceleration, as it can provide a threshold; the researcher can be 99% confident that this threshold will not be exceeded during normal driving conditions. The opposite applies to positive values, i.e. acceleration, where 99^{th} percentile is more useful as it represents the value which 99% of the values did not exceed. Figure 3.7 illustrates the above argument presenting the left (negative) and right (positive) side of a normal distribution. Normal distribution has been used here as an example; no assumption has been made for the threshold setting regarding the indicators' distribution.

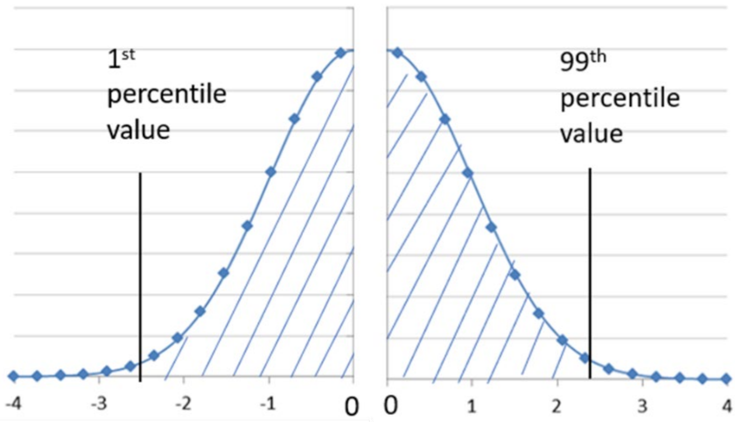


Figure 3.7 1st and 99th percentile in normal distribution

⇒ *Step-3(iii): Examine whether the thresholds are consistent across different driver groups (male vs female; young, middle-aged and older; etc.).*

Driving style varies by gender, age, culture and other operational conditions. Therefore, it would be interesting to see how the distributions of the values identified in Step-2 change with these factors. Composite variables of gender and age were computed in order to investigate whether the thresholds are consistent across different geo-demographic profiles, i.e. young drivers and old drivers, males and females. Indicators' values are better represented by their medians rather than their means; therefore, two non-parametric tests are employed:

- Mann-Whitney two-sample statistic (also known as the Wilcoxon rank-sum test): this test examines the hypothesis that two independent samples (unmatched data) are from populations with the same distribution (Wilcoxon, 1945, Mann and Whitney, 1947).
- Median test performs a nonparametric k-sample test on the equality of medians. It examines the null hypothesis that the k samples were drawn from populations with the same median.

Step 4: Modification of the "Initial Thresholds" based on the findings in Step-3.

The results of the analysis in Step-3 would provide important information whether the initial thresholds need to be modified. It is envisaged that thresholds may vary by the socio-demographic conditions of the drivers. In this task, thresholds will be modified based on the findings in Step-3.

Based on the analysis in Step-3, these indicators depend on many factors and therefore it may be difficult to identify a fixed threshold representing all conditions. In order to elucidate this argument, a 1st percentile (except acceleration for which it is actually 99th percentile) values of these indicators were derived for different speeds of the ego-vehicle. Based on these data, it is recommended that a functional equation, by developing a linear regression model, should be employed to determine a threshold value for an indicator. The functional relationship between an indicator and the speed can be written as follows:

$$\text{Indicator's threshold} = \mathrm{f}\ (\text{Speed})$$

e.g. for Acceleration threshold: Acc_{th}= f (Speed). The same relationship applies for every indicator of the study. The generic bivariate regression model takes the following form:

$$Y_i \ = \beta o + \beta_1 X_i \ + u_{\ i} \qquad (3.1)$$

Where Y_i is an estimate of the dependent variable; βo is the constant of the model; β_1 is the regression coefficient for the first predictor (X_i) and $u_{\ i}$ is the difference between the observed and predicted values of Y for the *i*th case. $u_{\ i} \sim (0,\sigma^2)$ as normally distributed is ignored in the model estimation. Therefore, the following linear regression equations for each indicator are formed:

$$\text{TTC}_{\text{th}} = \beta o + \beta_1 * Speed$$
$$\text{Acc}_{\text{th}} = \beta o + \beta_1 * Speed$$
$$\text{Dec}_{\text{th}} = \beta o + \beta_1 * Speed$$
$$\text{LAcc}_{\text{th}} = \beta o + \beta_1 * Speed$$
$$\text{YR}_{\text{th}} = \beta o + \beta_1 * Speed$$

The equations have been developed in a 95% confidence interval, except for the TTC (90%) and as the distributions of lateral acceleration and yaw rate were found to be symmetrical only the negative values (left-side) were used.

3.9 Safety indicators development for detecting deviations from normal driving (Objective 4)

In order to investigate the progression of the crashes and near crashes and formulate a picture of the state before the event, the need to produce representative indicators emerged. In this stage the research focus was on how the safety surrogate measures (TTC, longitudinal deceleration, longitudinal acceleration, lateral acceleration, yaw rate) for 127 crash and 647 near-crash events change with respect to the time progression from epoch 1 (normal condition) to epoch 1,500 (crash or near-crash condition). For instance, for the case of crash events, there are a total of 127 TTC values for each of the 1,500. In order to see how these 127 TTC values (from 127 events) changes from one epoch to the next, percentiles are calculated. Data were restructured in columns each containing all the values of each timepoint for all the events (e.g. 1st observation of the first event, 1st observation of the second event, …1st observation of last event) and the fifth percentiles of each timepoint of all the events sequence, were calculated. Since low values of TTC, deceleration, negative lateral acceleration and negative yaw rate imply unsafe conditions, the left-hand side of their distributions is of interest. Therefore, calculating a 5th percentile value is more logical than that of a 95th percentile value. For instance, the 95th percentile value of 127 TTC values is 168 seconds at epoch 1,500 whereas this is 0.8 seconds for the 5th percentile value. This means that only 5% of the 127 observations have TTC values less than 0.8 seconds. Figure 3.8 illustrates the process that was followed to generate the 5th percentile indicators for the crash relevant data. 1st percentile indicators were also calculated but they did not seem to depict any pattern; hence they were not useful towards the objective of this analysis. The procedure was the same for near-crash data and it was applied to TTC, deceleration, negative lateral acceleration and negative yaw rate. Lateral acceleration and yaw rate entire distributions were symmetrical, therefore only the negative values were chosen for analysis to comply with the choice of the 5th percentile explained above. Longitudinal acceleration exhibited a very few observations in many timepoints that rendered the generation of a valid indicator impossible.

Examining the distributions and the descriptive statistics of these indicators could give a picture of the onset of the deviation in the course of the crash or near crash development. It is worthwhile to examine the possible differences when the event severity changes from near crash to crash, as well.

The same method was applied to create indicators for drivers of different sex and age group (male and female drivers, young, adult and older drivers) or different kind of crashes (road departure events and all other) and event severity (crash and near-crash events).

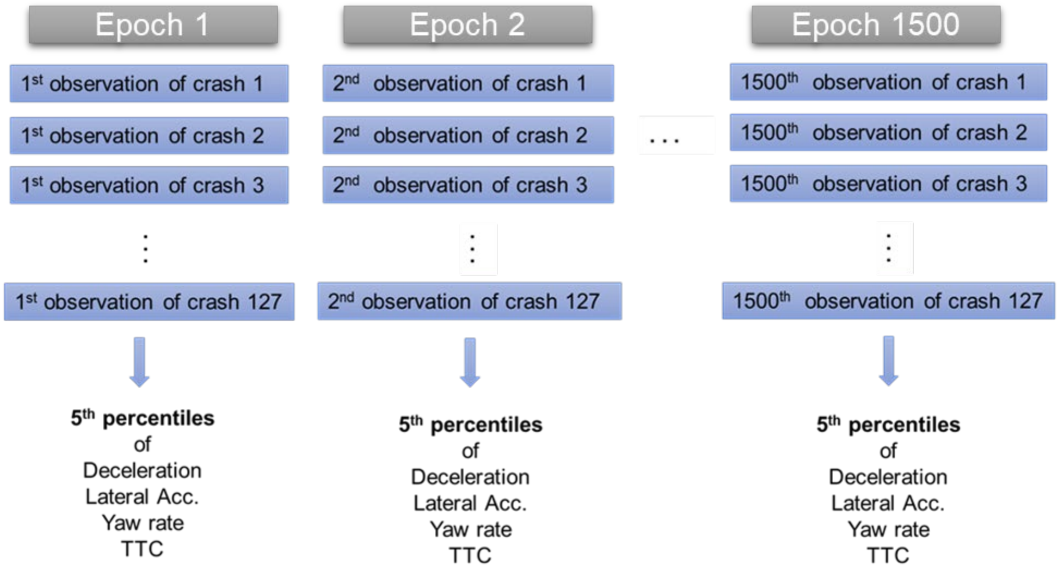


Figure 3.8 Generating process of 5th percentile indicators (127 crashes with 1500 epochs /observations per indicator)

3.10 Safety indicators during safety critical driving scenarios (Objective 5)

The generation of the 5th percentile indicators and the investigation of the whole crash and near-crash sequence development revealed that it would be useful to examine the last seconds before the events, where the drivers responses to the imminent situations take place and the transition from normal driving to safety critical scenarios can still be examined. For this analysis, the time period investigated was 30 seconds before the event. This was also the duration of the data that VTTI releases in case only the safety critical events are requested. It constitutes a significant data segment or time period as it includes the safety critical events and also provides the chance to investigate pre-event driver behaviour.

The investigation of pre-event driver behaviour can provide valuable insights regarding the pre-event conditions and the evolution of a risky situation. The analysis of 5th percentile indicators showed that some indications of deviation could be found when exploring the deceleration, yaw rate, and lateral acceleration. Exploring the braking and steering manoeuvres of the driver could aid in understanding and detecting deviations earlier in the crash development and for this purpose the safety indicators deceleration and yaw rate will be the focus of this analysis. Lateral acceleration was not part of this analysis as it is highly correlated with yaw rate, therefore the latter was decided to be used for drivers' steering manoeuvres detection. It should also be noted that for this part of the research, the absolute values of yaw rate were used due to the purpose of the analysis that was to capture all the drivers manoeuvres prior to the safety critical events and consider their correspondence with TTC.

3.10.1 Extracting events of interest

The important step is to identify braking and yaw rate events of interest in the available data. Towards this direction, algorithms were developed and applied with the use of MATLAB R2018. The specifications and criteria utilised along with the exact processes are described below.

3.10.1.1 Extracting deceleration or yaw rate events of interest - the algorithm of event duration identification

A deceleration event according to this analysis occurs when the deceleration exceeds the 3 standard deviation threshold for more than 10 timestamps, therefore for more than 1 second. The threshold has been calculated by the deceleration values of the same trip/driver. Likewise, a yaw rate event occurs when the yaw rate value exceeds the 3 standard deviation threshold for more than 7 timestamps, therefore for more than 0.7 second.

A driver personalised threshold of 3 standard deviations was used as this way values that represent extreme ends of deceleration or yaw rate distributions can be captured (0.3% are below these values). For the duration studies have used from 0.5s (Wang et al., 2005; Scanlon et al., 2015) to 5s (Chevalier et al., 2016) to investigate rapid deceleration events or capture events of interest. The duration for this analysis was set empirically at 1s for deceleration and based on the researcher's perception that drivers steering manoeuvre is quicker than the braking one, 0.7s for yaw rate. This is also considered adequate time for a braking or steering manoeuvre respectively in order to be further analysed.

The data extraction process that has been developed and adopted is described as follows (Algorithm 1):

1. Select normal driving data of every trip- Isolate the first 600 observations (1min) of every trip
2. Calculate for this dataset the 3 standard deviations deceleration threshold for every trip/driver
3. Back in the initial dataset isolate the last 300 observations of every trip
4. Match the last 300 observations dataset with deceleration 3 standard deviations values
5. Identify the deceleration values that are higher than the relevant value of 3 standard deviations consistently for at least 10 consecutive timestamps(1s)
6. Call it event k.
7. Calculate the duration, mean, max and min of the event k.
8. Keep the relevant deceleration values and the corresponding TTC values for the deceleration event.
9. Go to the next values exceeding 3 standard deviations for at least 1 s.
10. Call it event k+1.
11. Calculate the duration, mean, max and min of the event k+1.
12. Place it next to event k.
13. Repeat the same process until trip finishes.
14. Go to next trip.
15. Repeat 1-14 until all trips finished.
16. End.

Similarly, for the yaw rate (Algorithm 2):

1. Identify the yaw rate values that are higher than the relevant value of 3 standard deviations consistently for at least 7 consecutive timestamps (0.7s)
2. Call it event j.
3. Calculate the duration, mean, max and min of the event j.
4. Keep the relevant yaw rate values and the corresponding TTC values for the yaw rate event.
5. Go to the next values exceeding 3 standard deviations for at least 0.7 s.
6. Call it event j+1.
7. Calculate the duration, mean, max and min of the event j+1.
8. Place it next to event j.
9. Repeat the same process until trip finishes.
10. Go to next trip.
11. Repeat 1-10 until trips finished.
12. End.

Figure 3.9 presents the algorithm of deceleration event duration identification. The corresponding algorithm of yaw rate is similar therefore, it is not illustrated in a diagram.

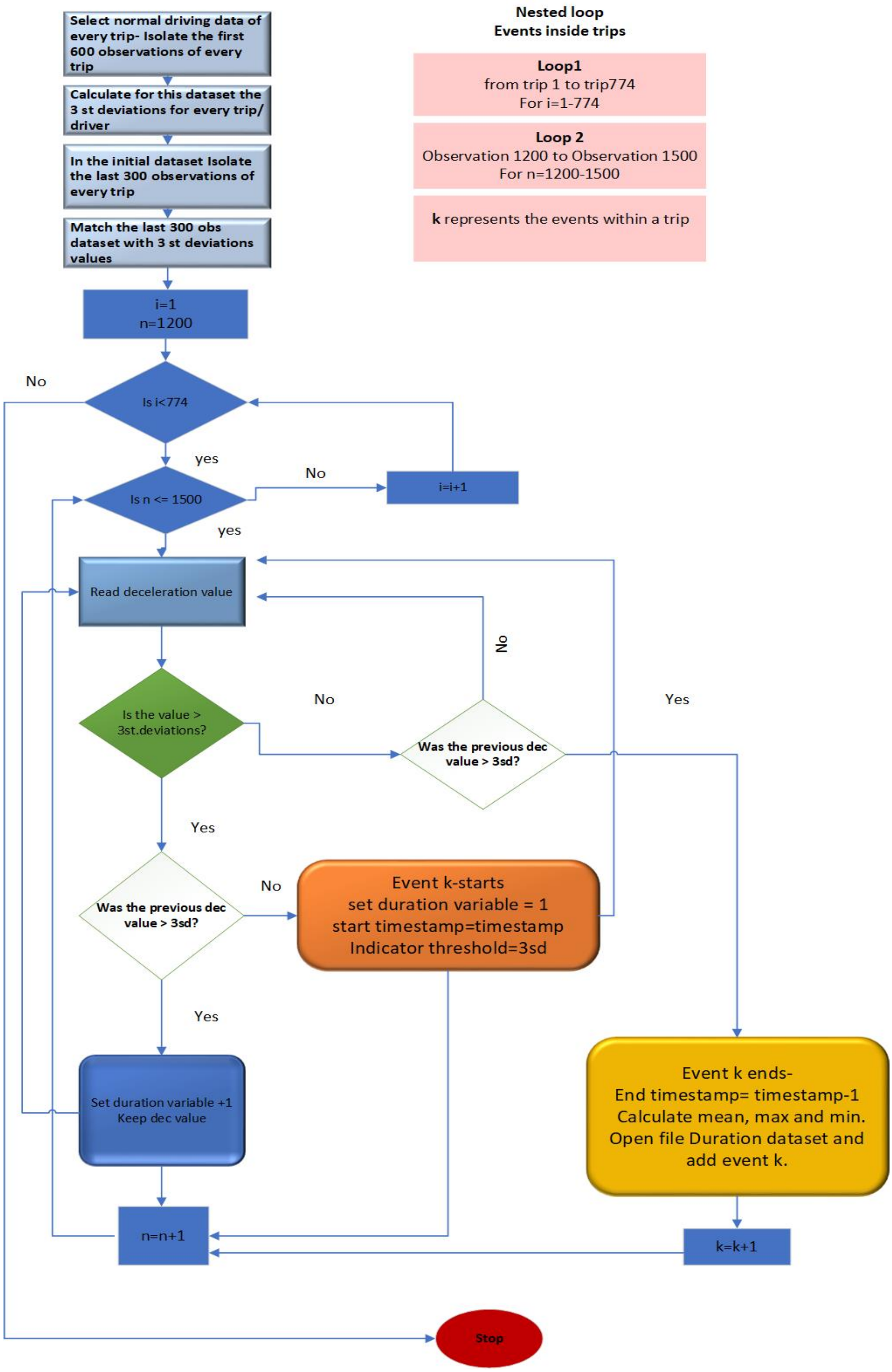


Figure 3.9 Deceleration Event duration identification algorithm

3.10.1.2 Extracting overlapping deceleration and yaw rate events - the algorithm of overlapping events' identification

After the investigation of the deceleration and yaw rate events separately, the overlapping events (events where both deceleration and yaw rate thresholds simultaneously occurred) were explored. The process that was followed to extract the overlapping events from the data is described in the steps below:

1. Apply Algorithm1 and Algorithm 2.
2. Identify timestamps where both deceleration and yaw rate values exceed 3sd deviations threshold (events simultaneously occurred).
3. Sort timestamps for both indicators in ascending order in the same vector/column.
4. Identify repeated timestamps within the same vector/column.
5. Determine which kind of event occurs before the first overlapping timestamp.
6. Find last overlapping timestamp.
7. Determine the kind of event occurred after last overlapping timestamps.
8. End.

Figure 3.10 visualises the algorithm used to extract the overlapping events and identify the order in which they occur.

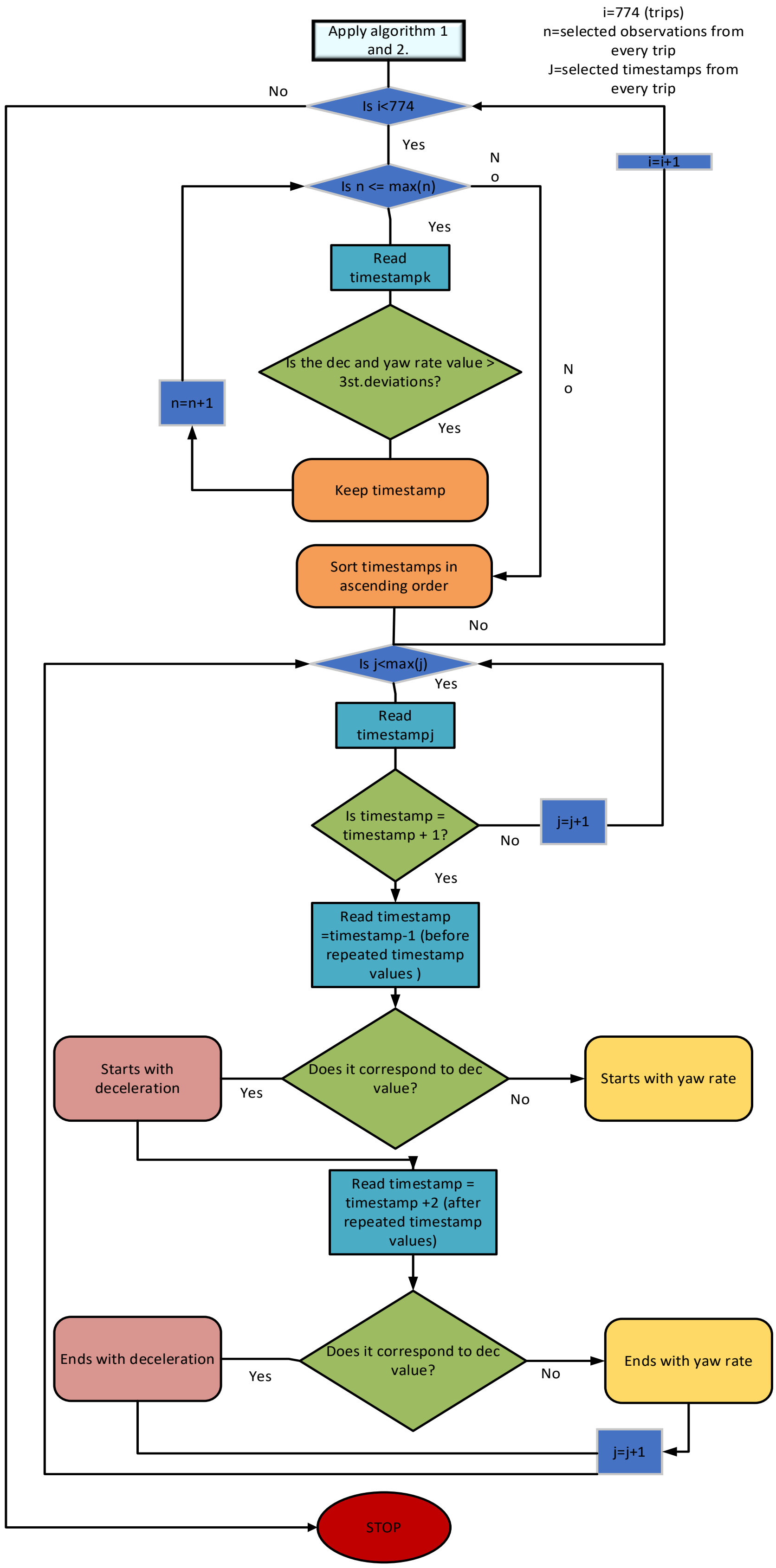


Figure 3.10 Algorithm of overlapping events

3.10.2 Descriptive analysis on event duration and kinematics

A thorough descriptive analysis was conducted including exploration of:

- **Duration of final and initial events**

The researcher assumes that the final event as it was defined (above 3 standard deviations for 1 s) is the critical one (crash or near-crash). Practically, the durations of each final no empty duration column, were gathered in one column, so a variable with the duration of the final critical events of all the trips has been created. On the contrary, events that were first detected during the time period of 30 seconds, were considered as initial events. The distribution of the events duration combined with the cumulative frequency can give a picture of the common event duration.

- **Kinematics of events prior event**

The distribution of the deceleration/yaw rate events means of the final events was explored (histogram and cumulative frequency), to be compared with the corresponding of the initial braking events.

- **Differences between crashes and near-crashes**

The event duration, the event deceleration/yaw rate mean, max and min, and also some demographic information is presented for events relative to crash and near crash trips.

- **Different percentile values of deceleration/yaw rate event mean and duration by incident type and severity**

Percentile values for events mean and duration are summarised for different event types (rear-end, striking, road departure, sideswipe) and event severity (crashes, near-crashes). 50th percentile representing the median could possibly be used to derive thresholds for emerging conditions. More specifically, the combination of the 50th percentile of the deceleration/yaw rate event mean and the 50th percentile of the event duration. As 50% of the deceleration/yaw rate event duration is less than a threshold value and 50% of the deceleration/yaw rate events means are below another threshold value, the concurrence of these two thresholds can be identical of a safety critical situation.

- **Categories of pre-crash behaviour and corresponding TTC**

The pre-crash behaviour has been categorised according to the number of deceleration /yaw rate events and characteristic examples have been plotted together with corresponding TTC values to visualise the relation between these indicators.

3.11 Statistical modelling (Objective 6)

TTC indicators did not show any explicit trend or pattern and thus, an in-depth analysis to explore the evolution of TTC values during the event sequence development and the factors that influence them would be constructive.

Crash prediction and hence, road safety has widely relied on statistical methods. Below, an introduction to Hierarchical Linear modelling gives the main features of the method and simultaneously, constitutes a justification for the specific model choice employed to further explore TTC. This multilevel regression modelling approach has been selected over other types of analysis due to its explanatory power as the focus of this analysis is not the forecasting but the identification of factors affecting TTC evolution. Nevertheless, Prais-Winsten AR(1) technique and latent growth modelling have been attempted with no significant difference or any improvement in the results.

3.11.1 Hierarchical Linear Modelling (HLM) or Multilevel mixed (fixed and random) effect modelling

Hierarchical levels of grouped data (data sorted into categories, classified) constitute a phenomenon that occurs regularly (Osborne, 2000). Analysis of hierarchical data is best performed by employing statistical techniques that take under consideration the hierarchy, such as Hierarchical Linear Modelling. Hierarchical Linear Modelling (HLM) is a complex form of ordinary least squares (OLS) regression that is utilised to analyse variance in the outcome variables when the predictor variables are at different hierarchical levels; for example, students in a classroom share variance due to their common teacher and common classroom. Before the development of HLM, hierarchical data was treated with fixed parameter simple linear regression techniques; nevertheless, as these techniques do not account for the shared variance, were inappropriate for this kind of analysis. HLM takes into account the shared variance in hierarchically structured data: The technique accurately estimates lower level slopes (e.g., student level) and their implementation in estimating higher-level outcomes (e.g., classroom level), (Hofmann, 1997).

While HLM is widespread across many domains, it is often used in the education, health, social work, and business sectors. As this statistical method was developed simultaneously within many fields, it is known by several names, including multilevel, mixed level, mixed linear, mixed effects, random effects, random coefficient (regression), and (complex) covariance components modelling (Raudenbush and Bryk, 2002). All these labels refer to the same advanced regression technique. As HLM simultaneously explores relationships within and between hierarchical levels of grouped data, it has proved to be more effective in considering the variance among variables at varying levels than other existing methods.

HLM can accommodate multiple continuous or discrete outcome variables in the same analysis (Raudenbush and Bryk, 2002) and the outcome variable of interest is always situated at the lowest level of the hierarchy (Castro, 2002). Each level-1 (X,Y) unit is identified by its level-2 cluster . Each level-2 cluster's slope is also identified and analysed separately. Using HLM, both the within-and between-group regressions are considered to relate the dependent and independent variable.

The analysis of nested data is suited by HLM as the latter describes the relationship between dependent and independent variables, by considering both level-1 and level-2 regression relationships. Except for the HLM's ability to estimate cross-level data relationships and unravel the effects of between- and within-group variance, it also requires fewer assumptions to be met in comparison with other statistical methods (Raudenbush and Bryk, 2002). HLM can cope with missing data, non-independence of observations, a lack of sphericity, small and/or discrepant group sample sizes, and heterogeneity of variance across repeated measures. Nevertheless, HLM depends upon large sample sizes to have adequate power especially when identifying effects at level-1. Higher-level effects are affected more by increases in groups than by increases in observations per group. Moreover, HLM takes out groups with missing data if they are at level-2 or above and accommodates missing data only at level-1.

3.11.1.1 Statistical features of HLM.

Two-level hierarchical data structures with continuous outcome (dependent) variables constitutes the most comprehensive but simple way to exhibit the statistical features of HLM. The notation that will be used here, has been employed by Raudenbush and Bryk 2002 (see Raudenbush and Bryk, 2002 for three-level models; see Wong and Mason, 1985 for dichotomous outcome variables). As mentioned before, hierarchical linear models permit the simultaneous exploration of the relationship within a given hierarchy and across the various levels. For this purpose, two models are employed: the first that depicts the relationship within lower level units, and another one that reflects how the relationship within lower level units differs between units (therefore correcting for the violations of aggregating or disaggregating data; Hofmann, 1997). This modelling technique can be implemented to any case where there are lower-level units nested within higher-level units and the lower-level units can be considered as individuals while the higher-level units as groups.

In two-level hierarchical models, different level-1 models are developed for each level-2 unit. These models show the effects in the context of a single group and called within-unit models (Gill, 2003). They take the form of simple regressions applied for each individual *i*:

$$Y_{ij} = \beta oj + \beta_{1j} X_{ij} + r_{ij} \quad (3.2)$$

where:

Y_{ij} = dependent variable measured for i^{th} level-1 unit nested within the j^{th} level-2 unit,

X_{ij} = value on the level-1 predictor,

βoj= intercept for the j^{th} level-2 unit,

β_{1j} = regression coefficient associated with the j^{th} level-2 unit, and

r_{ij} = random error associated with the i^{th} level-1 unit nested within the j^{th} level-2 unit.

Similarly to most statistical models, an important assumption of HLM is that any level-1 errors (r_{ij}) follow a normal distribution with a mean of 0 and a variance of σ^2 (see Equation 3.3; Sullivan, Dukes and Losina, 1999). This applies to any level-1 model utilising continuous outcome variables.

$$E\left(r_{ij}\right) = 0; \quad var\left(r_{ij}\right) = \sigma^2 \quad (3.3)$$

In the level-2 models, the level-1 regression coefficients (βoj and $\beta 1j$) are used as outcome variables and are related to each of the level -2 predictors. Level-2 models are also called between-unit models as they reflect the variability across multiple groups (Gill, 2003). The case of a single level-2 predictor will be modelled using Equations 3.4 and 3.5:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\ G_j\ + U_{0j}\ (3.4)$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}\ G_j + U_{1j};\ (3.5)$$

where:

β_{0j} = intercept for the j^{th} level-2 unit;

β_{1j}= slope for the j^{th} level-2 unit;

G_j = value on the level-2 predictor;

γ_{00} = overall mean intercept adjusted for *G;*

γ_{10} = overall mean intercept adjusted for *G;*

γ_{01} = regression coefficient associated with *G* relative to level-1 intercept;

γ_{11} = regression coefficient associated with *G* relative to level-1 slope;

U_{0j} = random effects of the j^{th} level-2 unit adjusted for *G* on the intercept;

U_{1j} = random effects of the j^{th} level-2 unit adjusted for *G* on the slope.

What discriminates HLM from a normal regression equation is that level-2 model presents two new terms (U_{0j} and U_{1j}) that are unique to it. Furthermore, the model developed would rely on the pattern of variance in the level-1 intercepts and slopes (Hofmann, 1997). For instance, if there was no variation in the slopes across the level-1 models, G_j would no longer be meaningful given that is equivalent across groups and would thus be removed from Equation 3.4 (Hofmann, 1997). Special cases of the two-level model Equations 3.2, 3.4 and 3.5 can be found in Raudenbush and Bryk (1992).

The assumption in the level-2 model (when errors are homogeneous at both levels) is that β_{0j} and β_{1j} have a normal multivariate distribution with variances defined by τ_{00} and τ_{11} and means equal to γ_{00} and γ_{10}. Furthermore, the covariance between βoj and β_{1j} (defined as τ_{01}) is equal to the covariance between U_{0j} and U_{1j}. As in the level-1 assumptions, the mean of U_{0j} and U_{1j} is assumed to be zero and level-1 and level-2 errors are not correlated. Finally, the covariance between U_{0j} and r_{1j} and the covariance of U_{1j} and r_{1j} are both zero (Sullivan et al., 1999). The assumptions of level-2 models can be summarized as follows (Raudenbush and Bryk, 2002; Sullivan et al., 1999):

$$
\begin{aligned}
&E(U_{0j}) = 0;\ E(U_{1j}) = 0 \\
&E(\beta oj) = \gamma_{00};\ E(\beta_{1j}) = \gamma_{01} \\
&var(\beta_{0j}) = var(U_{0j}) = \tau_{00}\ ;\ var(\beta_{1j}) = var(U_{1j}) = \tau_{11}\ ; \\
&cov(\beta_{0j}, \beta_{1j}) = cov(U_{0j}, U_{1j}) = \tau_{01}\ ; \\
&cov(U_{0j}, r_{1j}) = cov(U_{1j}, r_{1j}) = 0 \qquad (3.6)
\end{aligned}
$$

In order to allow for the classification of variables and coefficients regarding the level of hierarchy they affect (Gill, 2003), a combined model (i.e., two-level model; see Equation 3.7) is developed by substituting Equations 3.3 and 3.4 into Equation 3.2:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} G_j + \gamma_{11} G_j X_{ij} + U_{1j} X_{ij} + U_{0j} + r_{1j} \quad (3.7)$$

The combined model incorporates the level-1 and level-2 predictors (X_{ij} and G_j), a cross-level term ($G_j X_{ij}$) as well as the composite error ($U_{1j} X_{ij} + U_{0j} + r_{1j}$). Equation 3.7 is usually referred to as a mixed model because it includes both fixed and random effects (Gill, 2003). A comparison between Equation 3.7 and the equation for a normal regression (Equation 3.8) further pinpoints the uniqueness of HLM.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_n X_{in} + \varepsilon_i \quad (3.8)$$

As stated before, the HLM model presents two new terms (U_{0j} and U_{1j}) that enables the model to estimate error that normal regression cannot. In Equation 3.7, the errors are no longer independent across the level-1 units. The terms U_{0j} and U_{1j} show that there is dependency among the level-1 units nested within each level-2 unit. Furthermore, U_{0j} and U_{1j} may have different values within level-2 units, leading to heterogeneous variances of the error terms (Sullivan et al., 1999). This dependency of errors has important implications for parameter estimation.

3.11.1.2 <u>Estimation of effects</u>

Two-level hierarchical models include the estimation of three types of parameters. The first type of parameter is fixed effects, and these do not vary across groups (Hofmann, 1997). The fixed effects are represented by γ_{00} , γ_{01} γ_{11} and γ_{10} in Equations 4 and 5. The level-2 fixed effects could be estimated via the Ordinary Least Squares (OLS) approach, but as it requires the assumption of homoscedasticity, it is not a suitable estimation method. The accuracy of level-1 parameters are likely to vary across groups, therefore this assumption is violated in hierarchical models (Hofmann, 1997). The technique for fixed effects estimation is called a Generalized Least Squared (GLS) estimate. A GLS produces a weighted level-2 regression which secures that groups with more accurate estimates of the outcome variable (i.e., the intercepts and slopes) are allocated more weight in the level-2 regression equation (Hofmann, 1997).

The second type of parameter is the random level-1 coefficients (β_{0j} and β_{1j}) which are allowed to vary across groups (Hofmann, 1997). Hierarchical models provide two estimates for random coefficients of a given group: (1) computing an OLS regression for the level-1 equation representing that group; and (2) the predicted values of β_{0j} and β_{1j} in the level-2 model (Equations 3.3 and 3.4).

The estimation strategy that yields the most accurate values of the population slope and intercept for the given group is very important (Hofmann, 1997). HLM software programs use an empirical Bayes estimation strategy, which takes into account both estimation strategies by computing an optimally weighted combination of the two (Raudenbush and Bryk, 2002; Raudenbush et al., 2006). This strategy generates the best estimate of the level-1 coefficients for a particular group because it concludes in a smaller mean square error term (Raudenbush, 1988).

The final type of parameter estimation concerns the variance-covariance components which include: (1) the covariance between level-2 error terms [i.e., $cov(U_{0j}, U_{1j})\ or\ cov(\beta_{0j}, \beta_{1j})$ defined as τ_{01}]; (2) the variance in the level-1 error term (i.e., the variance of $r1$j denoted by σ^2); and (3) the variance in the level-2 error terms (i.e., the variance in U_{0j} and U_{1j} or βoj and $\beta 1j$ defined as τ_{00} and τ_{11} , respectively). When sample sizes are equal and the distribution of level-1 predictors is the same across all groups (i.e., the design is balanced), closed-form formulas can be used to estimate variance-covariance components (Raudenbush and Bryk, 2002). In reality, however, an unbalanced design is more likely to occur. In such cases, variance-covariance estimates are made using iterative numerical procedures (Raudenbush and Bryk, 2002). Raudenbush and Bryk (2002) recommend the following conceptual approaches to estimating variance-covariance in unbalanced designs: (1) full maximum likelihood; (2) restricted maximum likelihood; and (3) Bayes estimation.

3.11.1.3 TTC modelling

In the literature, Chapter 2, factors that affect TTC were reviewed. Here, a theoretical model is employed to express TTC affecting factors and set the basis for the Hierarchical Linear modelling process of the TTC values during the crash or near crash development. Furthermore, SHRP2 NDS data is utilised to practically investigate the factors affecting TTC values during the event sequence and especially the change of TTC values over time with the objective to identify the timepoint where the first change in the pattern of TTC values occurs.

According to the literature, the TTC values could be modelled by the following equation (Papazikou et al., 2019):

> "TTC = f (road characteristics, traffic conditions, vehicle type & kinematics, driver behaviour)"

The objective here is to develop a statistical model which can examine how some of these factors are related to TTC within the crash sequence and if there is a time point where a change in TTC values can be usefully detectable. Considering the structure and the nature of the data, Hierarchical Linear Modelling (or multilevel mixed effects linear regression modelling) has been chosen as suitable analysis tool.

More specifically, three types of factors are considered: (1) driver factors (e.g. age, gender, miles driven in the previous year), (2) factors relating to the traffic, vehicle and vehicle kinematics (e.g. flow, vehicle type, speed, yaw rate, lateral and longitudinal acceleration) and (3) factors related to the time within the event sequence (e.g. System timestamp, System timestamp squared). Figure 3.11 constitutes an annotated version of Figure 2.11 showing the variables that were finally considered in the modelling process.

Since each driver had one or two trips and each trip had multiple observations (1500), the TTC values during the event progression can be modelled using two level analyses i.e. the trip level and the observation level. More thoroughly, the research question involves a hierarchy of 2 levels. At the highest level of the hierarchy (level-2) are vehicle type variable and driver related variables such as gender and age while in the lowest level (level-1) are vehicle kinematics and time variables such as speed, yaw rate, accelerations and system time stamp. Level 1 variables, observations, are impacted by level 2 variables, the trips. More specifically, the observations are nested within each trip-file in the dataset. The outcome variable is measured at level 1 as

Hierarchical Linear Modelling (HLM) requires. The modelling process will reveal whether the driver's gender and age, vehicle type, longitudinal and lateral acceleration, yaw rate, speed and time influence TTC values in the event sequence development and whether a change to TTC values is early detectable in order to understand when a hazardous deviation starts. Figure 3.12 visualises the hierarchy structure of a three level mixed effect model, involving observations, trip and driver level.

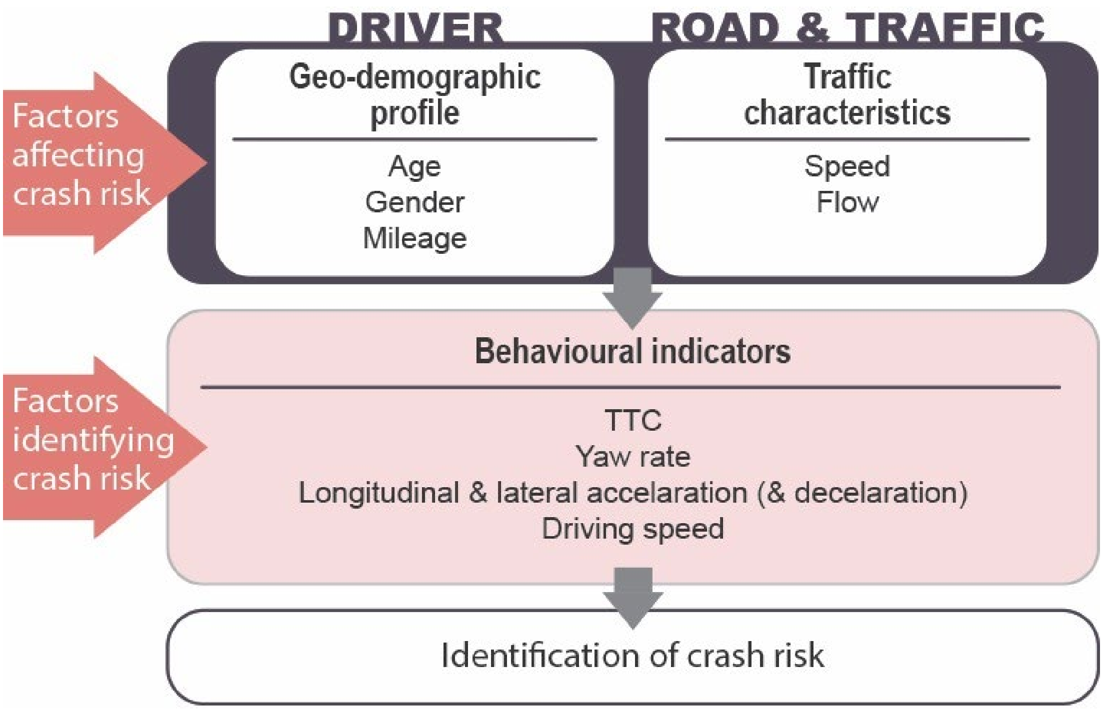


Figure 3.11 Annotated version of Theoretical framework for factors affecting or identifying crash risk

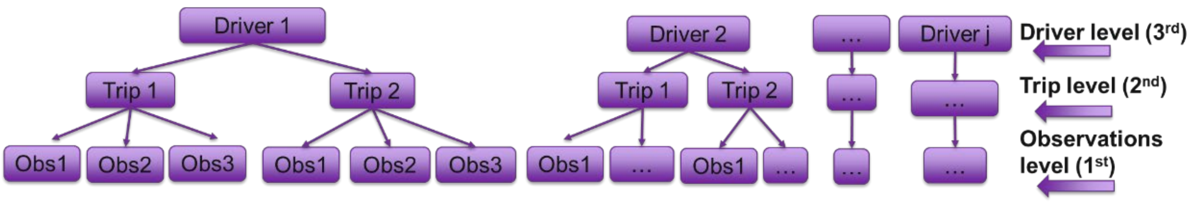


Figure 3.12 Three level mixed effect model hierarchy structure

Vehicle kinematics observations from the same driver may have common characteristics (i.e. within-cluster correlation) or there might be a variation between the observations from different vehicle types (i.e. between-cluster variation). The use of two level mixed effects linear regression model is suitable as it allows for dependency of the observations coming from the same driver and within the same trip and examines the variation of observations for different drivers and different trips by the same drivers. Moreover, it deals with the issue of consistency since not all the drivers have executed multiple trips. A two-level mixed-effects linear regression model can be developed for a single explanatory variable (x) as (Woltman et al., 2012):

Observations-level (level 1):

$$Y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij} \tag{3.9}$$

Trip-file level (level 2):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\, w_j + \, u_{0j}; \quad \beta_{1j} = \gamma_{10} + \gamma_{11}\, w_j + u_{1j}; \tag{3.10}$$

The composite equation can be expressed as:

$$Y_{ijk} = \gamma_{00} + \gamma_{10}\, x_{ij} + u_{1j}\, x_{ij} + \gamma_{11}\, w_j\, x_{ij} + \, \gamma_{01}\, w_j + u_{0j} + e_{ij} \tag{3.11}$$

In which Y_{ij} is the TTC values for ith level-1 observation nested within j[th] level-2 trip-file, x_{ij} is the value for the level-1 predictor, β_{0j} is the intercept for the j[th] trip-file, β_{1j} is the regression coefficient (slope) associated with the x_{ij} for the j[th] trip-file, e_{ij} is the random error associated with the i[th] observation (level-1) nested within the j[th] trip-file (level-2), w_j is the value on the level-2 predictor, γ_{00} is the overall mean intercept adjusted for w, γ_{10} is the overall mean intercept adjusted for w, γ_{01} is the regression coefficient associated with w relative to level-1 intercept, γ_{11} is the regression coefficient associated with w relative to level-1 slope, u_{oj} is the random effects of the j[th] trip-file adjusted for w on the intercept, u_{1j} is the random effects of the j[th] trip-file adjusted for w on the slope.

All random components are assumed to follow a normal distribution with a mean of zero and a constant standard deviation. Equation 3.11 represents a two-level mixed-effects linear regression model for a single explanatory variable, but this can be similarly extended for multiple explanatory variables. This model can be estimated using the restricted maximum likelihood (REML) estimation method. REML constitutes a particular form of maximum likelihood (ML) technique for estimating variance components while considering that the estimation of fixed effects results in loss in degrees of freedom (Patterson and Thompson, 1971).

3.12 Summary

This chapter reported the employed empirical and analytical methods for this research. Initially, the choice of the research approach (NDS) and the data utilised (SHRP2 NDS) was denoted while the correspondence between the methods and the objectives of this thesis were indicated in the research design section. Then, the methodology of quantifying normal driving and identifying critical thresholds was described, followed by the empirical process of generation of indicators for detecting deviations from normal driving conditions. Furthermore, the algorithms developed for data of interest extraction were demonstrated and the chapter ends with a detailed description of the multilevel mixed effects modelling technique that was employed to explore TTC during event sequence. In summation, Figure 3.13 presents the flowchart of the overall methodology of this thesis.

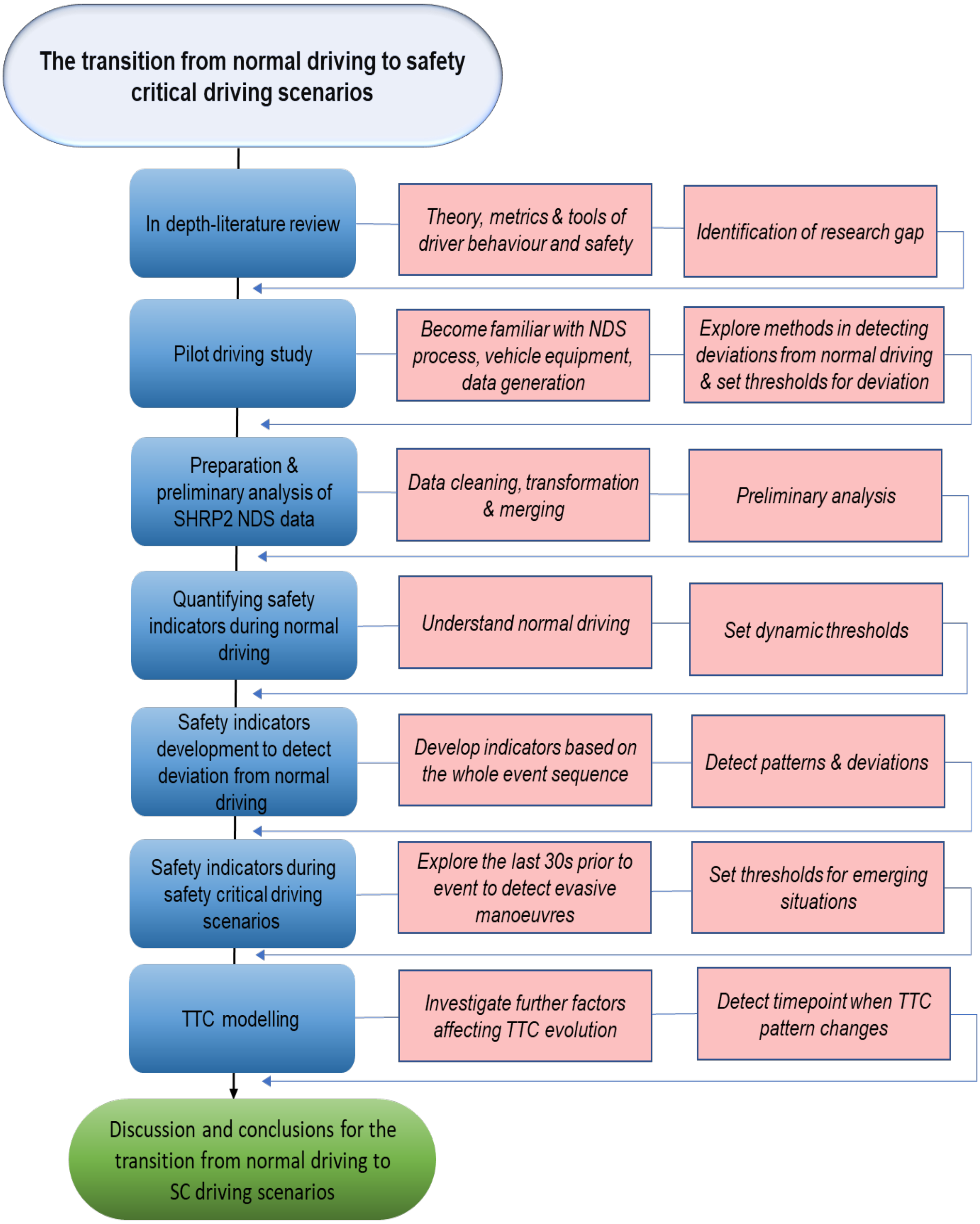


Figure 3.13 Thesis Methodology flowchart

4 Data preparation and preliminary analysis

4.1 Introduction

Once the SHRP2 NDS data was received, several steps needed to be completed in preparation for data modelling. These steps constitute the initial part of the data mining process and they are crucial for the quality and the depth of the study.

Pre-processing is a process of cleaning selected data for better quality. Cleaning, manipulating and transforming the data was a fundamentally significant part of this research and constituted a really challenging task due to the structure, the complexity, the size and the continuous nature of the data.

It should be mentioned though, that the radar data was not received raw but heavily post-processed by the VTTI. This process involved adjusting timestamps, removing erroneous data points and targets, smoothing the data to reflect what was physically possible, classifying the data into meaningful categories and reorganising the data into a more intuitive format.

In this chapter, data manipulation, data pre-processing and the relevant challenges are presented. The chapter describes all the first steps of analysis followed regarding the drivers, the events, and other variables in a thorough way and several diagrams illustrate the process.

4.2 Data Understanding

Data comprehension takes into account data requirements. This stage includes the exploration and the description of the data, as well as, the verification of data quality. With data exploration the researcher tries to gain a better understanding of the data set and this procedure helps to filter and redirect the discovery process (Olson et al., 2008). Basic data exploration can be done by viewing summary statistics (which also includes the visual display of categorical variables). Below the most important results of the SHRP2 NDS data exploration are presented.

4.2.1 Initial data form

SHRP2 data were obtained from VTTI in the form of four folders: (i) Time series (i.e. highly disaggregated kinematic data for the ego-vehicle), (ii) Demographic questionnaire of the driver, (iii) Event detailed and (iv) Event ID data key. Each folder, apart from the data, contains an extra file that provides all the variables' names and the relevant clarification. The times series data folder consists of 5813 files representing the SHRP2 NDS events.

4.2.2 Time series data

5,813 files were received totally for crashes and near crashes events. From them, 1,485 were events consisting of "fast" type tables. The "fast" tables are only generated when the instrumented vehicle gets into a crash/near-crash situation thereby triggering certain kinematic thresholds. The fast tables are then produced resulting from activation of higher data rate sensors to reflect fast-evolving kinematic dynamics which would not have been captured with the normal data rate sensors. No data is available for other periods except for the time window that it is activated. Therefore, the "fast" tables include the data of approximately 5-6 seconds (in some cases much more) after accelerometer kinematic spike detection. Figure 4.1 presents the whole crash sequence (2.5 minutes before the event) at the frequency of 10 Hz, therefore 10 observations per second (150 seconds), while Figure 4.2 constitutes the plot of the corresponding fast type data of this event. Here, the frequency is at 100Hz and 30 seconds have been recorded (3000 observations).

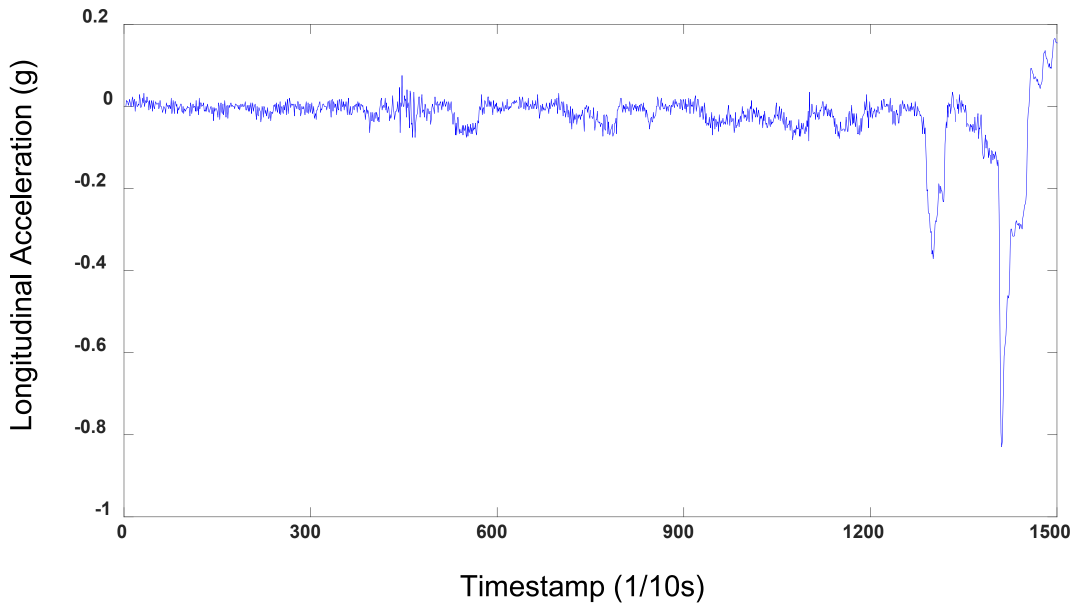


Figure 4.1 Longitudinal Acceleration during crash sequence (10Hz)

(Case number:152207960)

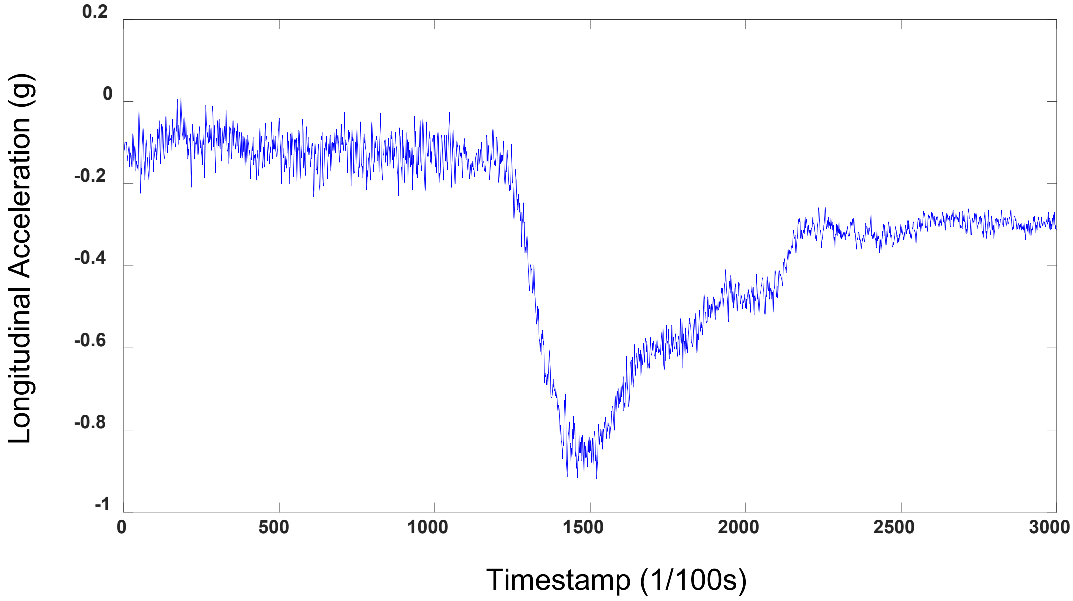


Figure 4.2. Longitudinal Acceleration during crash sequence (100Hz)

Moreover, from a first visual exploration it was obvious that some files had small size. This happens because some events occurred in the start of the trip, so not adequate data has been gathered before. An example of this type of event is illustrated in the Figure 4.3 below.

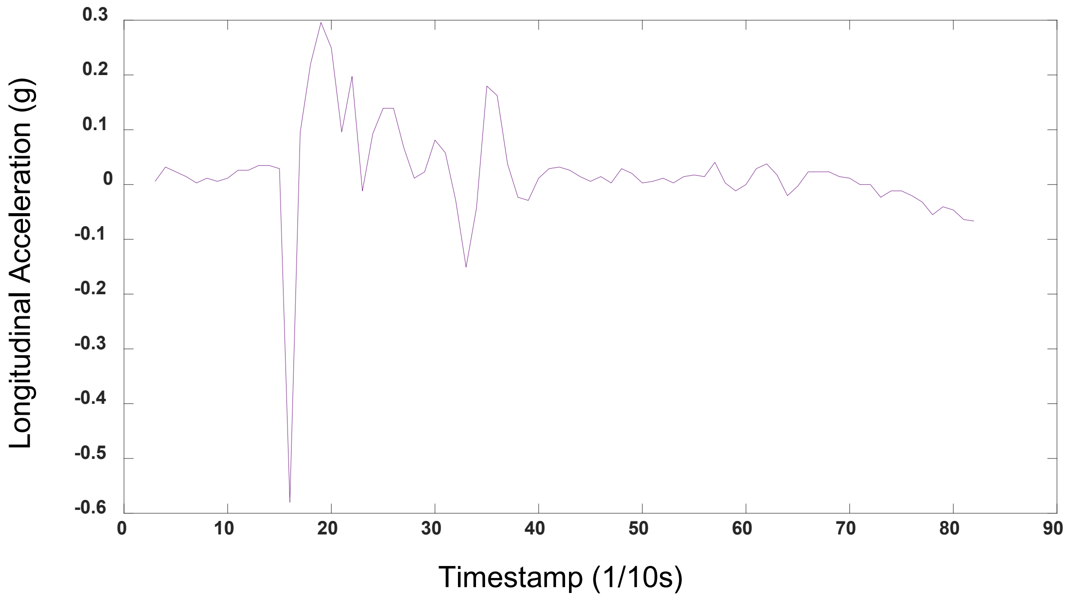


Figure 4.3. Longitudinal acceleration during a trip where the event occurred in the first seconds.

(Case number:15221171)

Figure 4.3 indicates that 1.5 seconds after the trip had started, a harsh braking occurred, denoting a near-crash or minor crash event at the very start of the driving process.

4.2.3 Events

A separate file containing information about the events in general was also received. From this, it is possible to investigate several aspects of the events, such as their severity level, the type, the liability, the locality, the weather, the traffic and lightning conditions. Table 4.1 displays the incident type by event severity.

Table 4.1. Incident type by event severity

Events	**Event severity**		
INCIDENT TYPE	**Crash**	**Near-Crash**	**Grand Total**
Animal-related	3.3%	4.7%	4.3%
Backing into traffic	1.1%	1.1%	1.1%
Backing, fixed object	1.2%	0.0%	0.4%
Opposite direction (head-on or sideswipe)	0.1%	1.5%	1.0%
Other	4.9%	2.2%	3.1%
Pedalcyclist-related	0.1%	1.2%	0.8%
Pedestrian-related	0.1%	2.9%	1.9%
Rear-end, striking	7.6%	50.4%	35.9%
Rear-end, struck	5.5%	0.6%	2.3%
Road departure (end)	6.4%	0.1%	2.2%
Road departure (left or right)	63.5%	2.8%	23.3%
Sideswipe, same direction (left or right)	1.4%	13.1%	9.1%
Straight crossing path	1.2%	2.9%	2.3%
Turn across path	1.6%	5.5%	4.2%
Turn into path (opposite direction)	1.2%	4.6%	3.5%
Turn into path (same direction)	0.6%	6.6%	4.6%
Unknown	0.1%	0.0%	0.0%
Grand Total	100% (1464)	100% (2865)	100% (4329)

Road departure crashes constitute almost the 70% of the total number, the rear-end 13.1%, leaving only a 17% for all the other incident types. For near-crashes, rear-end events are the most prevalent with 51% while sideswipe and turning into path incidents are following with 13.1% and 11.2% respectively.

Table 4.2 exhibits the event severity by weather and lighting conditions.

Table 4.2. Event severity by weather and lighting conditions

Events		**Lighting**					
EVENT SEVERITY	**Weather**	**Darkness lighted**	**Darkness, not lighted**	**Dawn**	**Daylight**	**Dusk**	**Grand Total**
Crash	Fog	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%
	Mist/Light Rain	0.3%	0.0%	0.0%	1.4%	0.1%	1.8%
	No Adverse Conditions	5.1%	1.3%	0.3%	21.6%	0.9%	29.2%
	Other	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Raining	0.6%	0.1%	0.0%	1.1%	0.0%	1.9%
	Snow/Sleet and Fog	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Snowing	0.1%	0.1%	0.0%	0.5%	0.0%	0.7%
Crash Total		6.2%	1.5%	0.5%	24.6%	1.0%	33.8%
Near-Crash	Fog	0.1%	0.0%	0.0%	0.1%	0.0%	0.2%
	Mist/Light Rain	0.6%	0.1%	0.0%	2.3%	0.1%	3.2%
	No Adverse Conditions	8.0%	1.5%	0.5%	47.2%	1.6%	58.8%
	Rain and Fog	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Raining	0.9%	0.1%	0.1%	2.3%	0.0%	3.4%
	Sleeting	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Snow/Sleet and Fog	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
	Snowing	0.1%	0.0%	0.0%	0.4%	0.0%	0.5%
Near-Crash Total		9.7%	1.8%	0.6%	52.3%	1.8%	66.2%
Grand Total		15.9%	3.3%	1.1%	76.9%	2.8%	100.0% (4329)

68.8% and 13.1% of the events happened during daylight or darkness with lighted street conditions respectively, and under no adverse weather. There were also a considerable number of events (10.3%) occurred during mist/light rain and rain while there were a few during snowing (1.2%).

As can be seen in Figure 4.4 and Figure 4.5, in terms of traffic conditions, most of the events occurred in free flow or slightly restricted flow and regarding the locality, the business/industrial category with moderate residential and interstate one seem to present more crashes in the sample.

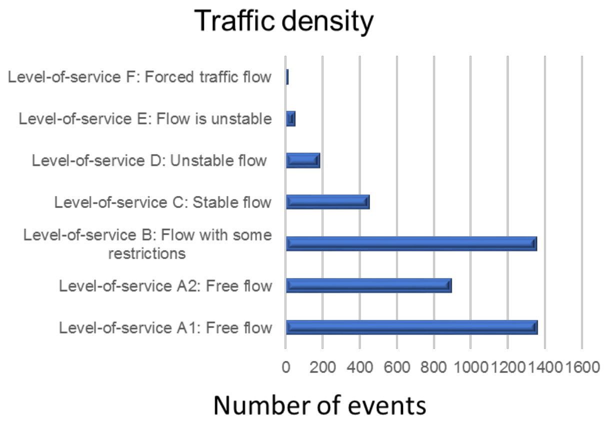


Figure 4.4. Traffic density **Figure 4.5. Locality**

In summary, there are 1464 crashes and 2865 near-crashes in the dataset and most of them are rear-end striking or road departure events, they occurred in non-adverse weather and free flow conditions, during daylight, in a business industrial region and the driver of the ego-vehicle seems to be the responsible.

4.2.4 Drivers

The dataset includes 1830 drivers, slightly more than half of them are females and the age groups have a range from 16-24 up to 95-99 with the majority of the drivers to be less than 25 years old. Moreover, 50% of them have a 10.000 annual mileage (drive over 10,000 miles a year).

Most of the vehicles are private cars with no advanced technology, with vehicle models dating from 1987 to 2013. Table 4.3 exhibits the drivers age group by gender.

As can be observed, the sample is biased towards young drivers that constitute almost the half of it. There is also a slight overrepresentation of females.

Table 4.3. Drivers by age groups and age

Drivers		**Gender**		
Age Group	N/A	Females	Males	Grand Total
N/A	0.1%	0.4%	0.3%	0.8%
16-19	0.0%	10.9%	9.9%	20.8%
20-24	0.0%	14.5%	11.2%	25.7%
25-29	0.0%	4.5%	3.6%	8.1%
30-34	0.0%	2.2%	2.2%	4.5%
35-39	0.0%	1.6%	1.6%	3.2%
40-44	0.0%	2.0%	1.5%	3.5%
45-49	0.0%	1.8%	2.2%	4.0%
50-54	0.0%	1.9%	2.0%	3.8%
55-59	0.0%	2.0%	1.6%	3.6%
60-64	0.0%	1.7%	1.6%	3.4%
65-69	0.0%	2.2%	2.2%	4.5%
70-74	0.0%	1.7%	1.9%	3.6%
75-79	0.0%	2.1%	3.0%	5.1%
80-84	0.0%	2.0%	1.9%	3.8%
85-89	0.0%	0.5%	0.8%	1.3%
90-94	0.0%	0.0%	0.2%	0.2%
95-99	0.0%	0.1%	0.1%	0.1%
Grand Total	0.1%	52.1%	47.8%	100.0%

**N/A:Not available*

Drivers’ involvement in crashes and near crashes was investigated and Table 4.4 presents the kind of the events (event severity) and how many drivers had involved.

Table 4.4. Events by drivers (%)

	Near-crashes			
Crashes	**0**	**1**	**2**	**3+**
0		28.6	9.4	5.2
1	18.9	10.5	5.3	5.4
2	4.0	2.4	1.8	2.6
3+	1.6	1.1	0.8	2.5
Total	24.5%	42.6%	17.3%	15.7%

The interesting point is that there is a 2.5% of the drivers that have involved to more than 3 crashes and near crashes, subsequently to more than 6 events, while 43.2% avoided a crash successfully at least once. Furthermore, 24.5% involved only in crash events.

4.2.5 Trips

The file contains aggregated data from 4171 trips that correspond to the events received. Here it should be noted that some trips involve more than one events. In this file, time spent between specific speed bins, time where TTC or Time headway present specific values, times that the indicators were activated, trip duration spent in different types of roads and many other details about the entire trips can be found. In the framework of this thesis though, this information has not been finally considered.

4.3 Data Preparation

In order for the data to be ready for analysis and the model development, it needs to be cleaned and built into the suitable form. At the end of this stage, data cleaning and data transformation should be completed. Figure 4.6 visualises the generic steps undertaken for data reduction.

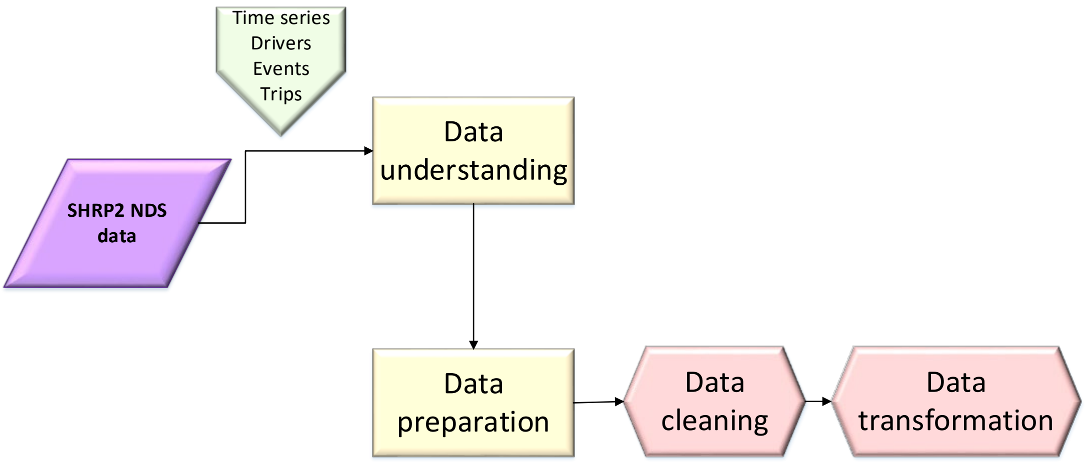


Figure 4.6. Data manipulation

4.3.1 Data Cleaning and transformation

Data Cleaning attempts to identify missing, inconsistent, or mistaken values or smooth the noise of the data. The exploration process helps as graphical tools can provide a picture of distributions, and statistics such as maxima, minima, mean values and makes easier any discrepancy detection.

Data transformation consists of the selection, creation and transformation of the data variables towards the model construction process. Based on the discoveries in the exploration phase again and on the research purposes, new variables can be introduced, or variables can be narrowed down to the most significant ones. Here it is worth to mention that different datasets can be merged or aggregated in order to exploit and get all the suitable information (data integration).

In this stage, some decisions were taken. The 1,484 “fast type” events were dropped as they were not adding anything at least at this stage of the research. It was decided that a frequency of 10 Hz already provides all the necessary information. Data has been, later in the analysis process, aggregated in the temporal resolution of second to check for sensitivity to data resolution and the results proved to be similar.

Moreover, 664 event files (trips where the event occurred in the first seconds) were also dropped, as they had less than 1500 observations and this inconsistency would probably cause problems during the processing of the data. Furthermore, as it was mentioned before, the times series data folder consists of 5813 files representing the SHRP2 NDS events. After the fast type tables and the events with less than 1500 observations were dropped, the next step was to combine all the remaining events in one single file. Therefore, a large time series data file of 3,665 events was created. The intention was to develop an even larger file inserting all the information about the events and the drivers, as well. The next action was to search for linking variables that would help in the data matching. These were found in the Event ID Key data file.

As the anonymous Demographic questionnaire table and Event ID data key table both include the Participant ID variable, they were the first datasets to be merged. Then, using the File ID as a common point, the Time series table and Event detailed table were also merged into the initial data table to form an aggregated dataset for analysis. Figure 4.7 presents the data aggregation process.

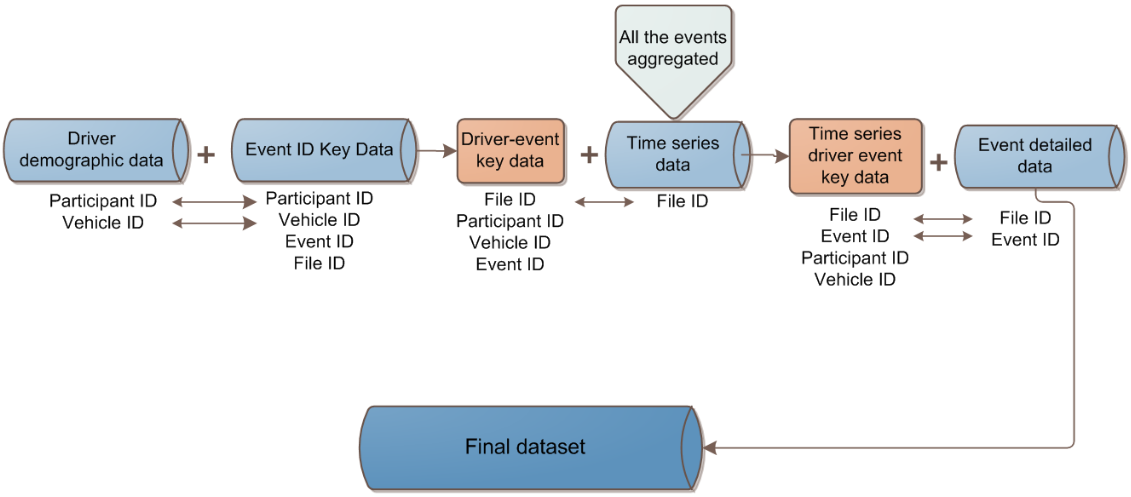


Figure 4.7. SHRP2 Data aggregation

The fact that the File ID was not unique for all the cases, in other words, in the same trip more than one events occurred, provoked problems to the matching process. Therefore, for these trips, only one of the events was kept for the data merge and the further analysis.

After the aggregated dataset was created, familiarisation with the new data structure was needed along with an extra data quality check. Exploration of the recently created dataset reveals that there is a number of problematic events where the timestamp variable exhibited inconsistences or there were dubious values in a few other variables. VTTI confirmed by checking the relevant video data that there were some cases where an error occurred during the data export or cases where a slight "ding" in the data was caused by vehicle hitting lane divider/markers or a pothole in the middle of the intersection or cases where simply the instruments needed recalibration. These events were excluded from the sample which at this moment was at 3596 events.

➢ ***<u>Calculating TTC indicator from existing variables:</u>***

TTC is a continuous variable and it can be calculated for any moment as long as the road users are on a collision course. TTC is not provided as a variable in the SHRP2 NDS data that was received. There are however a large number of relevant variables that could be employed. More specifically, the sensor platform within an ego-vehicle is capable of simultaneously tracking up to eight different targets that the ego-vehicle encounters within its radar field of view (range ~ 250m). Each of the targets is individually tracked as Track 0 through 7. With the raw radar variables, these tracked objects can sometimes switch across different tracks making the calculation of TTC difficult. To resolve this, a post-processing method was developed by VTTI in order to ensure that that the same target is being tracked consistently. This is to identify cases where the target vehicle is moving to different lanes while in front of the ego-vehicle or if another new vehicle comes into the radar's field of view. These post-processed range and relative velocity values were used in this study to calculate TTC for which the following process is developed and adopted:

1. Identify whether the ego-vehicle and a target vehicle are stationary;
2. Determine the lead target vehicle from the distances between target vehicles and the front bumper of the ego-vehicle, projected onto the x-axis (longitudinal) of ego-vehicle. The target vehicle with the smallest distance was identified as the lead vehicle;
3. TTC was calculated only if the lead vehicle was identified to be in the same lane of the ego-vehicle. This results in the calculation of minimum TTC (min TTC). The relevant variable (i.e. Lane in the NDS data) has five options coded as follows:

    2: there exists at least one lane to the right of the ego-vehicle that separates the ego-vehicle from the target;

    1: the target is in the adjacent lane to the right of the ego-vehicle vehicle;

    0: the target is in the same lane as the ego-vehicle;

    -1: the target is in the adjacent lane to the left of the ego-vehicle vehicle;

    -2: there exists at least one lane to the left of the ego-vehicle that separates the ego-vehicle from the target.

It should be noted that oncoming targets are not classified into lanes and these targets were given null values for "Lane"

4. Finally, the travel direction of the lead vehicle was identified and TTC was calculated only if the travel direction of the lead vehicle is known or the target is traveling in the same direction in relation to the ego-vehicle at the time of first detection and also at all other times that the object is being tracked. It

should be noted that the direction of the target was not known for about 70% of the cases in the sample data.

After completing the above process, a new dataset has been created containing only 139,914 observations (about 13% of the total observations initially obtained from the VTTI; these observations represent 1,033 trips, 683 drivers and 689 vehicles) suitable for calculating TTC. This is calculated as follows:

$$TTC_{ld} = \frac{d}{V_t - V_e} = \frac{d}{\Delta\vartheta} \tag{4.1}$$

In which: TTC_{ld} is the TTC between the ego-vehicle and the lead target vehicle travelling in the same lane and direction in relation to the ego-vehicle, V_t is the speed (in m/s) of the lead target vehicle, V_e is the speed of the ego-vehicle, d is the distance (in m) between the lead target vehicle and the front bumper of the ego-vehicle, projected onto the x-axis (longitudinal) of ego-vehicle, $\Delta\vartheta$ is the x-axis (longitudinal) component of range rate (relative velocity) between the lead target and the ego-vehicle. $\Delta\vartheta$ can be directly obtained from the SHRP2 NDS data.

The key values of TTC derived from the dataset are shown in Figure 4.8. An examination of the calculated TTC values along with the corresponding *d* and $\Delta\vartheta$ revealed that the primary reasons for some small values of TTC relate to: (i) low values of d (when *d*<2m), perhaps representing the scenarios when both the ego-vehicle and the target vehicle about to stop at a traffic light or a junction and (2) $\Delta\vartheta$ is large with a relatively small *d* value; this represents the scenario when the speed of the target vehicle is higher than that of the ego-vehicle at a reasonable space distance between them. The relationship between them is presented in Figure 4.8.

At a particular instant, two vehicles are actually considered to be in a collision course if the speed of the ego-vehicle (i.e. the following vehicle) is higher than the speed of the lead target vehicle i.e. $\Delta\vartheta < 0$. Therefore, the conditions applied to obtain reliable values for TTC are: (i) *d* is larger than 2m and (ii) relative speed ($\Delta\vartheta$) less than 0.

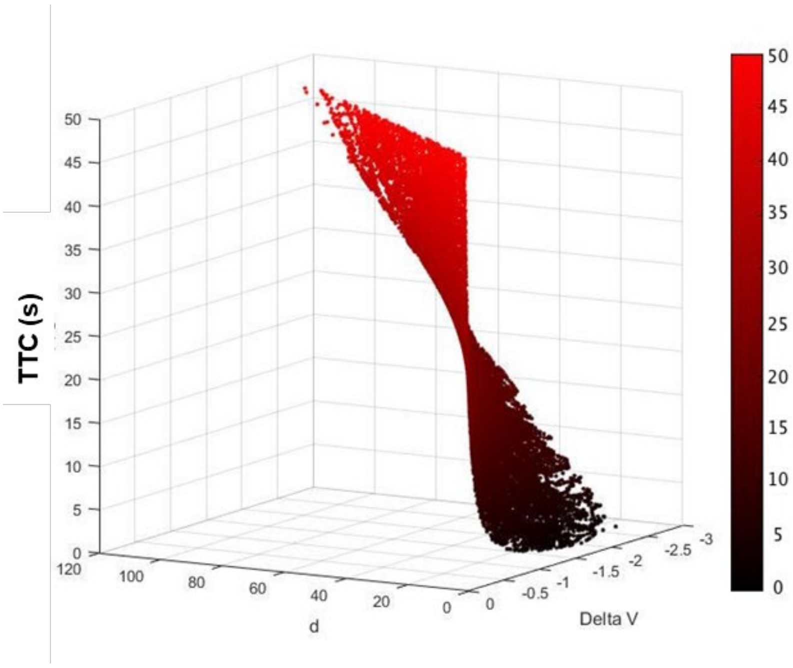


Figure 4.8. Relationship between d, $\Delta\vartheta$ and TTC

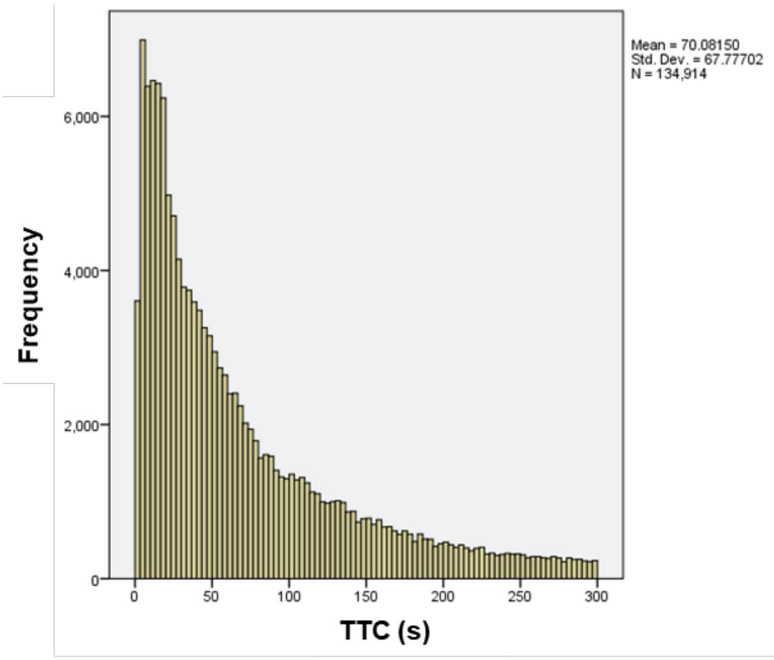


Figure 4.9. Distribution of TTC (when TTC<300)

This results in a total of 49,739 (only 4.6% of the original data obtained from the VTTI) valid TTC observations which have then employed for further analysis. Some of the TTC values are very large (due to very small $\Delta\vartheta$ and relatively large d) indicating that they are in a collision course.

With further exploration of TTC variable, it was noticed that not only there were many events with few TTC values, but also there were events that presented no TTC values at all. In particular, 1929 events (53.6%), present no TTC values at all while 86.6% of them present below 100 TTC values out of 1500 observations. Missing values could mean: i) there is no vehicle in front or any vehicle within the radar line-of-sight, or ii) the tracked vehicle is out of radar range or moved to another lane. In order for the analysis to be fruitful though, there is a need for as many (and valid) TTC values per event as possible. To resolve this issue, the algorithm used for the TTC calculation was questioned and as it was revealed, another calculation approach should be employed. More specifically, the initial algorithm set many criteria to be met in order for TTC to be calculated and this resulted in a dataset with few TTC values that could not be exploited properly. TTC recalculation process involved the test of many algorithms until the most appropriate to be found. This is translated into an algorithm that provides not only valid TTC values, but also many cases per event. The initial restrictions were gradually lifted and TTC was eventually given by the following equation:

$$\mathrm{TTC} = \frac{d}{|\Delta\upsilon|}\ , \quad \Delta\upsilon < 0 \tag{4.2}$$

In which: TTC is the TTC between the ego-vehicle and the closest target vehicle, d is the distance (in m) between the target vehicle and the front bumper of the ego-vehicle, projected onto the x-axis (longitudinal) of ego-vehicle and $\Delta\upsilon$ is the x-axis (longitudinal) component of range rate (relative velocity) between the target and the ego-vehicle (in m/sec). $\Delta\upsilon$ can be directly obtained from the data.

The new calculation algorithm resulted in a TTC variable with increased number of cases. Now, only 14.1% of the events present zero TTC values while 32% of them display more than 1000 values out of 1500 observations. The histograms below, Figure 4.10 and Figure 4.11 show how much the situation altered after the recalculation of TTC.

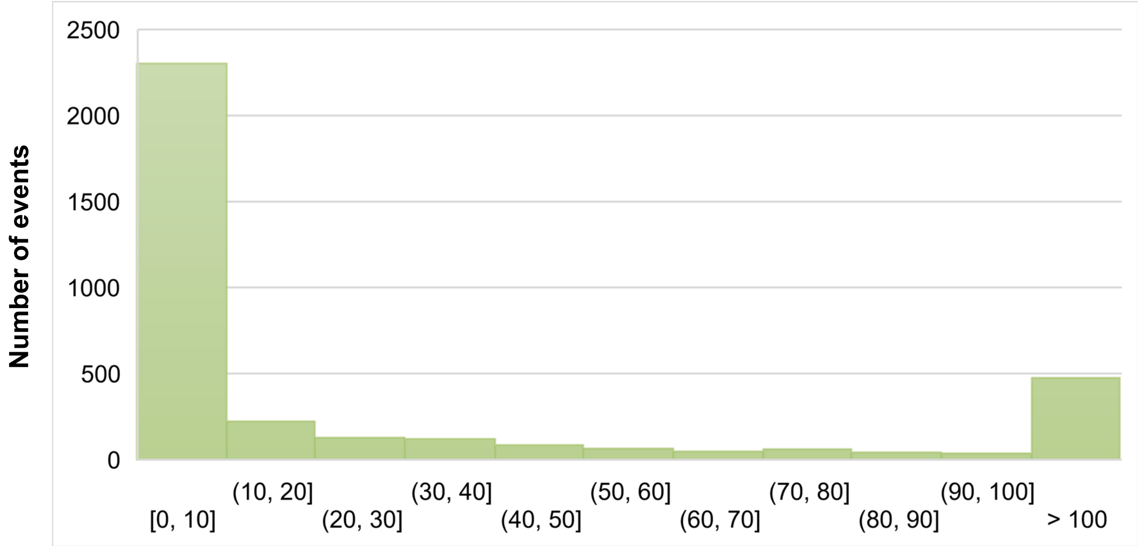


Figure 4.10. Frequencies of TTC values (s) per event (initial calculation)

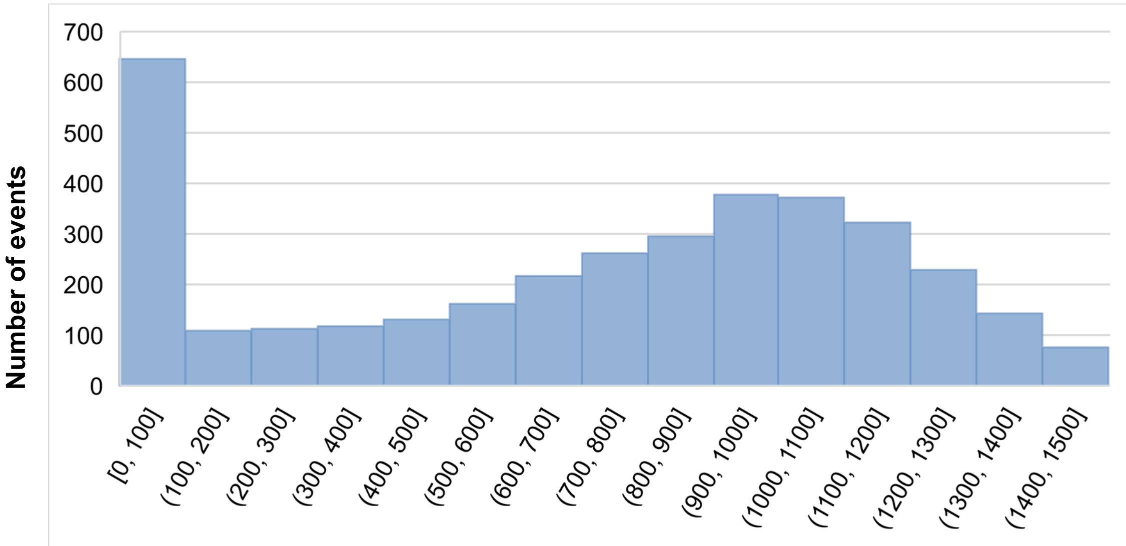


Figure 4.11. Frequencies of TTC values (s) per event (final calculation)

For the final dataset, only the events with more than 1100 TTC values (out of 1500) were chosen (774 events). This was decided with the criterion to have at least the 70% (73.33% eventually) of the TTC values per event. Despite this choice and that the final algorithm generated the most TTC values possible, there were still a lot of “missing values” per event that could cause problems in the next phases of the analysis and modelling. Therefore, another decision was taken, to fill in these values after exploring the replacing missing values techniques.

The available methods were series mean, the median and mean of nearby points, linear interpolation and linear trend at the point. All of them were tried out and the descriptive statistics were explored in the filled in variable to check if TTC still reserves its validity and if there are any mentionable changes introduced. These were some of the criteria for the method that was finally chosen. Table 4.5. below exhibit the descriptive statistics for TTC variable before and after the imputation of the missing values.

Table 4.5. Descriptive statistics before and after imputation

Descriptive Statistics						
	N	Min	Max	Mean	Std. Deviation	Variance
initial TTC	961713	0	16056368.6	235.4	24693.9	609788555.2
LINEAR TREND (TTC)	1161583	0	16056368.6	235.3	22469.2	504864258.8
SERIES MEAN (TTC)	1161583	0	16056368.6	235.4	22469.2	504864031.1
LINEAR INTERPOLATION(TTC)	1161582	0	16056368.6	583.5	40090.2	1607225862.3
MEAN OF NEARBY POINTS (TTC,SPAN2)	1161582	0	16056368.6	442.1	28839.7	831725822.8
MEDIAN OF NEARBY POINTS (TTC,SPAN5)	1161582	0	16056368.6	222.9	22558	508863234.5
MEDIAN OF NEARBY POINTS (TTC, SPAN15)	1161582	0	16056368.6	205.8	22469.4	504875463.5
MEDIAN OF NEARBY POINTS (TTC,SPAN10)	1161582	0	16056368.6	214.1	22531.4	507662556.6

The minimum and maximum values are not affected to any of the methods, but the variance seems to change considerably to some of them, e.g. linear interpolation or mean of nearby points. As it is observed linear trend at the point and series mean methods give almost the same mean and similar standard deviation with the initial one. However, further exploration of each method's influence in the variable values has been done and the median of nearby points (span 10) was eventually selected as the most appropriate method. Figure 4.12 shows the TTC values before and after the imputation of missing values with the method of median of nearby points (span 10) for only one event selected (presents 397 missing values).

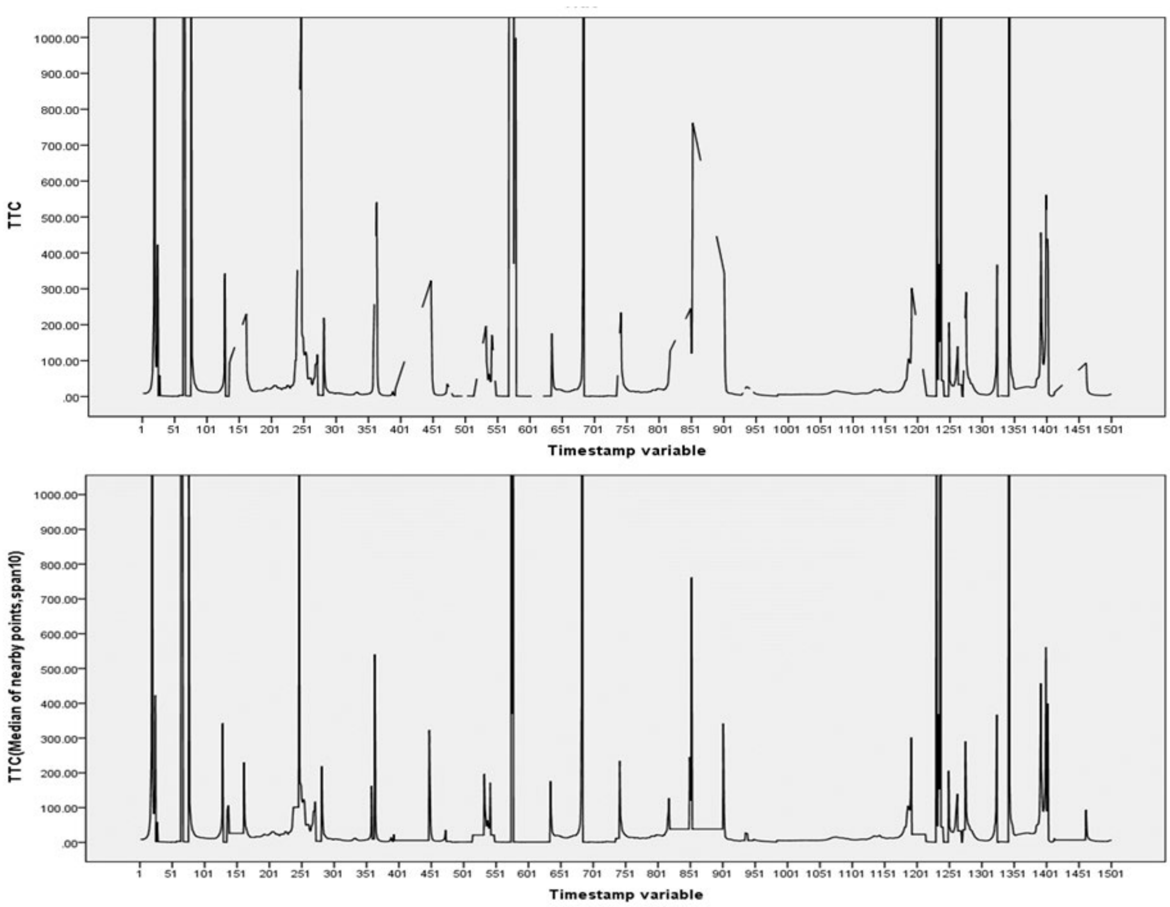


Figure 4.12 TTC values before and after the imputation of missing values

To become clearer, a closer look into the specific parts of the missing values was taken. More specifically, a smaller part of the diagram representing time where many missing values occurred is presented below captured for the different methods (Figure 4.13). Observing, hence, the several imputation methods, it was decided that the median of nearby points is the most suitable as it keeps the trend of the variable without inserting any abnormalities, extreme values or knees.

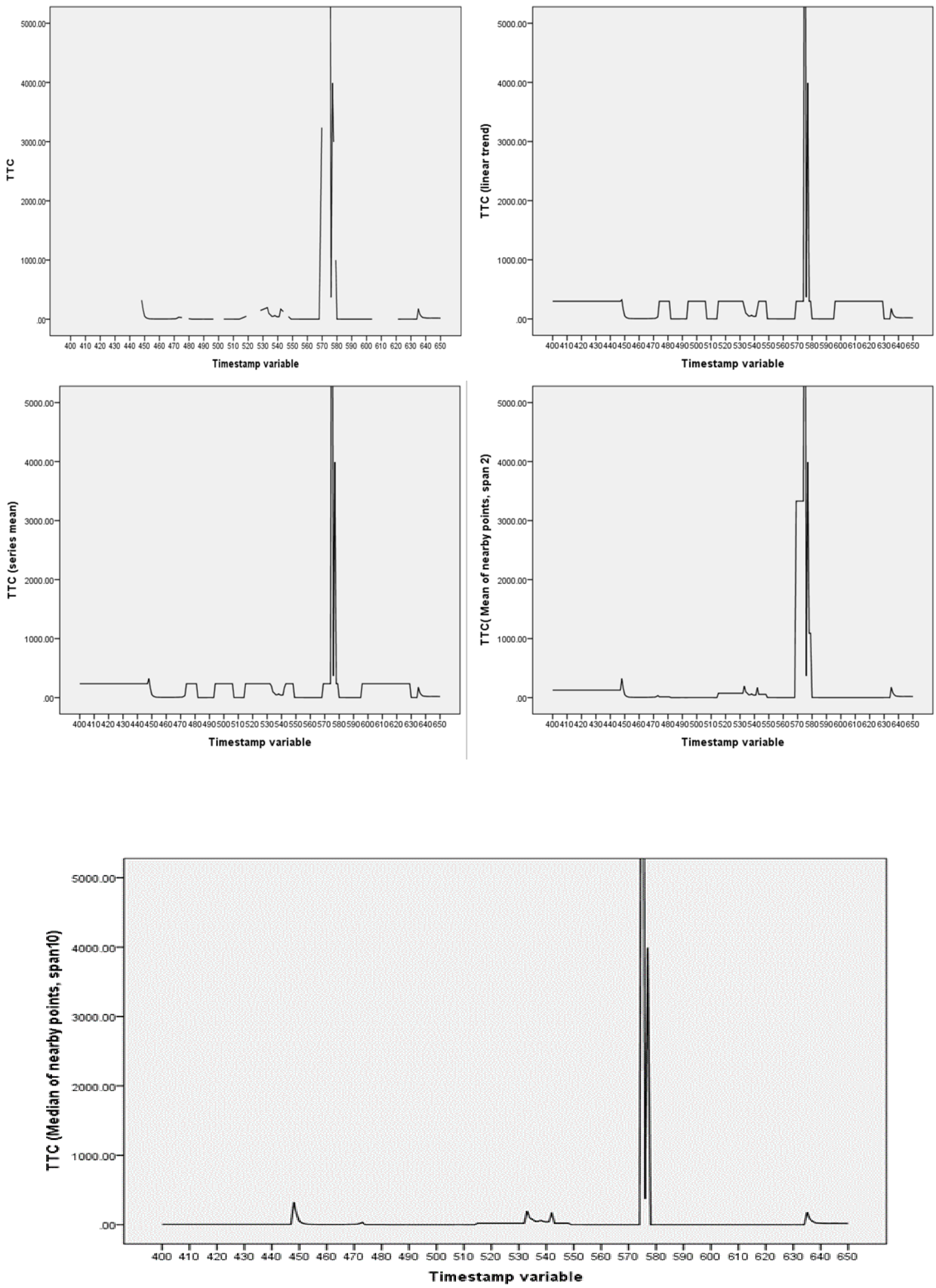


Figure 4.13. Event time-window for different TTC data imputation methods

As observed from the diagrams, but also confirmed by a descriptive analysis of TTC variable, TTC still presents some really high values that constitute a significant problem as they were at least misleading. In Table 4.6, the summary statistics for the new TTC are presented.

Table 4.6. Summary statistics for new TTC

Summary Statistics		
TTC		
N	Valid	1161582
Mean		214.0656
Median		15.58
Std. Deviation		22531.37
Variance		507662556.60
Minimum		0
Maximum		16056368.60
Percentiles	5	.89
	10	1.44
	25	4.27
	50	15.58
	75	48.36
	85	88.04
	90	135.42
	95	256.36
	99	1213.99

From the percentiles' investigation, there is a 1% that presents values higher than 1214 seconds! To address this issue, the extremely high values extending 256 seconds, were replaced with the 95th percentile value. In other words, 5% of the TTC values are replaced with the value of 256 seconds, that is undoubtfully still high.

Therefore, initially the dataset included 3,604 events with 989 drivers involved in a crash or near-crash according to the SHRP2 threshold specifications, but due to the cleaning, matching and TTC selection limitations, this was restricted to 774 driving events, with 553 different drivers. With the resulting driving data sampled at 10 Hz, over 1 million data points were available for analysis.

Crash and near-crash events have considered separately for some initial parts of the analysis, but jointly for the TTC modelling and the driver pre-event behaviour investigation. After the data reduction, the sample size has been greatly decreased especially in the case of crashes, therefore for the modelling and the algorithm development process both crash and near-crash events were employed as safety critical scenarios and under the assumption of similar kinematics in the moments before the final events.

The final dataset comprises 127 crashes and 647 near-crashes and most of them are rear-end striking or road departure events, they occurred in non-adverse weather and free flow conditions, during daylight, in a business industrial region and ego-vehicle's driver seems to be the responsible. The vehicle types available and finally used for the analysis are: Car, Truck, SUV and Crossover, VAN and minivan.

The pathway leading to the final dataset used for the rest of analysis is more clearly indicated in Figure 4.14.

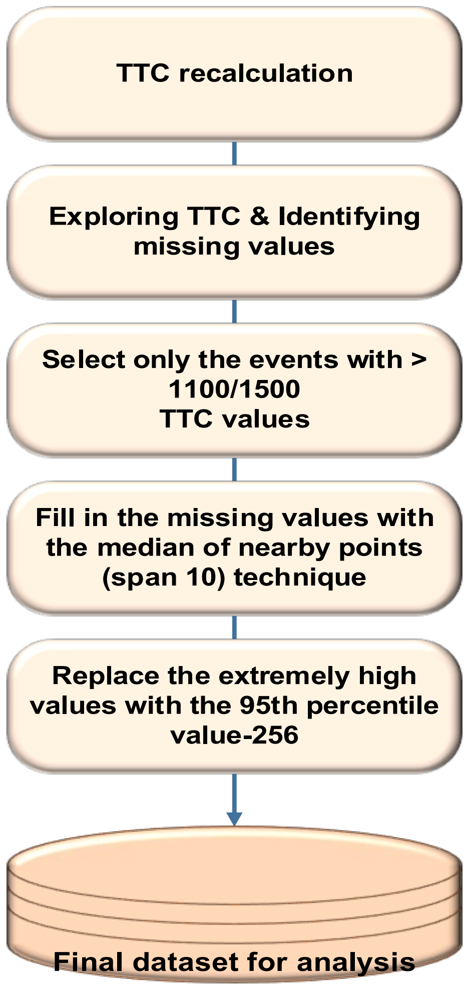


Figure 4.14. Dataset reduction process

4.4 Summary

The chapter presented all the data pre-processing and the preliminary data analysis that was essential for the data understanding and further sophisticated analysis. The SHRP2 NDS data is considered the more comprehensive dataset available until now and has been heavily processed by VTTI to provide numerous variables and plenty of relevant information about the drivers and the events that along with the road infrastructure data render it a very rich dataset to use. Nevertheless, an extensive data manipulation needed for the purposes of this research as the parts of the data had to be carefully combined and most importantly, various difficulties occurred towards obtaining valid TTC values. Therefore, while in the beginning the dataset consisted of 3,604 events with 989 drivers, after the data reduction process that followed, the sample was reduced to 774 events with 553 drivers. TTC is widely used as a collision proximity indicator and it was important to produce a valid dataset where it could be analysed meaningfully.

5 Results

5.1 Introduction

This chapter presents the results produced by employing the methods and datasets discussed in Chapters 3 and 4. Firstly, the analysis of the pilot driving study is presented and then, following the same order as in the methodology chapter, the outcomes are illustrated.

As mentioned in the methodology chapter, for the analysis 1st and 99th percentile are used depending on the sign of the indicator. For instance, for the distribution of deceleration the 1st percentile is sensible to be employed as it denotes the value from which only 1% of the values are more extreme, while for acceleration the 99th percentile shows the value that 1% of the values are above.

5.2 Safety indicators from the pilot study

Data was collected at 100 Hz and as the drivers followed the chosen route for a half of an hour, thousands of observations were gathered for each of them. The most important descriptive statistics are presented below in Table 5.1.

Table 5.1 Summary statistics for longitudinal and lateral acceleration and deceleration

Drivers		**Indicators**		
(g)	Longitudinal Acceleration	Deceleration	Lateral Acceleration (+)	Lateral Acceleration (-)
1st driver		0-31.8 mph		
Min	-	-0.32	-	-0.34
Max	0.29	-	0.44	
1st percentile	-	-0.24	-	-0.23
99th percentile	0.23	-	0.25	-
2nd driver		0-27.9 mph		
min	-	-0.23	-	-0.34
max	0.24	-	0.33	-
1st percentile	-	-0.14	-	-0.19

Drivers	**Indicators**			
99th percentile	0.18	-	0.21	-
3rd driver	0-30.8 mph			
min	-	-0.37	-	-0.34
max	0.35		0.38	
1st percentile		-0.22		-0.24
99th percentile	0.23	-	0.25	-
4th driver	0-37.7 mph			
min	-	-0.55		-0.49
max	0.38		0.43	
1st percentile		-0.24		-0.27
99th percentile	0.22	-	0.3	-
5th driver	0-38.4 mph			
min	-	0.44		-0.45
max	0.3		0.4	
1st percentile		-0.26		-0.27
99th percentile	0.22	-	0.27	-

As can be observed from Table 5.1, the minima and maxima of the longitudinal and lateral accelerations in combination with the 1st and 99th percentiles can give a good picture of the range of these indicators during normal driving. As the researcher was in the car with the drivers can confirm that the entire driving sample represents normal driving conditions except for an incident that the 4th driver had. This is also depicted in the quite high deceleration value occurred (-0.55g).

More specifically, the aforementioned driver braked rapidly when he noticed the leading motorbike decelerated unexpectedly. With this evasive braking manoeuvre, he managed to stop the car in time before a rear-end collision occurs. The deceleration value generated by this abrupt braking can be taken into consideration as a threshold for detecting deviation from normal driving. Below in Figure 5.1 the longitudinal acceleration regarding the time is presented for the trip where the event occurred.

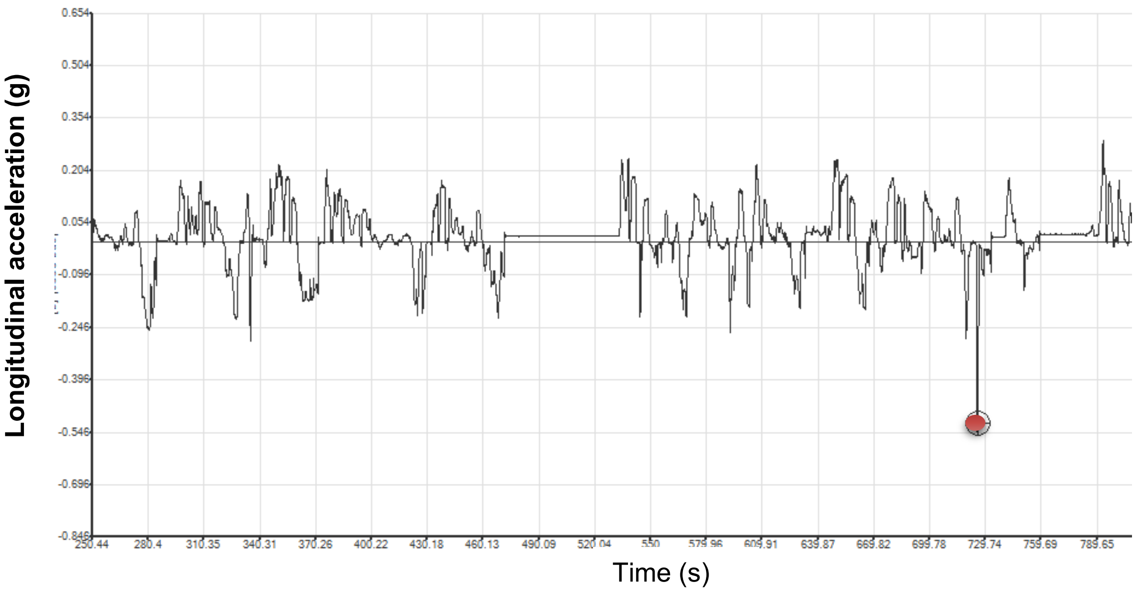


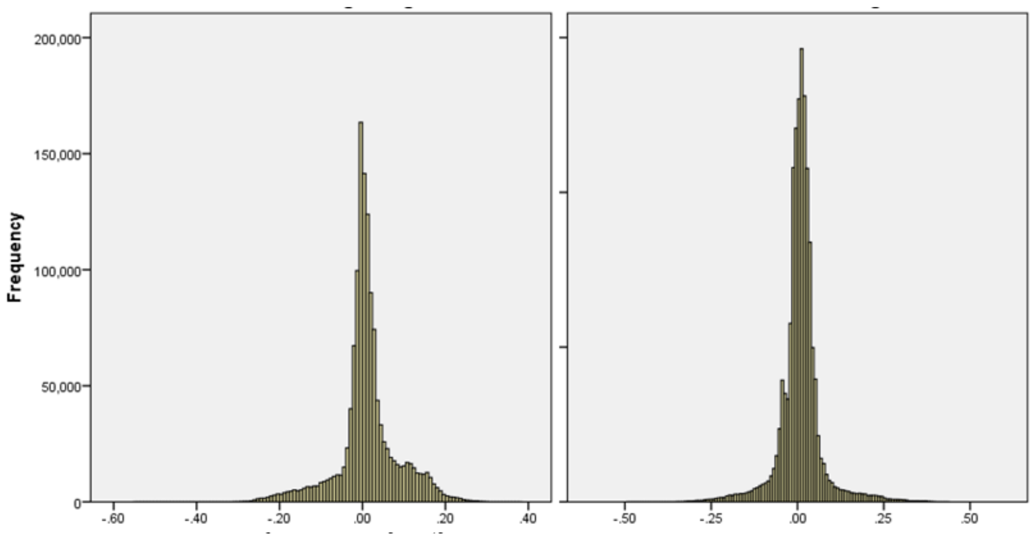
Figure 5.1 Time slice for Longitudinal Acceleration during the trip

In Figure 5.1 it is easy to detect the extreme acceleration value and the existence of a hazardous situation was confirmed from the video frame. The driver's reaction and the sight of the motorbike with the braking light on (Figure 5.2), reassert the harsh and abrupt braking.

Figure 5.2 Video frame with captured driver reaction

Apart from this value, the deceleration did not exceed -0.24g for the 99% of the cases in a speed range of 0-37.7 mph, while accordingly longitudinal acceleration was under 0.22g. Similar results occurred for the lateral acceleration as it can be seen in

Table 5.1. Moreover, the data of the five drivers were combined and the corresponding descriptive statistics are presented.



Longitudinal acceleration (g) **Lateral acceleration (g)**

Figure 5.3 Histograms of longitudinal and lateral acceleration

The most longitudinal acceleration values are between -0.1 and 0.2 while the lateral acceleration values mostly vary between -0.1 and 0.1. Table 5.2 summarises the indicators' summary statistics for the whole pilot study.

Table 5.2 Indicators summary statistics in the pilot study

Indicator (g)	Min	Max	Percentiles 1st	99th	Speed range (mph)
Longitudinal Acceleration	-	0.38	-	0.22	0-38.4
Deceleration	-0.55		-0.24	-	0-37.7
Lateral Acceleration (+)	-	0.44	-	0.26	0-38.2
Lateral Acceleration (-)	-0.49		-0.25	-	0-38.4

In summation, 5 drivers participated in the study producing 2.5 hours of driving data. There was an incident occurred by hard braking due to the suddenly reduced speed of a leading motorbike. Apart from this, all the data represents baseline, normal driving conditions. Longitudinal and lateral acceleration are the two variables that were investigated. It is worth mentioning that despite all the drivers follow the same route; the indicators' values differ between them, indicating different driving styles.

Key findings of the pilot study:

- 2.5 hours of normal driving data
- Familiarity with the instrumented car and experience of producing Naturalistic Driving data
- Knowledge of how to process NDS data and acknowledgement of the difficulties involved
- The ranges for two safety indicators
- A possible threshold for longitudinal acceleration

5.3 Quantifying safety indicators during normal driving

This analysis aimed at understanding normal driving. For this purpose, and as it is more thoroughly described in the methodology chapter, only the first 600 observations (first minute of data) of every trip were utilised and combined in a file as representative of normal driving conditions.

Step-1: Identification of indicators

As it was mentioned in Chapter 3, four indicators have been widely used in previous studies to detect hazardous situations and could be a useful foundation to determine normal driving thresholds. In the framework of this PhD, these four indicators, longitudinal and lateral acceleration, yaw rate and TTC were also employed and selected in order to be explored.

Step-2: Determination of Initial Thresholds

From Table 2.4, in Chapter 2, it can be seen that the TTC values for the case of a safety critical event (SCE) is generally assumed to range from 1.75 seconds to 2 seconds. The initial threshold value of TTC to represent normal driving can be expected to be equal or more than 2 seconds. However, the threshold value varies as it depends on other indicator values as well, i.e. speed or yaw rate. Longitudinal deceleration triggering values randomly fluctuate from -0.2g to -0.65g varying with the speed of the ego-vehicle. The range is large and therefore it is difficult to find a fixed threshold. It however seems that the value for normal driving deceleration should not often exceed -0.6g. Likewise, lateral acceleration values vary from 0.25g to 0.7g and reported to depend on travelling speed of the ego-vehicle. Based on these arguments and relevant literature, the following values are proposed as Initial Thresholds:

TTC ≥ 2 seconds;

Longitudinal deceleration ≥ - 0.5g;

Longitudinal acceleration ≤ 0.5g;

Lateral acceleration (left or right) ≤ 0.65g or ≥ - 0.65g;

Yaw rate < $|45|^0$/s

These values are considered to be the first attempt in quantifying normal driving without currently taking into account all the operational conditions, e.g. different speeds and road networks.

Step-3: Validation of the Initial Thresholds

For the validation of the thresholds proposed, SHRP2 NDS data was employed and analysed. More specifically, normal driving data from 553 drivers was analysed to identify threshold values for the key indicators that can be used to detect deviations from normal driving. These are: TTC, longitudinal deceleration (i.e. braking), longitudinal acceleration, lateral acceleration and yaw rate.

5.3.1 Time-To-Collision (TTC)

The process followed to calculate TTC from the initial dataset that was received, is described thoroughly in Chapter 4. TTC was calculated for a new, suitable for this purpose, dataset containing 1,161,583 observations; these observations represent 773 trips and 553 drivers and vehicles.

From the normal driving dataset (first 600 observations of every trip; totally 464,400 observations) dataset, it was found that TTC was higher than 2 seconds for at least 96.4% of the cases. This increased to 10 seconds for at least 71.6% of the cases. However, 1% of the cases have TTC lower than 1.6 second and this was largely due to small *d*-values (i.e. *d* was as low as 2m) used in the sample data. The initial threshold value for TTC in detecting any deviation from normal driving was chosen as 2 seconds (See Step-2). From this analysis, it has been found that 3.6% of the cases have TTC values less than 2 seconds and that TTC values are related to $\Delta\vartheta$ and d. It is therefore very challenging to identify an optimal threshold for TTC and there are other influencing factors. Considering all these factors, it can be concluded that a threshold value of 2 seconds for TTC would provide over 95% confidence level that the corresponding driving may be regarded as 'normal'.

Non-parametric tests were conducted to examine whether driving behaviours with respect to TTC vary by gender. Both the Wilcoxon rank-sum test and the median test rejected the null hypothesis at the 95% confidence level indicating that their driving behaviours are different. The 1^{th} percentile value of TTC for the female drivers was 1.9 seconds and the same value for the male drivers was 2.6 seconds. Similar results on driving behaviours with respect to TTC were obtained from the same tests for young drivers (16-24) vs other drivers (25+) with a value of 2.1s and 2.5s respectively.

Since these scenarios were considered to be 'normal driving' in the NDS data, it can be concluded that identifying a single threshold value for TTC is very challenging as this largely depends on other factors.

5.3.2 Longitudinal acceleration-deceleration

Longitudinal acceleration may constitute the most popular kinematic criterion for detecting safety critical events from NDS data. It is recorded as a variable in the SHRP2 data; therefore, it does not need any additional calculation. The analysis was conducted for 464,400 observations. The negative values of this variable are considered deceleration and the positive ones are acceleration.

Examining the percentiles though for the deceleration, it can be seen that 99% of the cases are higher than -0.26g (less negative) while the minimum value reaches -0.75g. However, further investigation reveals that the values under -0.5g represent only the 0.015% of the total deceleration. Regarding the acceleration, 99% of the cases were under 0.24g while the maximum value was 0.45g.

In order to evaluate the differences in distribution of the longitudinal acceleration and deceleration variables between the gender and the age groups, the aforementioned non-parametric tests were conducted. The tests showed that the distribution of deceleration differs in a statistically significant way (95% confidence level) across the categories of age group (16-24, 25+) and the gender. This indicates different deceleration patterns between the male and female and between the younger and the other drivers. More specifically, the 1^{st} percentile deceleration value was -0.28 g for the female drivers and -0.25g for the male ones while the same value was again -0.28g for the younger drivers and -0.25g for the other age group, implying that women and younger drivers may brake harder. In terms of the acceleration, the 1^{st} percentiles were the same, 0.23g for male and female drivers while the younger ones presented slightly higher accelerations than the older (99^{th} percentile: 0.24g, maximum: 0.45g vs 99^{th} percentile: 0.23g, maximum: 0.43g).

5.3.3 Lateral acceleration

Lateral acceleration appears also as a recorded variable in SHRP2 NDS dataset. The results of the descriptive analysis showed that, regarding the negative lateral acceleration, 99% of the cases were above -0.23g and 99.9% of them were also higher than -0.41g. The minimum value was -0.63g. Correspondingly, 99% of the positive lateral acceleration cases were below 0.26g and 99.9% of them were lower than 0.44g. The maximum value was 0.63g.

The non-parametric tests reveal that this variable also differs regarding its distribution across the gender and the age groups of the drivers. Table 5.3 presents the 99th percentile values in every case.

Table 5.3 Lateral Acceleration-99th (positive values) and 1st (negative values) percentile values

Lateral acceleration	Male drivers	Female drivers	Younger drivers (16-24)	Older Drivers (25+)
99th or 1st percentile	0.29g/ -0.24g	0.25g/ -0.22g	0.32g / -0.25g	0.22g / -0.17g

5.3.4 Yaw rate

Finally, yaw rate is also provided in SHRP2 NDS dataset. It presents positive and negative values as the other variables, but as they were found to be symmetrical, for this analysis only the negative ones are explored. The absolute values could have been used, as well. The minimum value of yaw rate was -52.68 °/s while 99.9% of the values were higher than -27 °/s and 99% higher than -16 °/s. The distributions of this variable were also found to be different across gender and age group and the 1st percentile values are presented in Table 5.4.

Table 5.4 Yaw rate-1st percentile values

Yaw rate	Male drivers	Female drivers	Younger drivers (16-24)	Older Drivers (25+)
1st percentile	-15°/s	-16.3 °/s	-16.9 °/s	-14.3 °/s

Step 4: Modification of the Initial Thresholds based on the findings in Step-3.

The initial thresholds of selected variables were presented according to: a) the thresholds employed in several NDS b) the pilot study and c) reasonable assumptions. After the analysis conducted in Step-3, based on SHRP2 NDS data, modified thresholds of indicators representing normal driving can be proposed (Table 5.5):

Table 5.5 Thresholds for normal driving

Acceleration/ Deceleration	Lateral acceleration	Yaw rate	TTC
|0.35|g	|0.4|g	|25| °/s	2 s

The analysis of the data in step 3 showed that a fixed threshold is very difficult to represent all the conditions and that a more dynamic one should be determined. For this purpose, the 1st percentile values of these indicators (99th for longitudinal acceleration) were derived for different speeds of the ego-vehicle (Table 5.6). Figure 5.4 and Figure 5.5 depict the relevant indicators trends.

Table 5.6 Indicators vary by speed of the ego-vehicle

Speed bins (km/h)	TTC(s)	Longitudinal Acceleration(g)	Deceleration(g)	Lateral Acc.(g) (-)	Yaw rate(^{0}/s) (-)
(0-10]	2.71	0.3	-0.29	-0.29	-13.42
(10-20]	1.62	0.33	-0.32	-0.27	-25.04
(20-30]	1.48	0.29	-0.32	-0.33	-24.06
(30-40]	1.39	0.26	-0.33	-0.45	-20.49
(40-50]	1.32	0.23	-0.29	-0.32	-9.07
(50-60]	1.31	0.19	-0.26	-0.19	-6.18
(60-70]	1.4	0.16	-0.22	-0.13	-5.85
(70-80]	1.73	0.14	-0.21	-0.20	-6.18
(80-90]	2.15	0.13	-0.20	-0.20	-6.83
(90-100]	3.13	0.12	-0.15	-0.18	-5.20
(100-110]	3.6	0.1	-0.11	-0.12	-4.23
>110	2.8	0.11	-0.14	-0.13	-4.88

For example, the 1st percentile value of TTC for speed between 40 km/h to 50 km/h is 1.32 seconds indicating that at least 99% of the TTC values in the sample are greater than 1.32 seconds. TTC follows a declining trend until the speed bin of 50-60 km/h and increases for higher speeds to decrease again at the speed level of more than 110km/h.

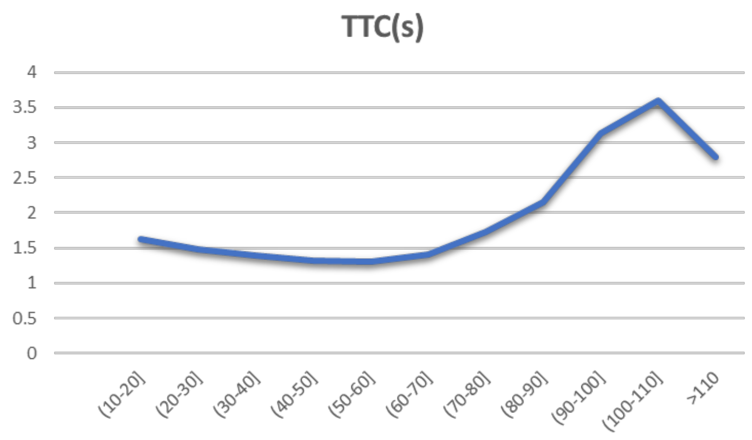


Figure 5.4 TTC 1st percentile values per speed bin

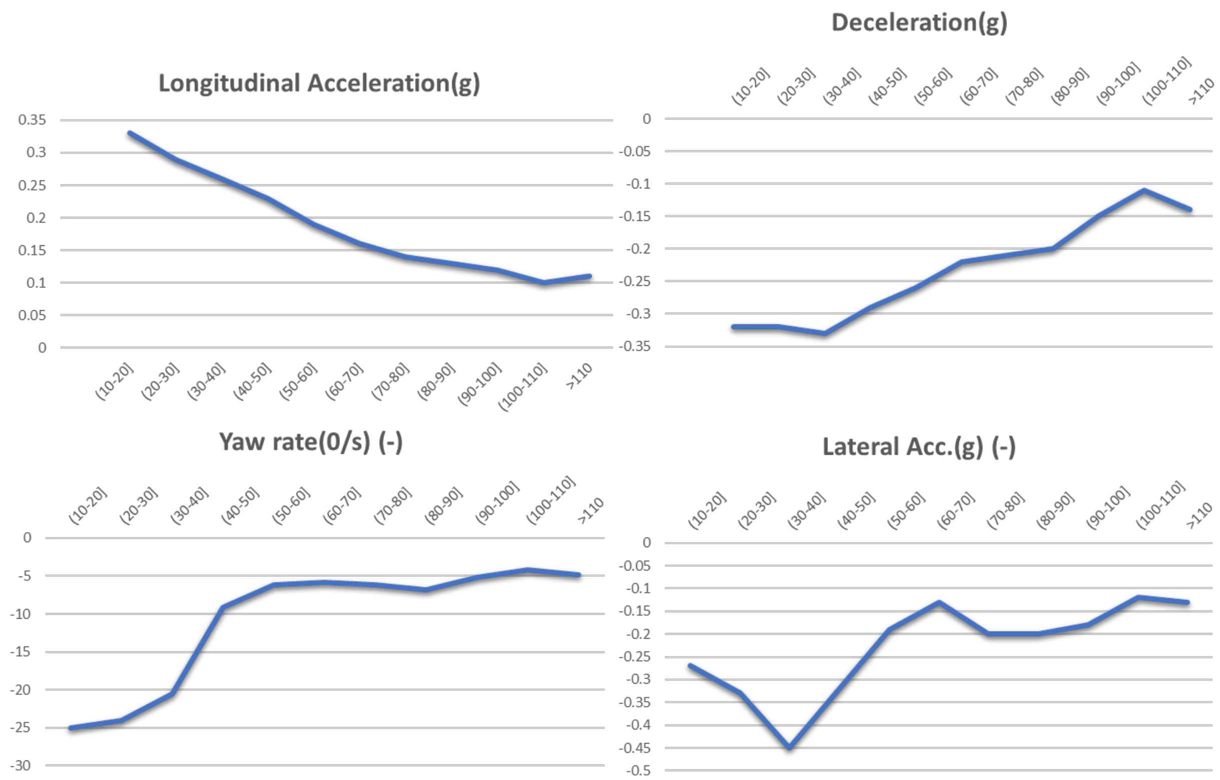


Figure 5.5 Indicators 1st percentile values per speed bin

Yaw rate indicates a clear decrease in relation to speed. Longitudinal acceleration (positive values) follows a declining course with speed increase, deceleration slightly increases up to 40 km/h and then also decreases while lateral acceleration does not seem to follow any clear pattern as indicates random fluctuations.

Based on the threshold values shown in Table 5.7, linear regression models were employed to determine threshold values for the indicators. Therefore, the functional linear regression equations for each indicator (95% confidence level except for TTC that the equation has been developed for a 90% confidence interval) have been formed as follows:

Table 5.7 Functional equations

Indicator	**Functional forms**	**Goodness-of-fit (R^2)**
Time to collision:	$TTC_{th} = 1.31 + 0.12 \text{ Speed}$	0.31
Longitudinal Acceleration:	$LongAcc_{th} = -0.329 - 0.002 \text{ Speed}$	0.94
Deceleration:	$Dec_{th} = -0.355 - 0.002 \text{ Speed}$	0.87
Lat. acceleration (-):	$LAcc_{th} = -0.358 + 0.002 \text{ Speed}$	0.55
Yaw rate (-):	$YR_{th} = -21.245 + 0.172 \text{ Speed}$	0.62

As observed, longitudinal acceleration and deceleration present high R^2 values. This can be explained by the fact that the models have been developed with aggregated data consisting of a few (12) datapoints (percentiles per speed bin). The residuals are normally distributed according to Kolmogorov-Smirnov (K-S) test and visual inspection and they have been also tested for serial autocorrelation using Durbin-Watson test.

It can be concluded that a detection of any deviation from normal driving would not only require the simultaneous measurements of multiple indicators, but also the different threshold values per indicator based on driver demographics and different speed levels as shown in Table 5.6. The indicators' thresholds for detecting deviation proved to be more dynamic than these presented in other studies. A multivariate analysis utilising vehicle kinematics data related to normal driving, as well as safety critical events, is therefore needed to identify the difference in thresholds within and across the indicators. Any future study shall also consider other indicators, such as time headway and also, driver state and environmental conditions.

Normal driving is a broad concept to comprehend and investigate, but this study has taken a step further in our understanding of driving and the difference in driving patterns between gender and age group. It will contribute to an understanding of the variables for characterising normal driving in order to model the relationships between them and provide a basis for investigation of different driving style patterns in different environments and situations. Moreover, this will introduce a new approach in detecting deviation from normal driving. The deeper the knowledge about it, the more effective detection systems will be developed. Future research can include more variables, i.e. time headway and lateral position and can be also enriched with driver clustering on different characteristics other than demographics.

5.4 Safety indicators for detecting deviation from normal driving

In order to detect early deviations from normal driving, the whole event sequence development needs to be explored. For this purpose, in this subsection, all the observations of every trip have been investigated; hence 2.5 minutes of time series data corresponding to every trip, from a normal driving situation until the event have been utilised for the analysis.

In Chapter 3, the methodology of deriving indicators has been thoroughly developed and explained. In this section, their usefulness regarding detection of early deviations is examined. As it is explained before, in case of right hand distributions (positive values), useful value would be the 95th percentile while for left hand distributions (negative values) the corresponding value is the 5th percentile. For consistency reasons regarding the generation and use of the indicators, only the negative values of them have been investigated at this stage and 5th percentile indicators were developed. Acceleration did not enable the calculation of a valid indicator due to limited values in many timepoints, but as the sample includes many rear-end events the deceleration may be a more suitable indicator for detecting abnormalities in driving.

5.4.1 Exploratory analysis of vehicle kinematics before a crash or near crash

Initially, raw data of the indicators from all the individual trips has been explored and plotted to acquire an understanding of the vehicle kinematics during the crash development. Figure 5.6 and Figure 5.7 constitute two examples of this process concerning the longitudinal acceleration and yaw rate from all the individual trips. As it can be observed, it is challenging to derive any conclusions from the diagram of the individual trips. However, combining them and generating the 5th percentile indicators, as described in previous chapter, could reveal patterns of the safety indicators during the development of the event.

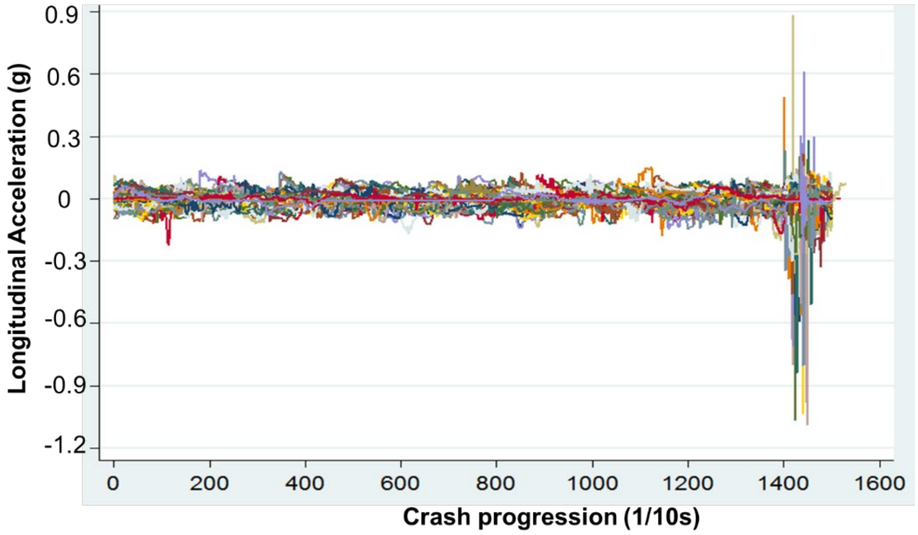


Figure 5.6 Longitudinal acceleration during crash progression (data from all the crashes)

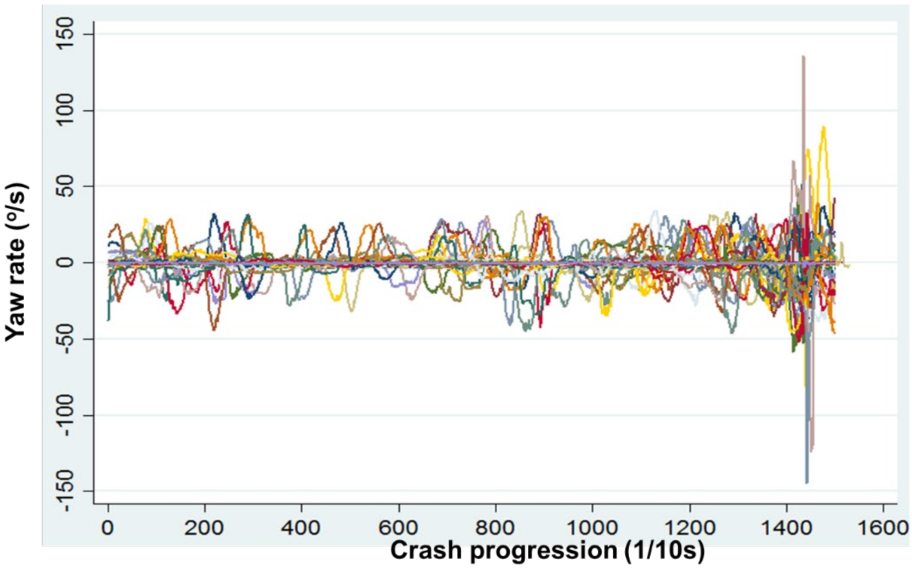


Figure 5.7 Yaw rate during crash progression (data from all the crashes)

Figure 5.8, Figure 5.10, Figure 5.12, Figure 5.14, represent the event sequence development from normal driving until the crash events (127) and near crash events (647) for a range of safety indicators (5th percentile) related to vehicle kinematics, namely, longitudinal deceleration, lateral acceleration (left side indicating negative values), yaw rate (negative values) and TTC. For all the four variables, vehicle kinematics data related to near crash events differ from the corresponding data for the crash events (Mann-Whitney U test at the 95% confidence level, $p<0.001$). The distributions and the corresponding descriptive statistics of these four indicators are also presented below (see Figure 5.9, Figure 5.11, Figure 5.13, Figure 5.15, Table 5.8, Table 5.9, Table 10, Table 5.11) to indicate that surrogate safety measures vary by event severity.

During the whole event development, the magnitude of the variation of the indicator values for crashes is greater than this of near-crashes, but the most significant element is that the deceleration, lateral acceleration and yaw rate indicator values for crash events tend to reduce sharply approximately 10-20 seconds before the event indicating a clear deviation from the baseline driving. However, this is not the case with TTC.

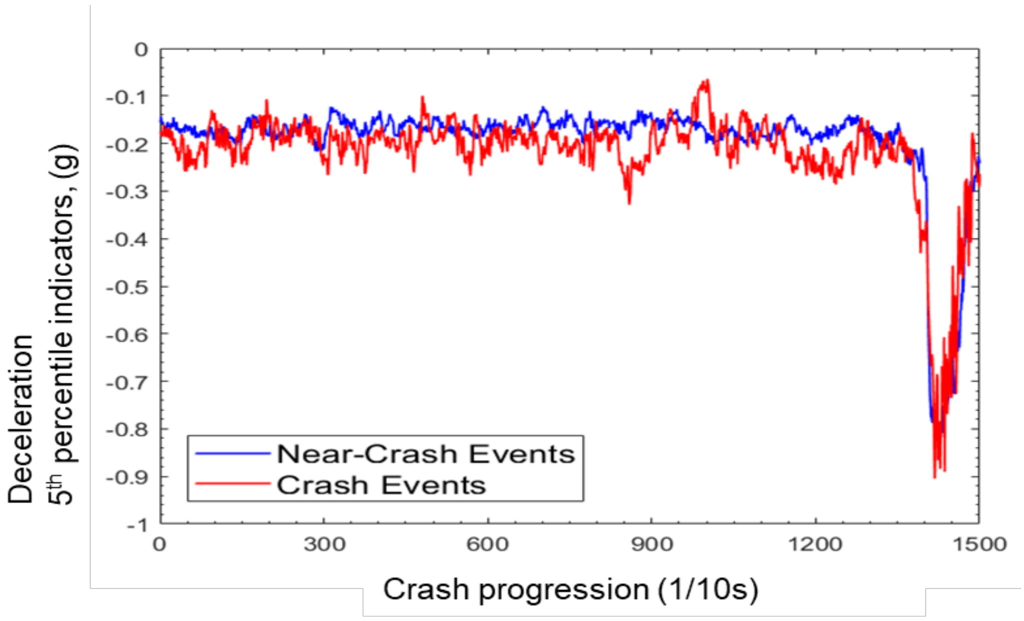


Figure 5.8 Deceleration 5th percentile indicators during crash & near-crash development

The deceleration 5th percentile indicators during the crash sequence show a slight abnormal sudden decrease in values approximately 5 seconds before the main event while in near-crash development, there is no indication of reaction before the actual evasive braking. From the descriptive statistics, it can be observed that the indicators for the crashes present more extreme values than those for the near-crash events and from the histograms displayed in Figure 5.9 the difference in their distribution is visible.

Table 5.8 Descriptive statistics of Deceleration 5th percentile indicators

Deceleration (g)	Crashes (127)	Near Crashes (647)
Mean	-0.22	-0.196
Median	-0.198	-0.169
Minimum	-0.905	-0.817
Maximum	-0.064	-0.121

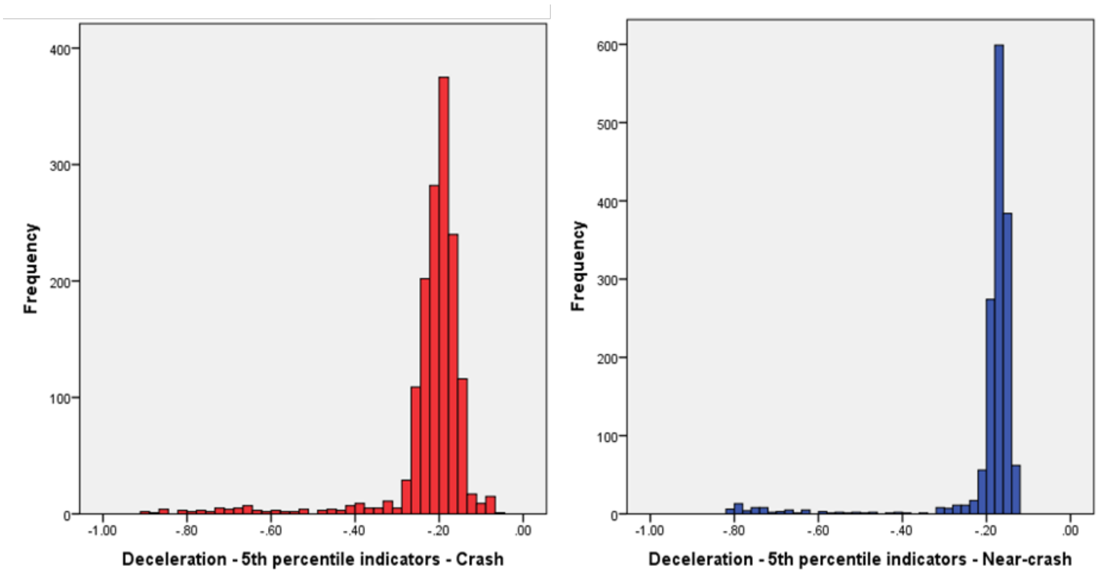


Figure 5.9 Distributions of Deceleration 5th percentile indicators for crash & near crash events

For lateral acceleration, the indicators minimum for crashes is two times larger than that of near-crashes and the distributions are also different. In crash development, there is a value reduction 5-10 seconds before the main event while in near-crash one, no significant decrease is observed before the event.

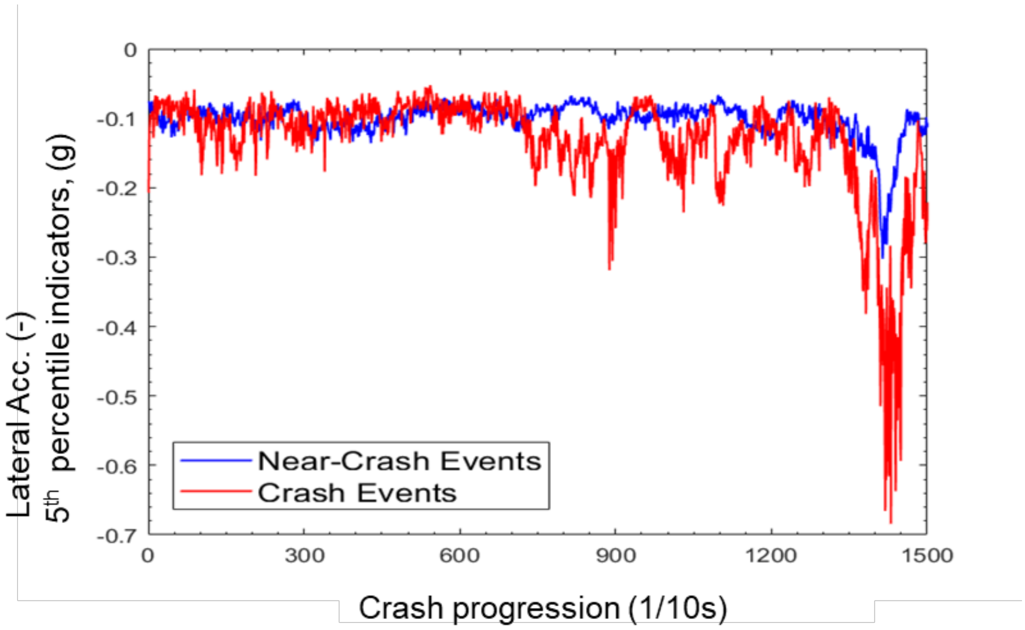


Figure 5.10 Lateral Acceleration (-) 5th percentile indicators during crash & near-crash development

Table 5.9 Descriptive statistics of Lateral acceleration (-) 5th percentile indicators

Lateral Acc. (-), (g)	Crashes (127)	Near Crashes (647)
Mean	-0.132	-0.102
Median	-0.112	-0.096
Minimum	-0.684	-0.303
Maximum	-0.052	-0.067

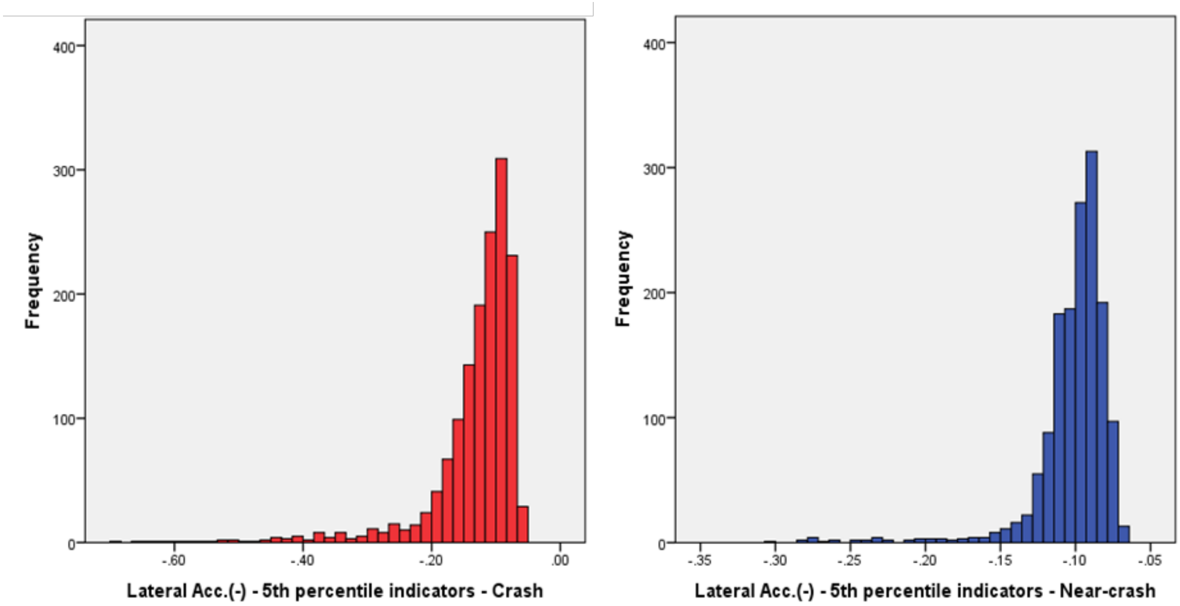


Figure 5.11 Distributions of Lateral Acceleration (-) 5th percentile indicators for crash & near crash events

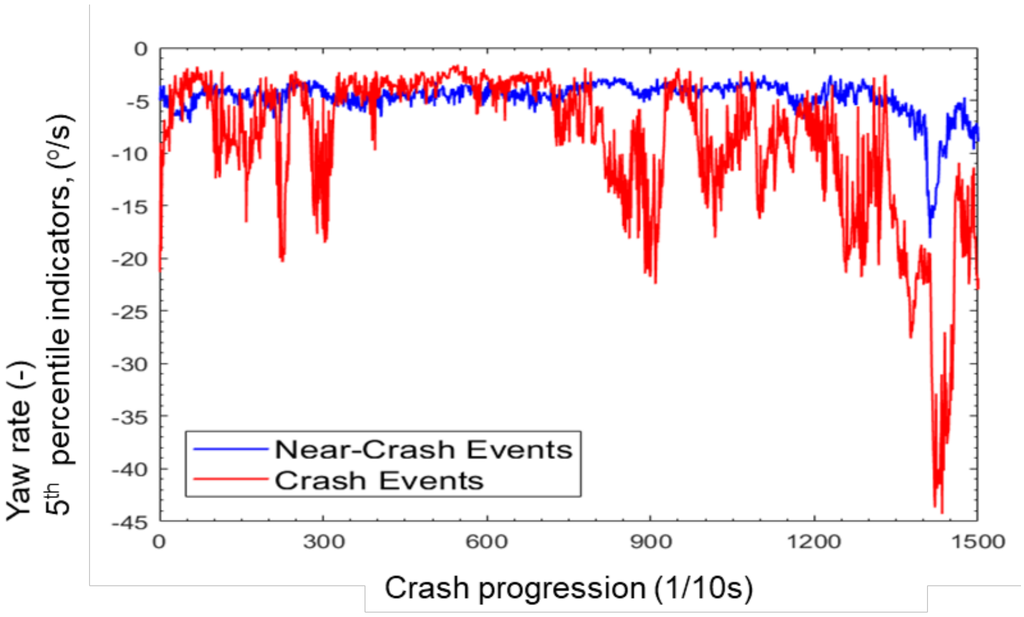


Figure 5.12 Yaw rate (-) 5th percentile indicators during crash & near-crash development

Regarding the yaw rate 5th percentile indicators, as it is observed from the descriptive statistics and the diagrams, their distribution differs significantly between crash and near-crash events. Moreover, 7-10 seconds before the main event in crash sequence, there is a detectable decline in the yaw rate indicator values suggesting a deviation from normal driving a considerable time before the event.

Table 5.10 Descriptive statistics of Yaw rate (-) 5th percentile indicators

Yaw rate. (-), (º/s)	Crashes (127)	Near Crashes (647)
Mean	-9.140	-4.751
Median	-7.057	-4.39
Minimum	-44.292	-18.097
Maximum	-1.626	-2.602

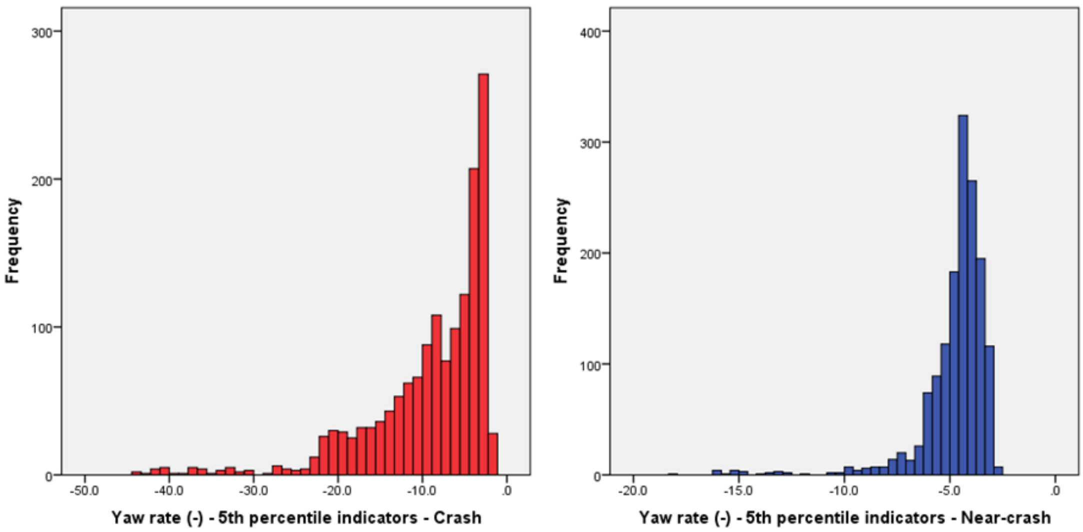


Figure 5.13 Distributions of Yaw rate (-) 5th percentile indicators for crash & near crash events

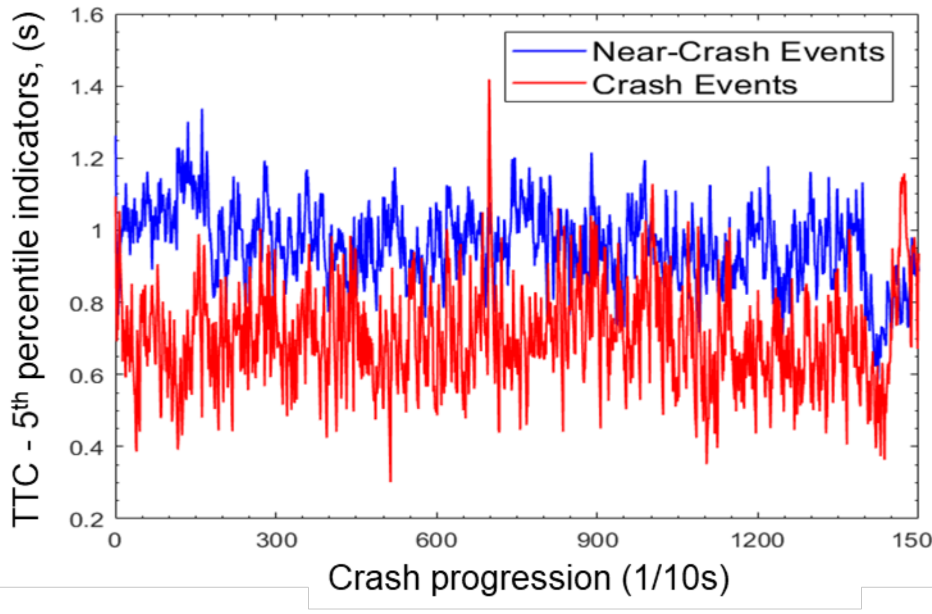


Figure 5.14 TTC 5th percentile indicators during crash & near-crash development

TTC 5th percentile indicators in crash sequence development are significantly lower than in near-crash one. However, the indicators values do not exhibit any clear pattern, rendering any conclusion impossible.

Table 5.11 Descriptive statistics of TTC 5th percentile indicators

TTC (s)	Crashes (127)	Near Crashes (647)
Mean	0.707	0.959
Median	0.696	0.963
Minimum	0.301	0.584
Maximum	1.419	1.337

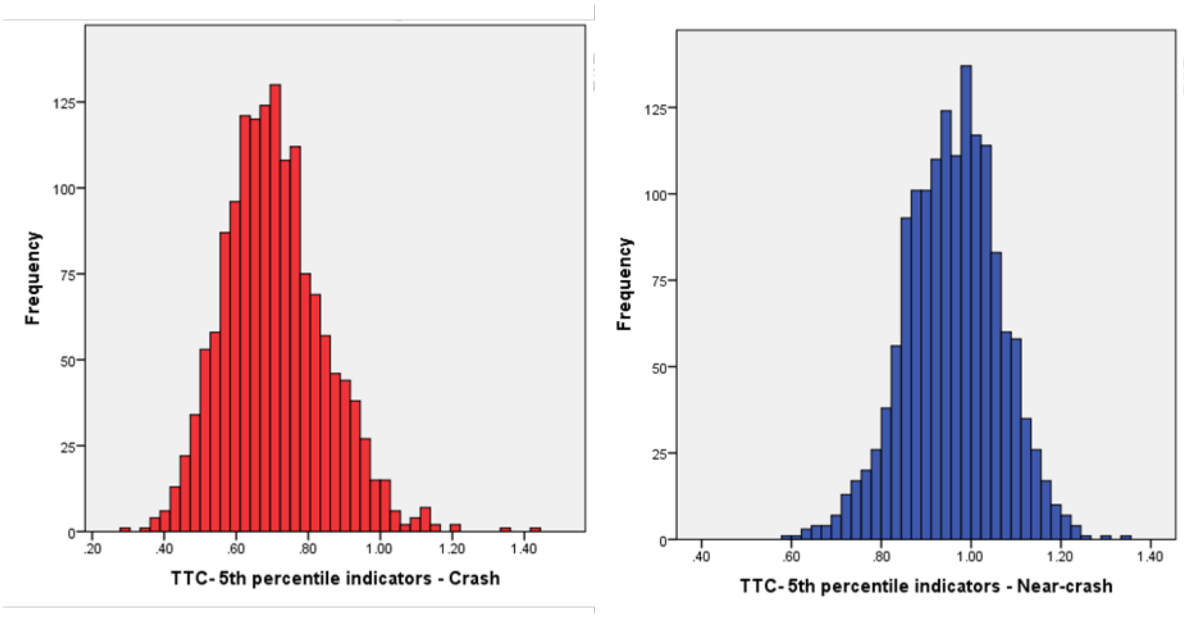


Figure 5.15 Distributions of TTC 5th percentile indicators for crash & near crash events

Above, the differences regarding the severity level were indicated, but it would be also interesting to explore the indicators considering male and female drivers. Figure 5.16 and Figure 5.17 show the 5th percentile indicators during near-crashes and crashes respectively allowing the comparison between males and females pre-event driving behaviour.

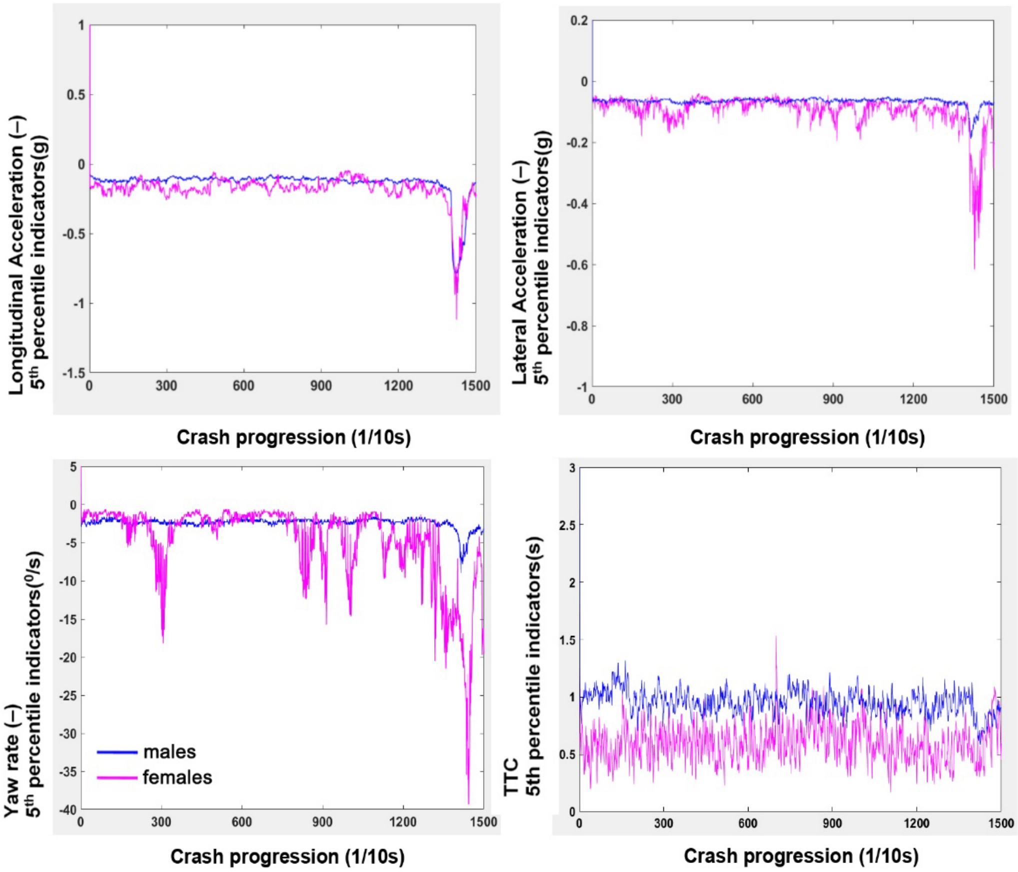


Figure 5.16 Safety indicators related to vehicle kinematics during the crash development for male and female drivers

During the crash progression, female drivers present more extreme values for all the indicators. As it can be easily noticed, longitudinal acceleration, lateral acceleration and yaw rate indicators present more fluctuations for females while TTC seems to be significantly lower than the males' one. During near-crash development, the differences between males and females are milder or minimal. The two lines present high similarity. These have been also statistically confirmed with the help of Mann-Whitney U test for 95% confidence interval ($p<0.0001$).

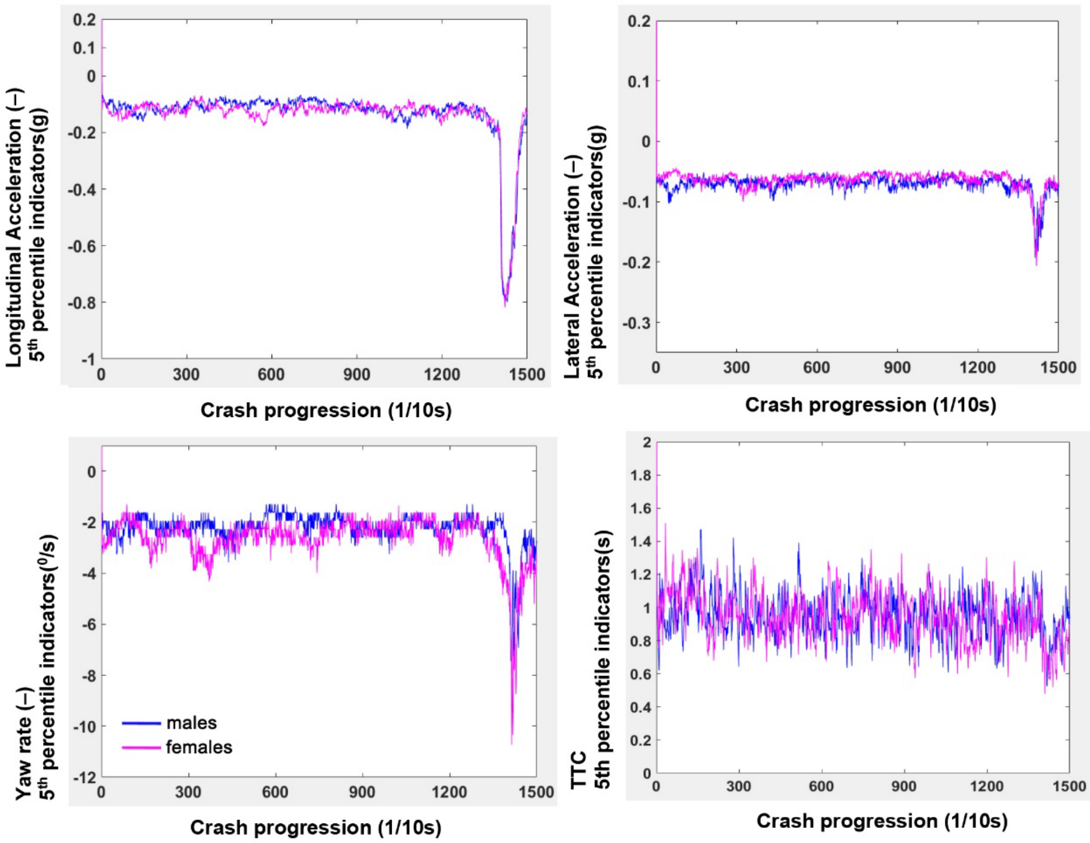


Figure 5.17 Safety indicators related to vehicle kinematics during the near-crash development for male and female drivers

Deceleration indicator was also explored regarding age groups and showed that the adult and older drivers present more extreme deceleration values than the younger drivers. This is the mostly the case near or during the safety critical event as in their normal driving phase do not exhibit great differences according to the indicators as shown in Figure 5.18. A statistically significant difference occurs between the young and the older drivers, and between the young and the adult drivers (Mann-Whitney U test, 95% confidence interval, p<0.0001).

.

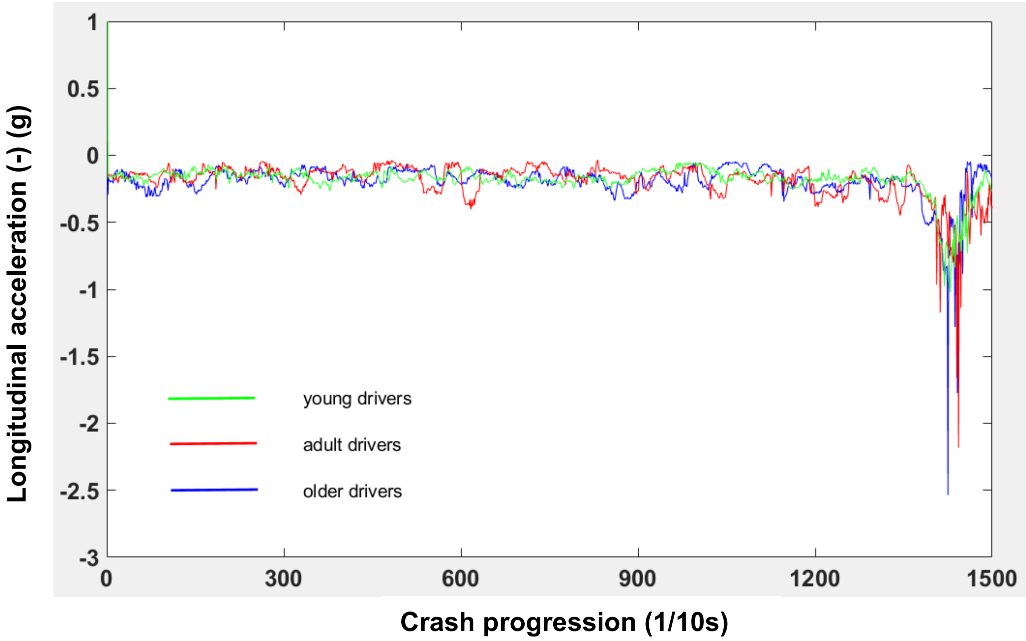


Figure 5.18 Longitudinal acceleration during the event development for young, adult and older drivers

Finally, the road departure crashes were checked regarding the yaw rate. In comparison with all the other types of crashes, they present more extreme values in yaw rate 5th percentile indicators as expected, Figure 5.19.

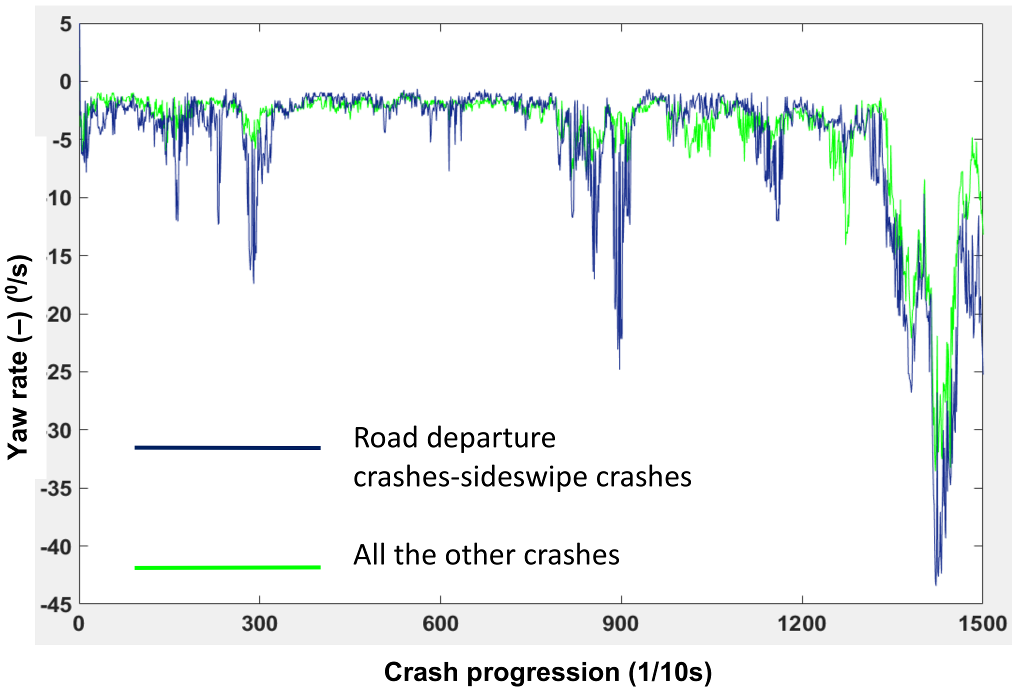


Figure 5.19 Yaw rate during the crash development for Road departure crashes-sideswipe crashes and all other crashes

In summary, in this section, indicators capable of detecting the departure from normal driving were presented. Aggregating the time series data in the timestamp variable, so in the form of the 10^{th} of a second, combining the relative 10^{th} of a second for all the events and calculating the 5^{th} percentiles of them led to the conclusion that near crash evolution differs from the crash one and possibly an early detection of deviation is feasible. Investigating further these indicators will provide valuable insights in comprehension of crash mechanisms and evolution of crash risk. As TTC did not show to follow any explicit pattern, an in depth analysis is required to determine the factors affecting the evolution of its values during the crash sequence development. The indicators suggested that the last 30 seconds have particular interest not only in terms of detecting deviation from normal driving but also regarding pre-event driver behaviour.

5.5 Safety indicators during safety critical driving scenarios

Examining the vehicle kinematics before the event in the previous subsection, revealed that an early indication of deviation could be detectable using the indicators of deceleration, lateral acceleration and yaw rate, while TTC did not provide any useful results in this analysis. The last 30 seconds of the event sequence development indicated drivers' reactions to events and arose interest regarding the feasibility of detecting deviations in early stages of crash or near-crash development. Investigating pre-event driver behaviour using the indicators explored in previous section could give valuable insights in the understanding of the transition from a normal driving situation to a safety critical event. Therefore, in this section the last 30 seconds time period before the events will be investigated using deceleration and yaw rate to determine drivers manoeuvring before the events. This analysis focused on deceleration and yaw rate as these indicators can characterise the magnitude of any avoidance action performed by a driver. The crash proximity indicator, TTC and how its values change during the crash development stages will be further and separately investigated in the next section.

In the methodology chapter, the process to extract the data of interest is thoroughly described. More specifically, a personalised threshold of three standard deviations was set for every driver and braking and yaw rate events during the crash sequence development were identified and extracted to be explored.

5.5.1 Deceleration

The dataset contains 774 events, 647 near-crashes and 127 crashes. Exploring the 30seconds pre-crash period, seven common cases were identified : i) trips where the drivers did not seem to implement any evasive braking reaction, in other words the threshold of 3 standard deviation was not exceeded, ii) trips where the drivers have just one braking event that was the critical one, iii) trips with 2 deceleration events, iv) trips with 3, v) trips with 4, vi) trips with 5 and finally, vii) trips with 6 deceleration events. A driver categorisation according to the number of pre-incident braking events is presented in Table 5.12.

Table 5.12 A summary of Driver deceleration behaviour 30 seconds before main incident

Cases	Trips	Event severity	Gender	Age-Group	Incident type
No evasive reaction - the threshold was not exceeded at any point	149	42 crashes, 107 near-crashes	82 F 67 M	65drivers≤24, 24>61drivers<65, 18 drivers≥65	68 rear-end striking, 29 road departure, 29 sideswipe, 16 other
1 deceleration event (critical)	343	40 crashes, 303 near crashes	186 F 156 M	171drivers≤24, 24>133drivers<65, 39 drivers≥65	226 rear-end striking, 21 road departure, 47 sideswipe, 49 other
2 deceleration events	169	24 crashes, 145 near crashes	76 F 91 M	86 drivers≤24, 24>66drivers <65, 17 drivers≥65	105 rear-end striking, 13 road departure, 20 sideswipe, 31 other
3 deceleration events	76	14 crashes, 62 near crashes	43 F 33 M	35 drivers≤24, 24>30drivers <65, 11 drivers≥65	41 rear-end striking, 6 road departure, 17 sideswipe, 12 other
4 deceleration events	24	5 crashes, (3 road departure, 1 rear end striking, 1 rear end struck), 19 near crashes	9 F 15 M	10 drivers≤24, 24>14drivers <65, 0 drivers≥65	17 rear-end striking, 3 road departure, 2 sideswipe, 12 other

Cases	Trips	Event severity	Gender	Age-Group	Incident type
5 deceleration events	11	1 crash (rear -end struck-65-69), 10 near-crashes	5 F 6 M	5 drivers≤24, 24>4drivers <65, 2 drivers≥65	7 rear-end striking, 1sideswipe, 3 other
6 deceleration events	2	1 Crash 1 Near-crash	2 F	1 driver→16-19 1 driver→30-34	1 road departure(crash) 1 rear-end striking(near-crash)

Investigating the trips that present 6 events during this period, the one is a road departure crash involving a very young female while the second is a near -rear end striking-crash dealt by female adult. Moreover, in the first categories, most of the trips are related to near-crashes, rear-end striking type of events and involve slightly more women than men. The age groups that prevail are these of young people (16-19, 21-24). That is expectable though as the sample is biased towards young drivers and rear-end striking events and involves slightly more women.

5.5.1.1 Final braking events

It is assumed that the final event as it has been defined (above 3 standard deviations for 1 s) is the critical one (crash or near-crash). The durations of each final no empty duration column were gathered in one column, so a variable with the duration of the final critical events of all the trips has been created. It should be mentioned that there are 149 trips with no exceeded thresholds though, therefore the variable contains 625 observations/events durations. Exploring the distribution of the durations, the histogram below reveals that most of the events have duration below 30 timestamps (3s) with a mean of 2.55 s and standard deviation of 1.79. It should be mentioned that the max deceleration (actually the min as they are negative values) does not always appear in the last event of the trip.

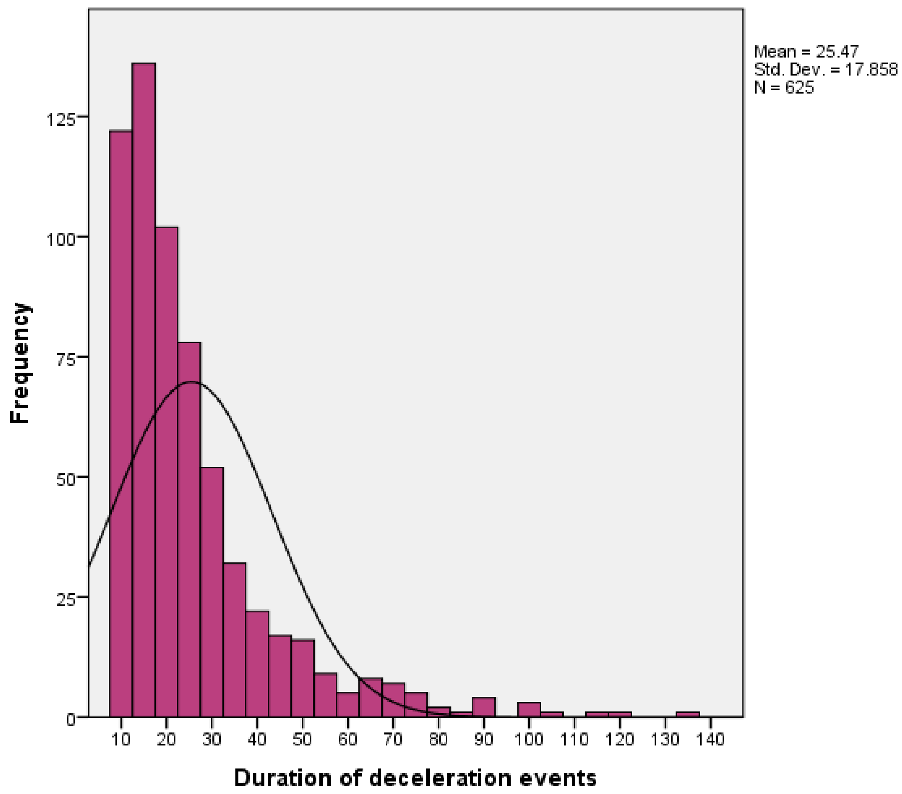


Figure 5.20 Histogram of deceleration events duration

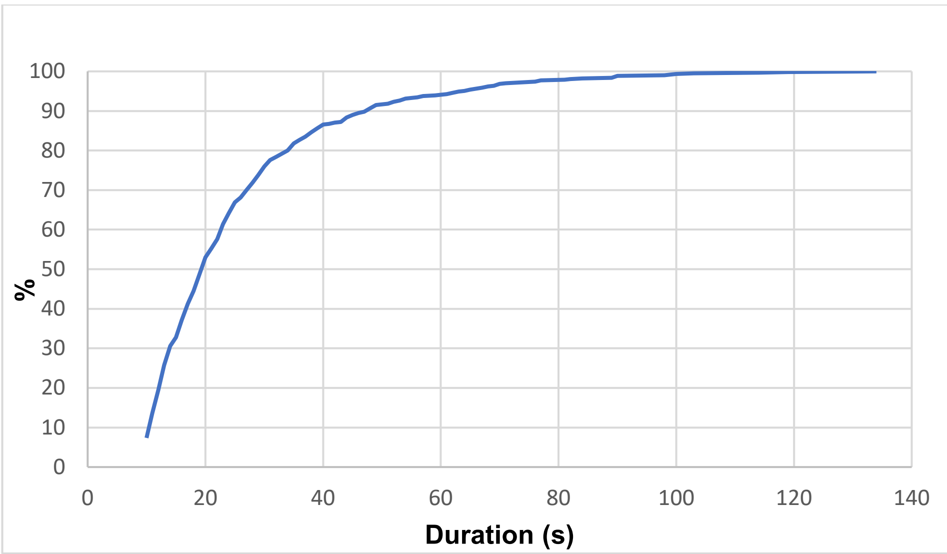


Figure 5.21 Cumulative frequency distribution of deceleration events duration

From the cumulative frequency distribution, it can be observed that 99.7% (within 3 standard deviations of the mean) of the events in the dataset have duration under 12,1 seconds, 95% (within 2 standard deviations of the mean) have duration under 6.5 seconds, 68% (within 1 standard deviation) under 2.6 s and 50% under 2 s.

From the events' extraction algorithm, mean values of the deceleration events were calculated. The distribution of the mean values of the final events is presented in the histogram that follows, Figure 5.22.

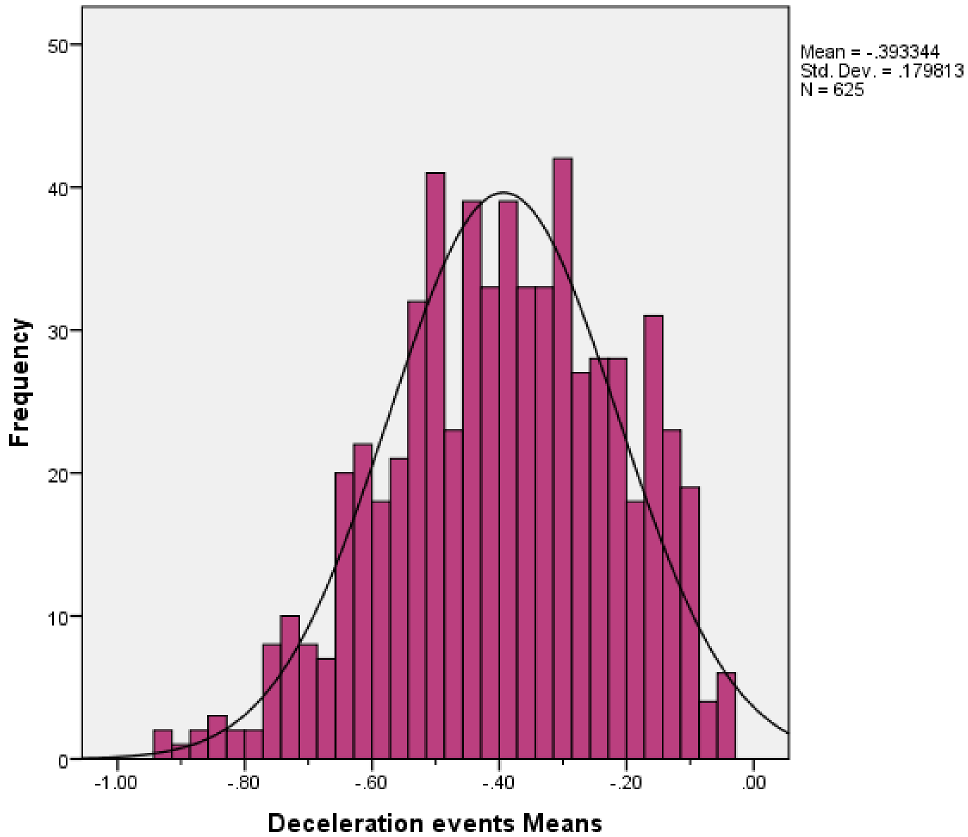


Figure 5.22 Histogram of the means of deceleration events

Most of the deceleration means seem to be between -0.3g and -0.55g with an average value of -0.4g. Plotting the cumulative distribution frequency of the means indicates that only 12% exceeds -0.6 g and this can be associated with the fact that the dataset contains many near-crashes. It would be interesting to examine the events according to their severity. Therefore, an analysis was conducted by separating them to crashes and near-crashes.

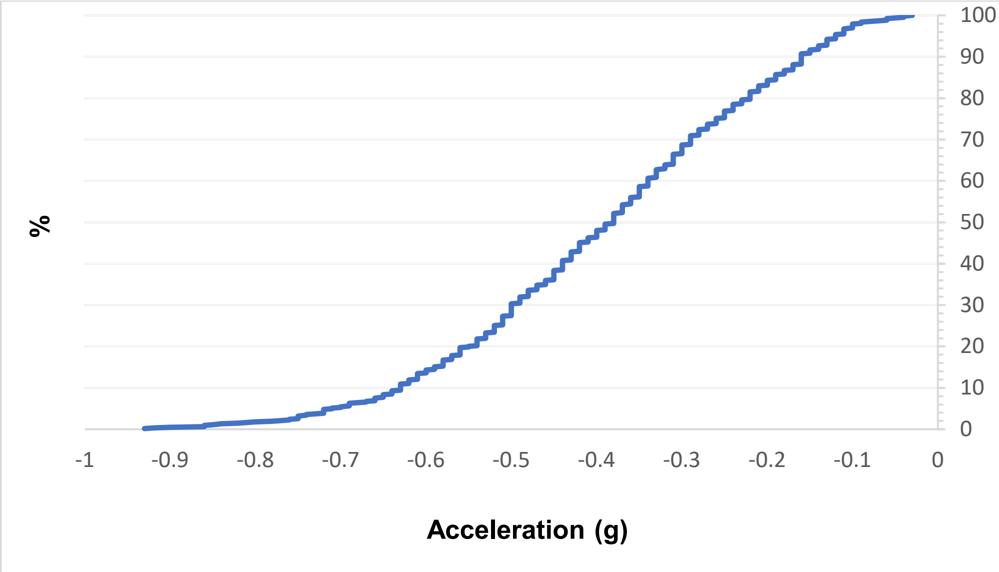


Figure 5.23 Cumulative frequency of the mean values of deceleration events

5.5.1.2 <u>Initial (first) braking events</u>

The initial events, these that have been detected firstly (in timestamps far from the main incident) have been also explored. The initial deceleration events exhibit lower means than the finals but slightly higher duration with mean 2.63 s. This could lead to the conclusion that the final deceleration events are sharper and can also justify up to a point the initial assumption that the final events are the critical ones. Below, the distribution of the initial deceleration events duration and means are presented.

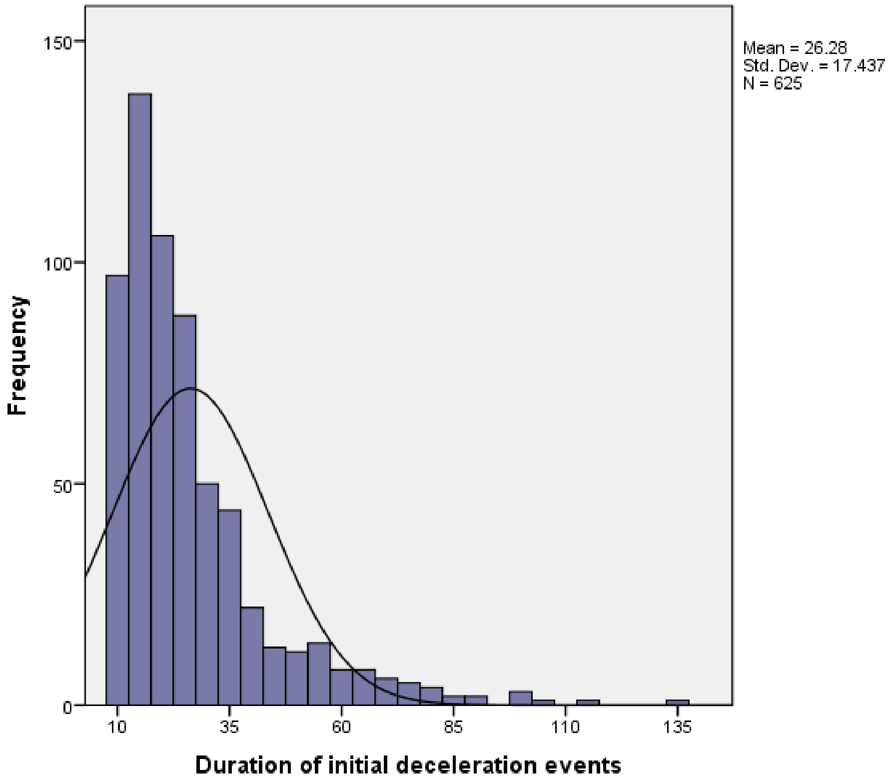


Figure 5.24 Histogram of initial deceleration events duration

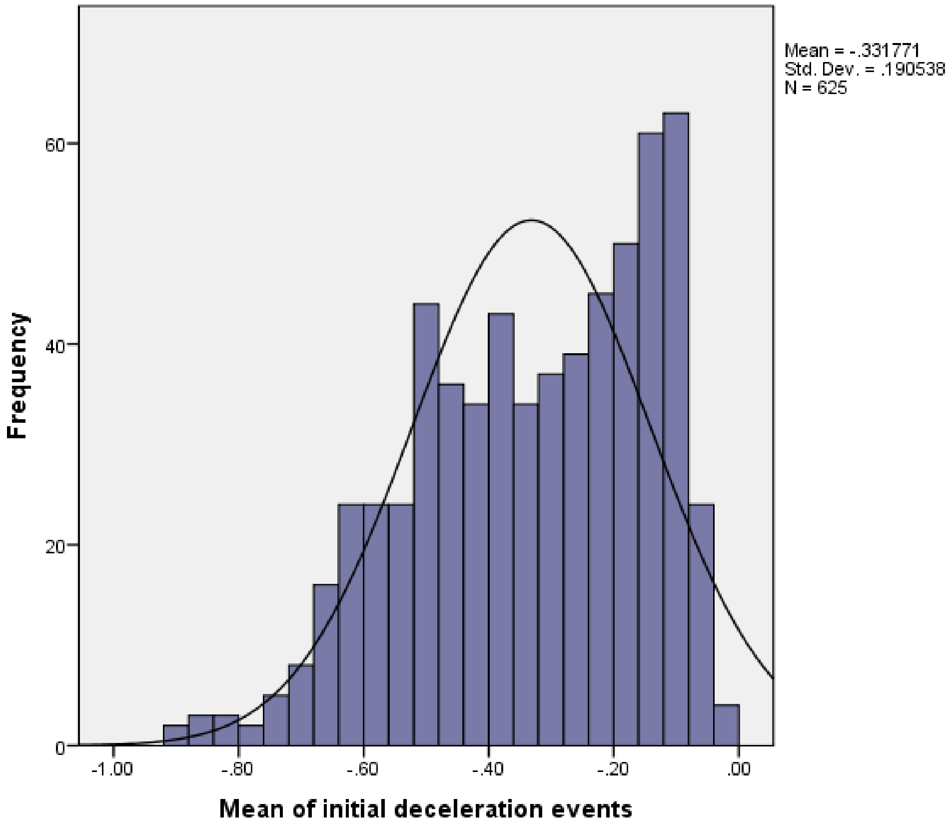


Figure 5.25 Histogram of initial deceleration events means

5.5.1.3 Exploring differences between Crashes and Near-Crashes

The event duration, the deceleration events mean and min value, and also main demographic information are presented in Table 5.13.

Table 5.13 Descriptive statistics for crashes and near crashes

	Crashes	Near crashes
Event Duration	2.69s	2.53s
Deceleration Mean	-0.32g	-0.4g
Deceleration Min	-3.55g	-1.15g
Age group	43drivers≤ 24, 25≤23 drivers≤64,18drivers≥65	265drivers≤ 24, 25≤221 drivers≤64, 51drivers≥65
Gender	40 F - 45 M	281 F – 256 M

The event duration in crashes is higher than this of near-crashes and surprisingly near-crashes exhibit higher deceleration values than crashes judging from the mean. In terms of crashes 51.8% of the drivers are under 24 years old and similar is the situation in near-crashes where 49.8% belong to this age group. Regarding the gender, women have more near-crashes from men than crashes.

5.5.1.4 Detecting deviation from normal driving

The percentile values of event duration and the deceleration event mean of all incidents are presented in Table 5.14.

Table 5.14 Percentile values of event duration and deceleration event means

Percentile	Event duration	Deceleration Events Means
25%	1.3s	-0.52g
50%	2s	-0.38g
68%	6.5s	-0.3g
85%	3.9s	-0.19g
95%	6.5s	-0.12g
99.7%	12.1s	-0.04g

50% of the deceleration events have duration less than 2s and 50% of the deceleration events means are below -0.38g, therefore the concurrence of these two thresholds can be identical of a critical situation. In other words, as 50th percentile represents the median, it can be supported that a critical situation occurs when the deceleration is lower than -0.38g for more than 2 s.

It is useful to investigate whether these thresholds change in case incident type and severity have been considered.

Table 5.15 shows how the percentiles change when only crashes are considered while Table 5.16 when only near-crashes are to be explored.

Table 5.15 Percentile values of event duration and deceleration event means for crashes

Percentile	Event duration	Deceleration Events Means
25%	1.35s	-0.42g
50%	**2s**	**-0.28g**
68%	2.85s	-0.19g
85%	4.4s	-0.13g
95%	7.19s	-0.061g
99.7%		

Table 5.16 Percentile values of event duration and deceleration event means for near-crashes

Percentile	Event duration	Deceleration Events Means
25%	1.3s	-0.52g
50%	**2s**	**-0.4g**
68%	2.6s	-0.31g
85%	3.89s	-0.2g
95%	6.3s	-0.13g
99.7%	12.2s	-0.05g

Moreover, percentile values of event duration and deceleration event mean by incident type are presented in Table 5.17. All the rear-end striking incidents appear to have larger duration than road departure and sideswipe ones while the latter present the lowest. The sample, as it has already been mentioned, is biased towards rear-end striking near-crashes. They exhibit higher values both in duration and deceleration means values than the sideswipe ones, showing that these events last more and present more extreme values. Regarding the rear-end crashes, they are not enough to derive safe conclusions, but relating to road departure crashes, all the percentile values have higher values with the median being -0.61g and 2.3s duration while the relevant of road departure crashes is -0.2g and 1.9s respectively. The road departure near-crashes and the sideswipe crashes are only 4, rendering any conclusions impossible.

Table 5.17 Percentile values of deceleration events mean and duration by Incident type

Incident type	Rear-end, striking	Road departure	Sideswipe	Rear-end, striking/ Crashes (13)	Road departure/ Crashes (37)	Sideswipe/ Crashes (4)	Rear-end, striking/ Near-Crashes (384)	Road departure/ Near-Crashes (4)	Sideswipe/ Near-Crashes (83)
25th percentile	1.4s	1.3s	1.2s	1.3s	1.3s	1.7s	1.4s	1.23s	1.2s
	-0.52g	-0.3g	-0.51g	-0.76g	-0.3g	-0.56g	-0.51g	-0.35g	-0.51g
50th percentile	2.2s	1.9s	1.6s	2.3s	1.9s	3.2s	2.2s	1.3s	1.6s
	-0.4g	-0.22g	-0.37g	-0.61g	-0.2g	-0.29g	-0.4g	-0.31g	-0.37g
68th percentile	2.8s	2.91s	2s	2.8s	3.08s	7.4s	2.8s	1.74s	1.9s
	-0.32g	-0.16g	-0.24g	-0.51g	-0.16g	-0.14g	-0.32g	-0.21g	-0.24g
85th percentile	4.26s	5.34s	3.32s	3.8s	5.46s	-	4.4s	-	3s
	-0.22g	-0.12g	-0.16g	-0.21g	-0.12g		-0.22g		-0.16g
90th percentile	4.94s	6.48s	3.7s	5.4s	6.84s	-	5s	-	3.56s
	-0.17g	-0.1g	-0.15g	-0.19g	-0.1g		-0.17g		-0.15g
95th percentile	6.81s	8.84s	5.48s	-	9.1s	-	6.88s	-	4.38s
	0.13g	-0.1g	-0.11g		-0.09g		-0.13g		-0.12g
99.7th percentile	13s	-	-	-	-	-	13.1s	-	-
	-0.05g						-0.05g		

The duration of deceleration events at the different incident types differs statistically at a 95% confidence level (median test, $p<0.0001$). Regarding the gender, the age and the incident severity no statistical differences were detected.

The deceleration means proved to be normally distributed and have also a statistically significant difference among the different incident types. Furthermore, they differ statistically between crashes and near-crashes (t-test, $p<0.0001$) but not across gender and age.

In an attempt to combine the information of all the above tables, the following one has been created.

Table 5.18 50th percentile values for event duration and mean by event type and severity

50th percentile	Events by type & severity	Duration	Deceleration event mean
	Crashes	2s	-0.28g
	Near-Crashes	2s	-0.4g
	All	2s	-0.38g
	Rear-end, striking	2.2s	-0.4g
	Road departure	1.9s	-0.22g
	Sideswipe	1.6s	-0.37g
	Total Mean	1.95s	-0.34g
	Total Median	2s	-0.37g

In this table, the critical combinations for incident severity and incident type have been identified. Calculating the total median (mean is biased towards the extreme values of the sample) the result is very similar with the initial where all the incidents have been considered together as it was expected.

5.5.1.5 Pre-event deceleration driver behaviour and corresponding TTC-some examples

In the following graphs the deceleration events 30 seconds before and until the main event are presented. The pre-crash behaviour has been categorised regarding the number of deceleration events a driver had, therefore there are up to six categories. One crash and one near-crash of each one has been plotted as a visualisation of the pre-crash deceleration driver behaviour. Together with the deceleration plots, the corresponding TTC values are depicted to visually check the extend that these two indicators relate.

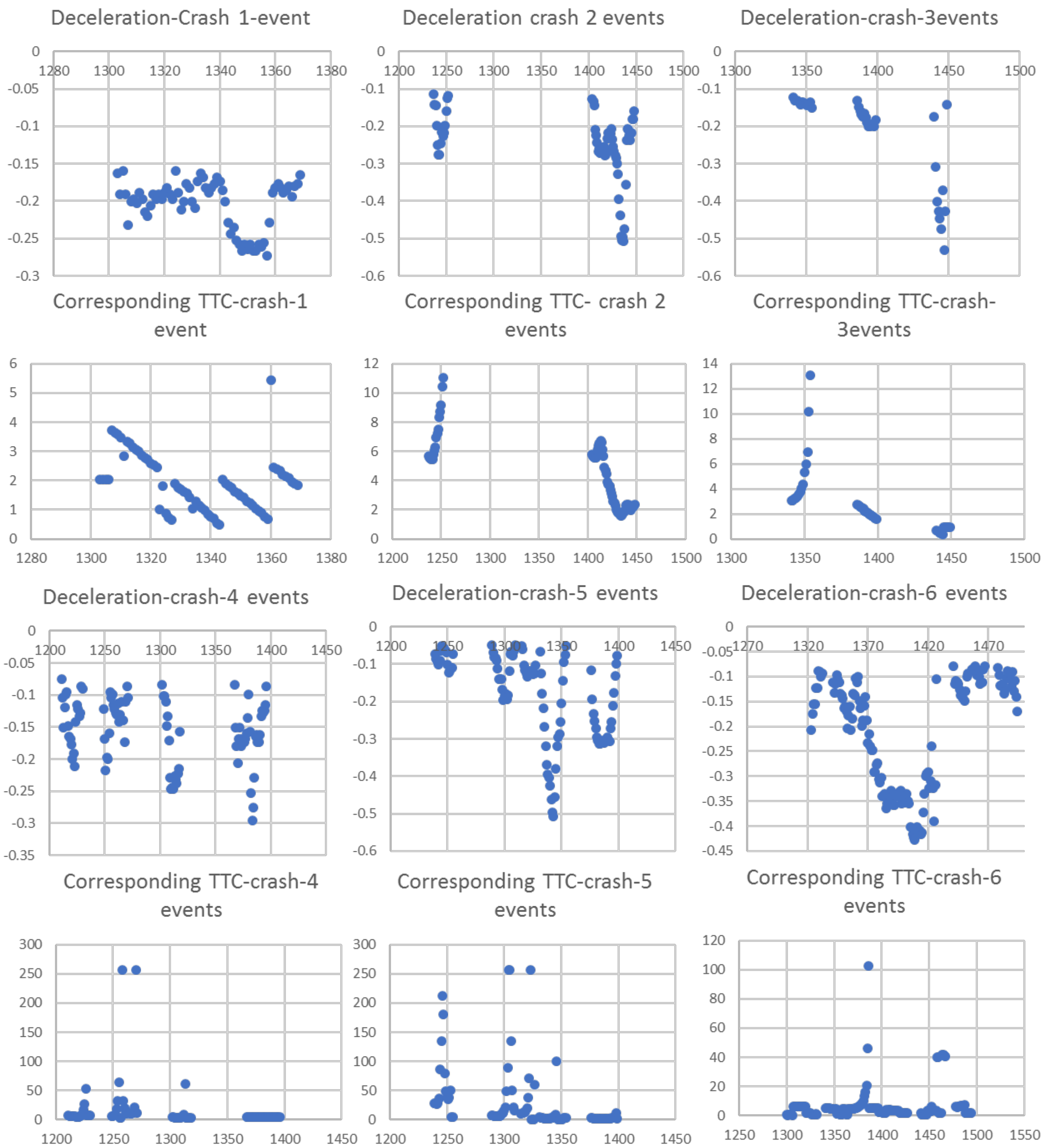


Figure 5.26 Pre-event braking behaviour- Crashes

As it can be observed from the plots, most of the times TTC values appear to comply with the deceleration events as they are decreasing respectively. Moreover, in some cases the final deceleration event does not exhibit the most extreme values. For instance, in the case of the crash that has 4 other braking events before the final, the fourth was the one that presents the lowest value (maximum absolute deceleration value). A similar situation can someone observe in the case of near-crashes where 3 out of 6 (in the examples presented in Figure 5.27) exhibit the most severe-extreme braking as a final event while the other 3 of them present the highest deceleration values earlier.

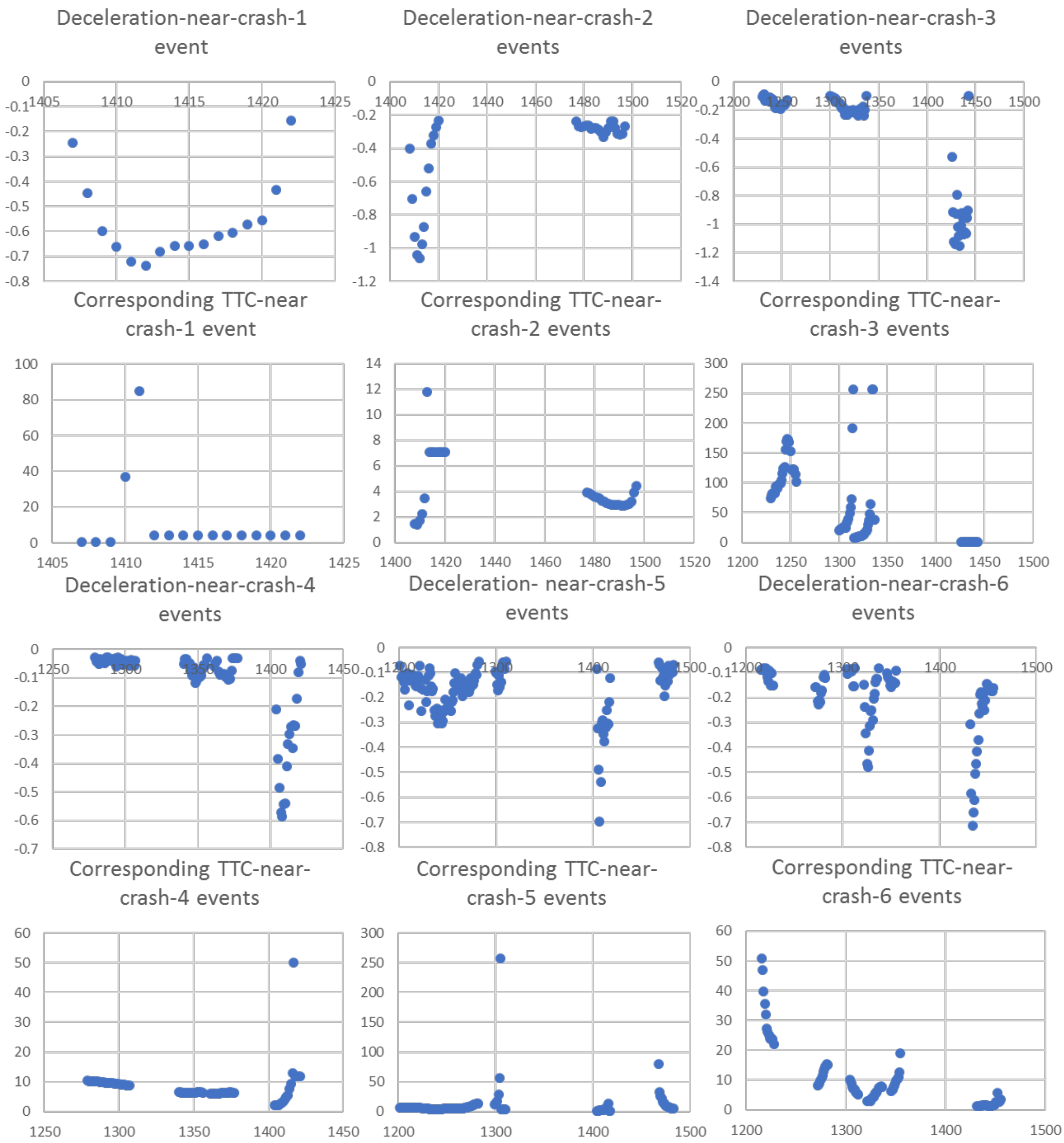


Figure 5.27 Pre-event braking behaviour - Near-Crashes

The plots are not following the same scale as they are pure representations of individuals event values that are derived from different incident types; therefore, the values can exhibit a large range. The purpose here, is to visualise the pre-incident events and generate a general impression of the driver pre-incident manoeuvres.

5.5.2 Yaw rate

The data extraction process regarding yaw rate was similar with this described in deceleration subsection, but the absolute values were used and instead of 1s, the observations should be exceeded consistently the threshold for 0.7 s (7 timestamps). The reason for this decision is that it probably takes less time for someone to steer than to brake. 1 sec (10 timestamps), 0.8 sec (8 timestamps) and 0.5 sec (5 timestamps) were checked as well, but 0.7 sec was decided to be the optimum choice for this analysis. The cases presented in Table 5.19 have been classified according to the results of the algorithm implementation.

Table 5.19 A summary of Driver yaw rate behaviour 30 seconds before main incident

Cases	Trips	Event severity	Gender	Age-Group	Incident type
No evasive reaction - the threshold was not exceeded at any point	437	47 crashes, 390 near-crashes	238 F 197 M	207drivers≤24, 24>184drivers<65, 46 drivers≥65	316 rear-end striking, 9 road departure, 59 sideswipe, 53 other
1 yaw rate event (critical)	125	21 crashes, 104 near crashes	61 F 63 M	68 drivers≤24, 24>44drivers<65, 12 drivers≥65	69 rear-end striking, 9 road departure, 22 sideswipe, 25 other
2 yaw rate events	87	14 crashes, 73 near crashes	44 F 43 M	33 drivers≤24, 24>39 drivers <65, 15 drivers≥65	40 rear-end striking, 12 road departure, 18 sideswipe, 17 other
3 yaw rate events	55	21 crashes, 34 near crashes	27 F 28 M	27 drivers≤24, 24>21drivers <65, 7 drivers≥65	41 rear-end striking, 6 road departure, 17 sideswipe, 12 other
4 yaw rate events	38	15 crashes, 23 near-crashes	20 F 18 M	22 drivers≤24, 24>11drivers <65, 5 drivers≥65	17 rear-end striking, 3 road departure, 2 sideswipe, 12 other
5 yaw rate events	9	2 crashes 7 near-crashes	4 F 5 M	5 drivers≤24, 24>2 drivers <65,	6 rear-end striking, 2 sideswipe, 1 other

Cases	Trips	Event severity	Gender	Age-Group	Incident type
				2 drivers≥65	
6 yaw rate events	10	5 Crash 5 Near-crash	2 F 8 M	6 drivers≤24, 24>2 drivers <65, 2 drivers≥65	1 rear-end striking, 1 road departure, 2 sideswipe, 2 other
7 yaw rate events	6	1 crash 5 near-crashes	2 F 4 M	3 drivers≤24, 24>3 drivers <65	Road departure
8 yaw rate events	1	Crash	F	16-19	Rear-end, striking
9 yaw rate events	1	Near-crash	F	20-24	2 rear-end, striking
11 yaw rate events	2	2 Near-crashes	2 M	30-34, 40-44	Rear-end, striking
26 yaw rate events	1	Near-crash	F	16-19	Rear-end, striking
27 yaw rate events	1	Near-crash	F	20-24	Rear-end, striking
30 yaw rate events	1	Near-crash	F	20-24	

The trips that have more than 8 events are rear-end, striking near-crashes involving young women except for the 2 trips that 11 events have been identified where two men from 30 to 44 were involved. It is worthwhile mentioning that in 437 cases out of 774 the drivers did not perform any evasive yaw rate manoeuvre. This fact is definitely affected by the type of the incident. In most of the cases, only one yaw rate event occurred that was expectably the final and critical one. It is impressive though that in 3 cases, drivers had 26, 27 or even 30 Yaw rate events in a 30second time period.

5.5.2.1 Final yaw rate events

Similar to the deceleration events, there is the assumption that the final events (above 3 standard deviations for 0.7s) are the critical ones. The process that was followed was exact the same as for the deceleration. The absolute values of yaw rate were used. The duration in the diagrams are in timestamps, therefore 10 timestamps= 1s.

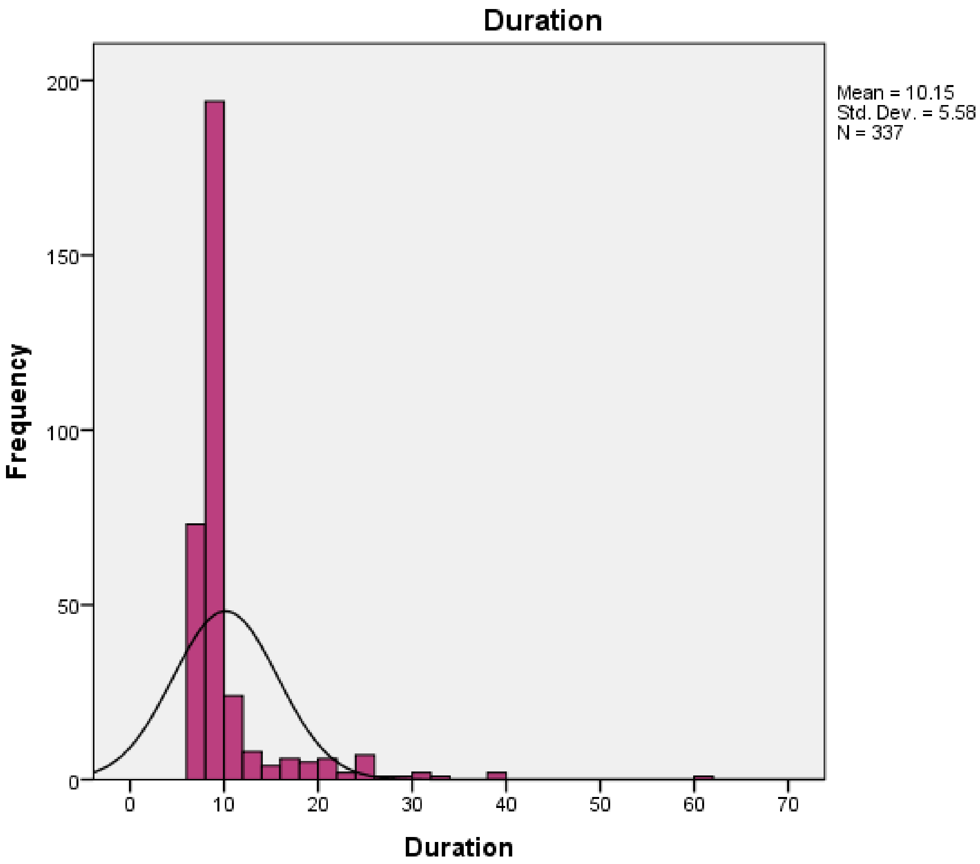


Figure 5.28 Histogram of yaw rate events duration

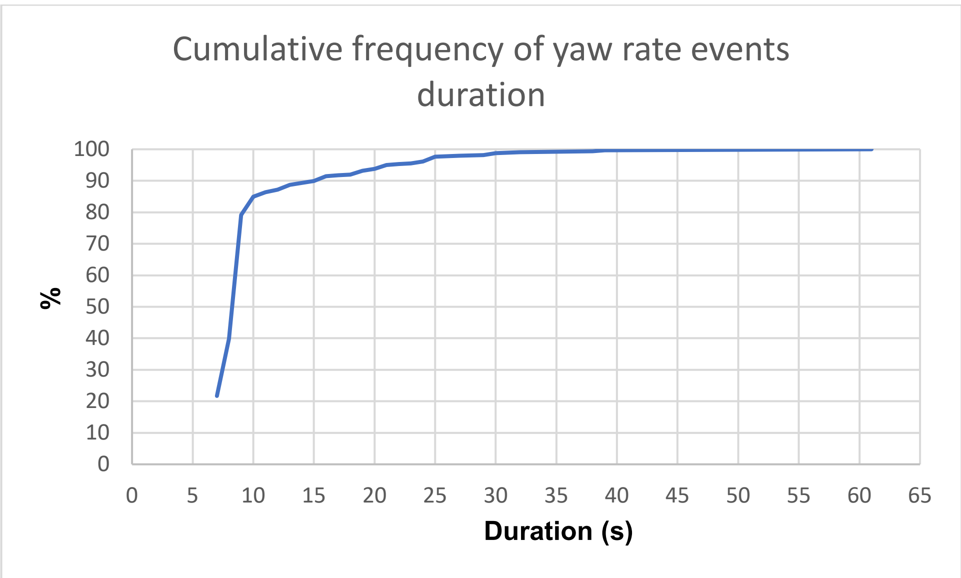


Figure 5.29 Cumulative frequency of yaw rate events duration

99.7% (within 3standard deviations of the mean) of the events in the dataset have duration under 6.7 seconds, 95% (within 2 standard deviations of the mean) have duration under 2.2 seconds, 68% (within 1 standard deviation) under 0.9 seconds and 50% under 0.9 seconds. The distribution of the mean values of the final events is presented in the histogram that follows, Figure 5.30.

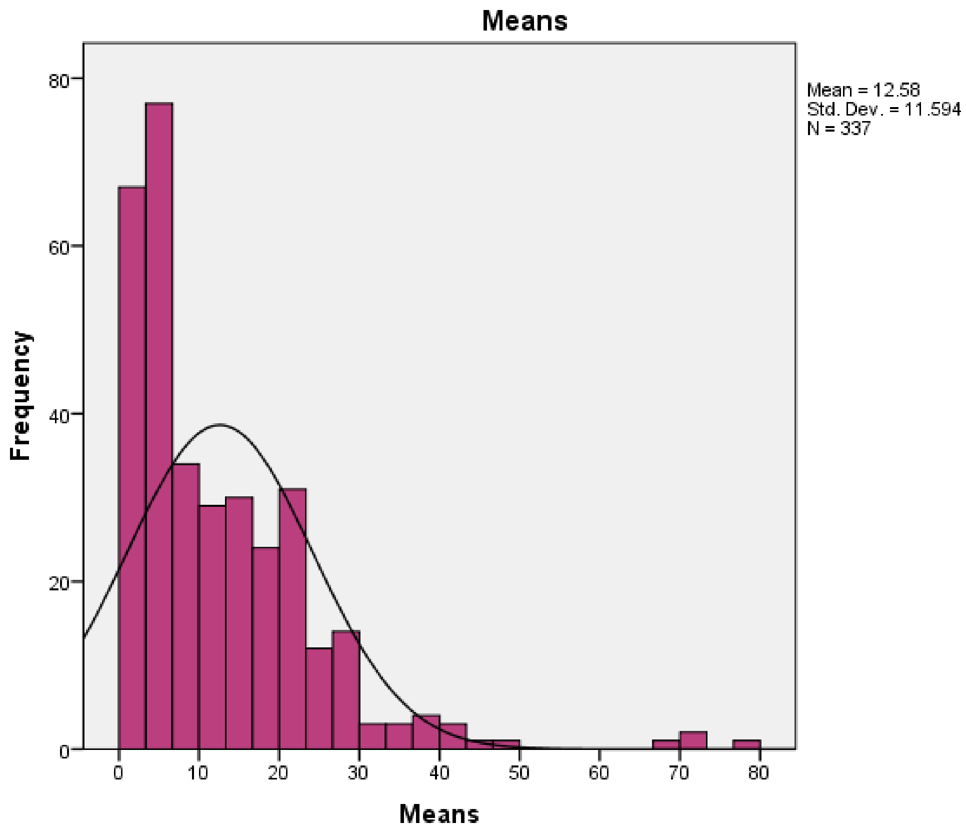


Figure 5.30 Histogram of yaw rate events means

Most yaw rate means seem to be between 0.7 and 10°/s with a mean of 12.6°/s. The maximum value is 78.74°/s and from plotting the cumulative distribution frequency of the yaw rate events means less than 5% of them exceeds 30°/s. It would be again interesting to examine the events according to their severity and type.

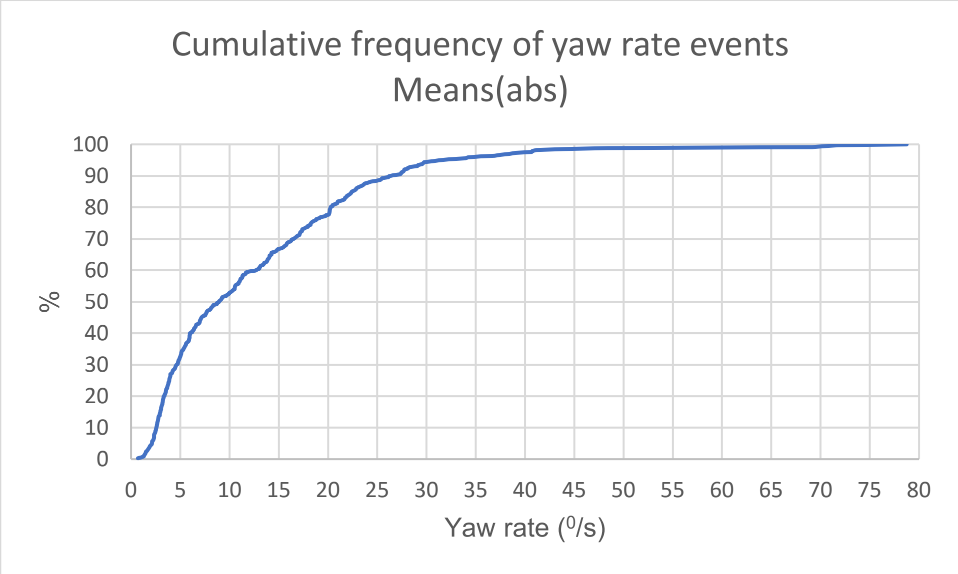


Figure 5.31 Cumulative frequency of yaw rate events Means

5.5.2.2 Initial yaw rate events

Exploring the initial yaw rate events, there were no significant differences detected except for a slight one in duration of the events. The final events appear to be longer than the initial ones (duration mean = 9.48s) while the yaw rate event mean is 12.46^0/s, almost the same with the corresponding of the final events.

5.5.2.3 Exploring differences between Crashes and Near-Crashes (regarding yaw rate events)

The event duration, the yaw rate event mean, and main demographic information are presented in Table 5.20.

Table 5.20 Descriptive statistics for crashes and near crashes

	Crashes	Near crashes
Duration	1.1s	0.9s
Yaw rate Mean	21.55 °/s	9.79 °/s
Max	71.8 °/s	78.7 °/s
Age group	45 drivers≤ 24, 25≤19 drivers≤64,16 drivers≥65	124 drivers≤ 24, 25≤108 drivers≤64, 25 drivers≥65
Gender	40 F - 40 M	125 F –131 M
Incident type	60 road departure, 5 rear-end striking, 4 sideswipe, 11 other	144 rear-end striking, 53 sideswipe, 4 road departure, 56 other (turn into path,etc)

Similarly to deceleration events, the yaw rate event duration in crashes is higher than this of near-crashes and again near-crashes exhibit higher yaw rate values than crashes judging from the mean. 57.5% of the drivers involved in a crash are under 24 years old and in near-crashes 48.6% belong to the same age group. In terms of the gender, there are not any noticeable differences. 75.1% of the crashes are road departure incidents as expected while in near-crashes only the 1.6% belong to this incident type. The majority of the near crashes (56%) are rear-end, striking incidents.

The duration of the events differs significantly (median test, $p<0.0001$) in near-crash and crash incidents, but not throughout the age groups, gender and incident type. Means do not differ statistically in the above categories.

5.5.2.4 Detecting deviation from normal driving

The percentile values of yaw rate event duration and the yaw rate events means of all incidents are presented in Table 5.21.

Table 5.21 Percentile values of event duration and yaw rate event means

Percentile	Event duration	Yaw rate Events Means
25%	0.8s	3.9 °/s
50%	0.9s	8.9 °/s
68%	0.9s	15.8 °/s
85%	1.1s	22.5 °/s
95%	2.2s	32.5 °/s
99.7%	6.1s	78.6 °/s

Following the same reasoning as for deceleration events analysis, 50% of the yaw rate events have duration less than 0.9s and 50% of the yaw rate events means are below 8.9°/s, therefore the concurrence of these two thresholds can be identical of a critical situation. In other words, as 50th percentile represents the median, it can be supported that a critical situation occurs when the yaw rate is lower than |8.9|°/s for more than 0.9 s.

It is useful again to investigate whether these thresholds change in case incident type and severity have been considered. Table 5.22 shows how the percentiles change when only crashes are considered while Table 5.23 when only near-crashes are to be explored.

Table 5.22 Percentile values of event duration and yaw rate event means for crashes

Percentile	Event duration	Deceleration Events Means
25%	0.8s	13.9 °/s
50%	**0.9s**	**19 °/s**
68%	1s	24.37 °/s
85%	1.5s	30.51 °/s
95%	2.5s	41.16 °/s
99%	-	-
99.7%	-	-

Table 5.23 Percentile values of event duration and yaw rate event means for near-crashes

Percentile	Event duration	Yaw rate Events Means
25%	0.7s	3.3 °/s
50%	**0.9s**	**6.2 °/s**
68%	0.9s	10.79 °/s
85%	1s	19.89 °/s
95%	2.1s	27.4 °/s
99%	3.84s	45.8 °/s
99.7%	-	-

Moreover, percentile values of event duration and yaw rate events means by incident type are presented in Table 5.24. All the Road departure incidents have higher duration than the sideswipe and rear-end, striking ones while the latter present the lowest. The yaw rate means of road departure incidents also show that they not only last longer, but they exhibit higher values during their occurrence. In near crashes the duration does not present remarkable differences, but the mean values are higher in road departure and sideswipe near crashes. It should be mentioned though that the rear-end, striking and sideswipe crashes and the road departure near-crashes are very limited, therefore any conclusion could be invalid.

Table 5.24 Percentile values of yaw rate events means and duration by Incident type

Incident type	Rear-end, striking (149)	Road departure (57)	Sideswipe (57)	Rear-end, striking/ Crashes (5)	Road departure/ Crashes (60)	Sideswipe/ Crashes (4)	Rear-end, striking/ Near-Crashes (144)	Road departure/ Near-Crashes (4)	Sideswipe/ Near-Crashes (53)
25th percentile	0.8s	0.9s	0.7s	0.75s	0.9s	0.7s	0.8s	0.9s	0.7s
	3.18 °/s	16.21 °/s	3.34 °/s	7.8 °/s	15.68 °/s	8.28 °/s	3.12 °/s	33.12 °/s	3.23 °/s
50th percentile	0.9s	0.9s	0.9s	0.9s	0.9s	0.75s	0.9s	0.9s	0.9s
	5.38 °/s	21.39 °/s	7.71 °/s	17.34 °/s	20.7 °/s	13.13 °/s	5.13 °/s	36.49 °/s	6.4 °/s
68th percentile	0.9s	0.94s	0.9s	1s	0.97s	1s	0.9s	1.1s	0.9s
	7.8 °/s	27.54 °/s	11.42 °/s	23.4 °/s	25.95 °/s	17.47 °/s	7.6 °/s	40.1 °/s	11.25 °/s
85th percentile	1s	1.9s	1s	-	1.9s	-	1s	-	1s
	13.84 °/s	37.81 °/s	17.76 °/s	-	33.6 °/s		13.49 °/s		17.1 °/s
90th percentile	1.3s	2s	1.7s	-	2s	-	1.3s	-	1.8s
	17.38 °/s	40.7 °/s	20.75 °/s	-	42.7 °/s		15.32 °/s		21.34 °/s
95th percentile	2s	2.5s	2.5s	-	2.5s	-	2.1s	-	2.6s
	21.45 °/s	46.47 °/s	24 °/s		49.57 °/s		20.83 °/s		25 °/s
99.7th percentile	-	-	-	-	-	-		-	-

Combining the elements of the previous tables, Table 5.25 is presented below:

Table 5.25 50th percentile values for event duration and mean by event type and severity

50^{th} percentile	Events by type & severity	Duration	Yaw rate event mean
	Crashes	0.9s	19 º/s
	Near-Crashes	0.9s	6.2 º/s
	All	**0.9s**	**8.9 º/s**
	Rear-end, striking	0.9s	5.38 º/s
	Road departure	0.9s	21.39 º/s
	Sideswipe	0.9s	7.71 º/s
	Total Mean	0.9s	11.43 º/s
	Total Median	0.9s	8.3 º/s

In this table, the critical combinations for incident severity and incident type have been identified as exactly for deceleration events. The result again is very similar with the initial where all the incidents have been considered together.

5.5.2.5 Pre-event yaw rate driver behaviour and corresponding TTC-some examples

In the following graphs the yaw rate events 30 seconds before and until the main event are presented. The pre-crash behaviour has been categorised regarding the number of yaw rate events a driver had, and some representative ones are presented. One crash and one near-crash of each one has been plotted as a visualisation of the pre-crash yaw rate driver behaviour. Together with the yaw rate plots, the corresponding TTC values are depicted to visually check the extend that these two indicators relate (Figure 5.32). The x axis represents the timestamp (1/10 of a second) and y axis the TTC measured in seconds.

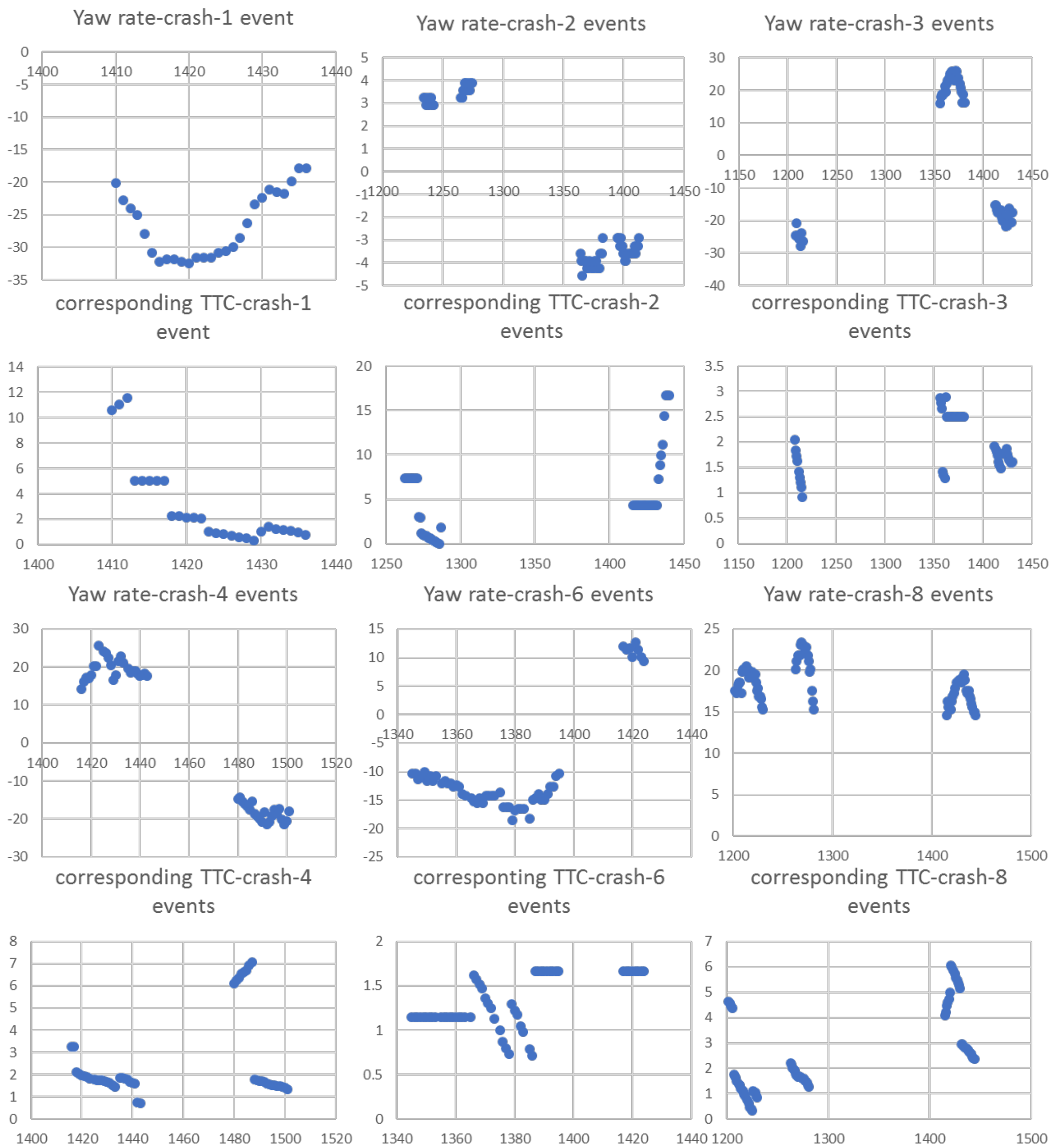


Figure 5.32 Pre-incident steering behaviour- Crashes

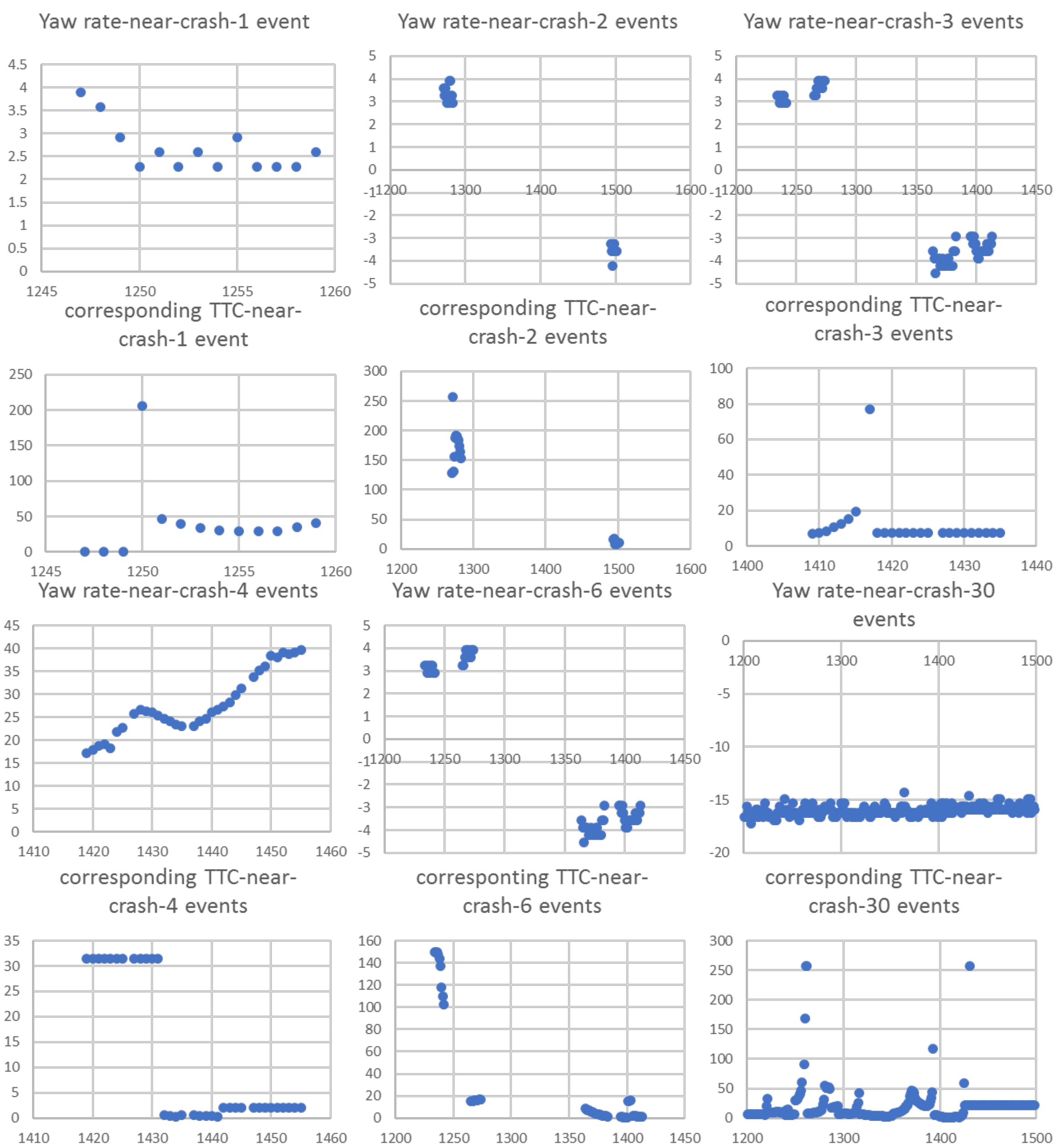


Figure 5.33 Pre-incident steering behaviour- Near-Crashes

The plots are not presented in the same scale similarly to these of the deceleration analysis.

5.5.3 Overlapping Pre-event manoeuvres

In some trips the braking and steering events overlapped and the order with which the driver reacted was investigated. The trips where there was an overlap of braking and steering events were chosen. Then, the dataset was merged with demographics and event detail data to generate the final dataset for analysis.

In 89 trips there was an overlap in braking and steering events -the driver braked and steered simultaneously beyond the threshold. All the overlapping events started with an exceeded threshold in deceleration, then both yaw rate and deceleration presented exceeded thresholds and finally the yaw rate was beyond 3 st.deviations. In other words, in these events the drivers were braking as first reaction, then were steering simultaneously and in the end, they were only steering beyond the threshold. It is interesting to explore these events regarding their type and severity that may affect the outcome. Table 5.26 presents the number of events by severity, type and vehicle classification along with basic demographic information.

Table 5.26 Number of incidents by severity and type, vehicle classification, age group and gender.

Incident Severity	Frequency	Percentage (%)
Crash	24	27.0
Near-Crash	65	73.0
Vehicle classification		
Car	65	73.0
Pickup Truck	7	7.9
Suv_crossover	17	19.1
Gender		
Female	42	47.2
Male	47	52.8
Age group		
16-24	46	51.7
25-64	35	39.3
65-89	8	9
Incident type		
Rear-end, striking	37	41.6
Road departure (left or right)	23	25.8
Sideswipe, same direction (left or right)	15	16.9
Turn into path (opposite or same direction)	7	7.9
Turn across path	2	2.2
Animal-related	1	1.1
Opposite direction (head-on or sideswipe)	1	1.1
Other	3	3.3

The table gives the picture of the sample, as near-crashes, cars, young drivers and rear-end, striking incidents are overrepresented. From 89 incidents though, 23 are road departure where the driver’s first reaction was braking and not steering. The same probably applies to the 15 sideswipe incidents. That reveals that drivers in these cases, even in incidents that someone would expect the opposite, they choose to brake first, maybe subconsciously trying to reduce the speed of the impact.

5.6 TTC modelling results

As analytically presented in the methodology chapter (Chapter 3), TTC values are analysed by using mixed effects hierarchical regression models. The predictor variables are gender, age, vehicle type, driven miles per year, longitudinal acceleration (both negative and positive values), lateral acceleration, yaw rate, speed, and time expressed with the timestamp and its quadratic term, timestamp squared, variable that was derived later for analysis purposes. This additional transformation constitutes an important choice as the aim is to investigate the evolution of TTC values within the crash or near crash development and detect any earlier change to them that could be associated with the event. Thus, a non-linear relationship between TTC values and time should be also examined. Different specifications for the predictor variables have been attempted during the model development including speed squared and log-transformations of acceleration, speed and timestamp. Traffic density was not eventually considered in the model estimation as it was not provided for every data point, but only for the moment of the event, so it would not have an effect in the evolution of TTC values.

Initially, an unconstrained (null) model was employed to check if there is a variability in the TTC values by trip. In other words, to confirm that there is a trip effect in TTC values, so the use of mixed effect model is supported against a simple regression analysis. The results indicate that there is overwhelming evidence at the 0.05 significance level, rendering this type of model appropriate for the analysis.

A random slope model was finally estimated with one explanatory variable (vehicle type) at level-2 and 4 variables (longitudinal acceleration, speed, timestamp and timestamp^2) at level-1. The aforementioned variables proved to be affecting the TTC values in a statistically significant way. On the contrary, yaw rate, lateral acceleration, age, gender and miles driven last year were not statistically significant and they were excluded from the model. The variables were also examined to show if their effects (slope coefficient) varies across the trip-files. The variance associated with slope coefficient of longitudinal acceleration, speed, timestamp and timestamp squared were found to be statistically significant. The results of the multilevel model are presented in Table 5.27 and reveal that the TTC values evolution during crash sequence, therefore from a normal driving situation until the event, is affected by vehicle type, acceleration, speed and time within crash development, where there is a timepoint when TTC values start to reduce until the final event.

The estimate of the residual variance represents the variability of the observations by trip around the regression lines. More specifically, if one best-fit line is drawn for a trip, their actual measurements would randomly vary around this line with about 95% of the values falling within 120.114 (i.e. square root of 3606.85 * 2) of the line.

As mentioned before, the vehicle classification proved to be significant and the Van&minivan category has been selected as reference. It can be observed that for all the other vehicle types the TTC values are significantly higher. Regarding the vehicle kinematics variables, longitudinal acceleration and speed of the ego-vehicle have been found to be randomly distributed with their coefficients value varying by trip.

Since the variance of the speed coefficient is quite large (i.e. 0.549) in relation to the mean of the speed coefficient, there is a high possibility that some values of the coefficients would be positive. Additionally, since the speed coefficient has assumed to follow a normal distribution, the mean is -0.2305 and the variance is 0.549647, then Z-statistic can be obtained to calculate the area under the normal curve between the mean and 0. This is given by:

$$Z = \frac{0 + 0.2305}{\sqrt{0.549647}} = 0.31$$

Z=0.31 represents a 12.2% of the area under the normal curve. This means that 50% + 12.2% = 62.2% of the speed coefficients show a negative sign (i.e. they are negatively associated with TTC) and 37.8% of the coefficients exhibit a positive sign (i.e. they are positively associated with TTC). The speed of ego vehicle has a mixed-effect on the TTC indicating that driver speeding behaviours play a key role in their TTC values.

Similarly, the variance of longitudinal acceleration coefficient is also large compared to the longitudinal acceleration slope itself. As the mean is 42.815 and the variance 4696.035, the z-score is -0.63. This represents a 26.43% of the area under the normal curve and shows that 73.6% of the longitudinal acceleration slopes affect positively the TTC values while 26.4 % of them have a negative effect (make values lower-riskier).

Timestamp variable that represents the time to a crash or a near crash (i.e. crash progression sequence starting 2.5 minutes before the event) has assumed to have a quadratic relationship with TTC. We therefore hypothesized that the TTC values just before the crash would be ‘low’ in relation to other TTC values and it would be interesting to see at which point in time the TTC values start to decline.

To achieve that, the Fermat’s Theorem has been employed to identify the possible extremum of the quadratic TTC function. The first-time derivative of the function is: $\frac{d(TTC)}{dt} = \beta_1 + 2\beta_2 t$. Therefore, the possible extremum will be the point within the function where $\frac{d(TTC)}{dt}$ = 0, in which β_1 is the coefficient for the linear timestamp term, b_2 is the coefficient of the timestamp squared and t is the timestamp. The critical timestamp is found to be t = 526.

Moreover, the second-time derivative of the function is: $\frac{d^{2}(TTC)}{dt^2} = 2 * \beta$ = 2* (-0.0000238) < 0 indicating that this timestamp value represents the maximum point of the function. In other words, at timestamp 526, the functional value changes and the TTC values start to drop.

This critical timestamp value represents 1.62 minutes (on a timescale between 0 and 2.5 minutes) before the crash indicating that TTC values, on average, starts to decline about 1 minute before the crash, *ceteris paribus*.

Table 5.27 Multilevel mixed effects model

Dependent variable	**TTC**	
Fixed effect	**Coefficient**	**t-stat**
Speed	-0.230543	-8.450
Longitudinal Acceleration	42.8148	15.723
Vehicle type:		
Car	31.8363	13.687
Truck	29.63	10.102
SUV and Crossover	31.6881	13.057
VAN and minivan (reference)		
Timestamp	0.0250	5.268
Timestamp^2	-0.0000238	-7.797
Intercept	20.2940	8.838
Random effect parameters		
Variance of speed	0.549647	18.35
Variance of Longitudinal Acceleration	4696.0346	15.875
Variance of Timestamp	0.016768	19.042
Variance of Timestamp^2	6.879815E-9	18.999
Variance of Residual	3606.8504	639.270
Statistics		
Number of observations	1,161,584	
Number of groups	774	
Bayesian Information Criterion (BIC)	8949558	

Figure 5.34 presents the TTC within the crash progression and across different speeds according to the model developed. The diagram confirms the impact of speed to TTC values during the event sequence development. It is observed that in higher speeds, TTC values stand lower.

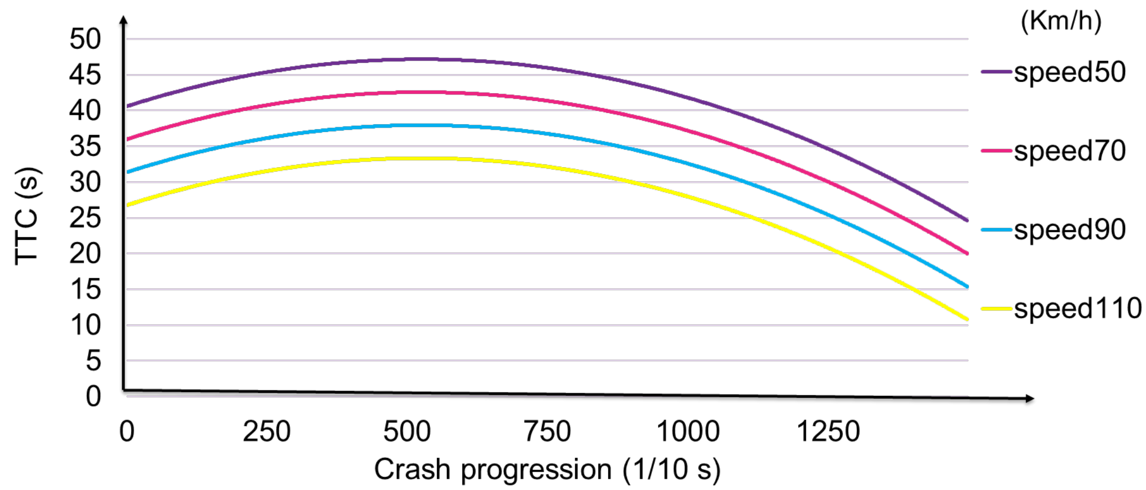


Figure 5.34 TTC evolution at different speeds

5.7 Summary

In this chapter, the analysis of the pilot diving study was presented, dynamic thresholds for quantification of normal driving were set, indicators for detecting deviations from normal driving were generated and the pre-event driver behaviour was explored. The significance and implications of these results will be discussed in the next chapter. During the course of this analysis, empirical and statistical tools described at the methodology chapter (3), were both used to address the research aim and questions (chapter 1) and provide results. The main findings are:

- The range of longitudinal acceleration in the pilot driving study for normal driving was -0.55g to +0.38g and the corresponding one for lateral acceleration -0.49g to +0.44g.
- Thresholds to quantify normal driving and detect deviations should be dynamic and can be given by linear regression equations of indicators' percentile values.
- Indicators generated by a process based on the 5th percentile of deceleration, lateral acceleration and yaw rate can help in detecting abnormalities in driving, while TTC does not provide a clear picture.

- TTC during crash sequence is affected by longitudinal acceleration, speed, vehicle type, and the time within crash progression.
- Investigating the pre-event driver behaviour, a critical situation can occur when the deceleration is -0.38g for more than 2s and accordingly, yaw rate 8.9^0/s for more than for 0.9s.
- Drivers first brake and then steer in events where both deceleration and yaw rate thresholds are exceeded.

6 Discussion and conclusions

This chapter presents a summary of the research conducted along with the findings produced, an examination of the results in relation to existing research and the implications of the research. The key findings and the relevant discussion are presented for each of the research objectives. Furthermore, this last chapter highlights the contribution to knowledge, cites the limitations of the study and provides recommendations for future research.

6.1 Summary of key findings and discussion

✓ *Objective 1 "To formulate a comprehensive theoretical framework of factors quantifying the transition from normal driving to a safety critical scenario and influencing crash risk".*

📖 *(§2.3.3, §2.3.4)*

The literature review focused on how previous research has investigated driver behavior and safety. Tools and metrics to determine compromising factors in driver safety have been identified, reviewed and categorized. Chapter 2 and in particular, sub-sections 2.3.3 and 2.3.4 discuss the relevant literature.

The theoretical framework (Figure 2.11) incorporates factors that influence the development of crash risk (traffic characteristics, environment, geo-demographic driver profile, human factors) and moreover, factors that can identify it and indicate crash proximity (including TTC, Time Headway, Yaw rate, Longitudinal and lateral acceleration).

Addressing this objective by reviewing the relevant literature supports: i) a consideration of the time element in crash development sequence, (e.g. TTC decreases with time) ii) a better understanding of the factors affecting the transition from a normal driving situation to a deviation and then to a safety critical event (a comprehensive set of risks), and ii) provided the knowledge for selecting the suitable indicators (preliminary selection of metrics) to examine the transition from normal driving to safety critical scenarios that is the scope of this thesis.

✓ *Objective 2 "To examine the potential contribution of NDS to an understanding of vehicle kinematics"*

🕮 *(Pilot study, §3.5.1, §5.2)*

A pilot study was designed and conducted involving 5 drivers driving for 30 minutes each. The study constituted a short trial and resulted in a considerable amount of data. This comprised 2.5 hours of normal driving along with the relevant video recordings and enabled the researcher to gain familiarity with NDS data collection and processing and gave insights for the range of two indicators, longitudinal and lateral acceleration.

A threshold for longitudinal acceleration was recommended, as one driver was involved in a safety relevant event that was easily identifiable from the data and videos and provided the opportunity for further investigation. A hard braking manoeuvre performed by a subject driver when a leading motorbike stopped unexpectedly, resulted in a longitudinal deceleration of -0.55g. The 100-car study (Dingus et al., 2006) and Naturalistic Teen driving study (Lerner et al., 2010) have set a thresholds of -0.6g and -0.65g respectively that are close to the threshold reported from the pilot study, although less conservative. The range of the longitudinal acceleration values in the study were - 0.55g to +0.38g with 99% of the observations not exceeding -0.24g while for the lateral acceleration were -0.49g to +0.44g with 99% of the observations not exceeding -0.22g. These results represent normal driving conditions and are in line with many previous studies that attempted to detect safety critical events setting kinematic thresholds (SHRP2 NDS, teen driver study, EuroFOT, 100-car-study).

This pilot study constituted a way to evaluate the potential contribution of NDS to this research and also, to learn the limitations that these studies entail. Defining the research questions and the specification of the recording equipment constitute an important part of an NDS design. The representativeness of the sample, the accuracy of the devices and the familiarity with the sensors type and output play an important role in NDS and aid in data quality control, better interpretation and greater generalisation power of the results. This pilot study contributed in the acknowledgement of the difficulties in NDS analysis and provided a sample in order to investigate ways of exploiting efficiently the data towards the aim of this PhD research.

✓ *Objective 3 “To characterise safety indicators during normal driving*

🕮 *(§5.3)*

A method to quantify normal driving was developed through analysis of SHRP2 NDS dataset. The data describing vehicle kinematics of the preceding 2.5 minutes before a crash or near-crash-event was requested from VTTI. This provided a large amount of data covering the period before the events and a methodological decision was made to isolate the first part of these data to produce baselines for every trip. Therefore data 2.5 minutes until 1.5 minute before each event were chosen and combined in a dataset representing normal driving conditions. Previous studies that attempted to detect SCE in NDS have set triggering kinematic thresholds and these were used as a basis for setting initial thresholds for deviation from normal driving. After a descriptive analysis of the dataset looking into extreme values and indicators' distributions and percentiles, initial normal driving thresholds were validated and proposed. Linear regression equations were also employed to provide more dynamic thresholds for normal driving including speed as an independent variable. Thresholds for departure from normal driving should not be considered as similar to the thresholds for triggering SCE as they correspond to a different stage of the crash sequence (see Figure 2.13, Chapter 2). Further examination of extreme values of indicators distributions (e.g.1st or 99th percentiles) showed that values of deceleration and lateral acceleration did not follow a clear pattern while yaw rate percentile values were decreasing with speed to sharply increase in speeds higher than 110km/h. The values of TTC decreased with speed up to a speed bin of 50-60km/h and increased for higher speeds. In accordance with these results, Kusano et al., (2015) and Chen et al., (2016) showed that driver behaviour is highly associated with vehicle speed and that TTC values increase with speed. The results could suggest that in case of high-speed driving scenarios, drivers may be more attentive and initiate braking earlier.

TTC, longitudinal and lateral acceleration and yaw rate distributions were found to be different between younger and other drivers and between female and male drivers during normal driving conditions. Women exhibit statistically significant lower TTC values than men and adult and older drivers higher than the young ones. The latter result supports the study by Kusano et al., (2015) where older drivers have been also found to underestimate TTC more than younger ones. This could be an evidence that older drivers avoid risky driving behaviour or are more experienced and apply the brakes earlier. Moreover, female and younger drivers present more extreme deceleration values than male and older drivers. Regarding their normal driving lateral behaviour, higher values of lateral acceleration and yaw rate have been observed for male and younger drivers. It is fairly discussed in the literature that female and male drivers perceive and evaluate risk differently (Harris et al., 2006), thus the differences in the indicators' distributions can be explained by women being more cautious or overreacting (harsh braking) on occasions. Gender and age differences in risk taking behaviour have been found in several studies before (e.g. Harré et al., 2000; Harré et al., 1996; Oltedal and Rundmo, 2006, Reason et al., 1990).

✓ *Objective 4 "To develop safety indicators for detecting deviations from normal driving"*

🕮 *(§5.4)*

In order to investigate the transition from normal driving to safety critical driving scenarios, indicators were generated to describe the whole crash or near-crash sequence. Exploring total (all events) raw values of deceleration, lateral acceleration, yaw rate and TTC was not capable to provide any clear indication of deviation or abnormality. Creating indicators from aggregated data where all the timepoints are represented enabled a check of the feasibility to detect deviations and provide a visualisation of the pre-crash conditions. The distributions and the descriptive statistics of the indicators helped in further understanding of the vehicle kinematics during the whole crash or near-crash sequence.

Deceleration, lateral acceleration, and yaw rate in crash data exhibited signs of abnormalities in the last seconds and especially lateral acceleration and yaw rate frequently showed deviations 15 seconds before the end of event. Deceleration did not prove to be the best indicator in this phase for detecting early deviations contradicting Dingus et al., (2006) (100-car study) results where it showed the best overall performance in detecting safety critical events with a success rate of 45% and a false rate of 66% against TTC that showed high success rate along with an even higher false rate though. Nevertheless, the deceleration false rate was still high and similarly, a high number of false alarms using deceleration as a marker for safety critical events was experienced by McLaughlin et al. (2008). Previous studies have also observed no statistically significant differences between evasive and normal driving braking (Nygård, 1999; Wahlberg, 2000; van der Horst, 1990; Várhelyi, 1998). The comparison of these studies with the current analysis could be misleading though, as the latter is mostly visual, but indeed the 5th percentile indicator plot did not show any detectable deviation especially in near-crashes prior the event. The corresponding TTC indicator did not follow any clear pattern and no conclusions could be drawn from the near-crash and crash data plots.

Exploring separately the female and male drivers, during the crash progression, the first ones present more extreme values for all the 5th percentile indicators. Longitudinal acceleration, lateral acceleration and yaw rate indicators present more fluctuations for females while TTC is significantly higher for male drivers. The latter contradicts the results of Montgomery et al., (2014) where women were found to brake at higher TTC than men, but it needs to be pointed out that the percentile indicators values are quite extreme ends of their distributions. During near-crash development, the differences across gender are not significant. Moreover, according to 5th percentile indicators the adult and older drivers were found to present more extreme deceleration values than the younger drivers. A similar result was found again by Hong et al., (2016) where the older drivers were speeding to the intersection and then decelerating rapidly.

✓ *Objective 5 "To formulate safety indicators during safety critical driving scenarios"*

(§5.5)

The indicators generated in the framework of objective 3, raised the interest for the last 30s prior to events to capture driver reaction during safety critical scenarios. Driver braking and steering behaviour has been examined using the safety indicators longitudinal deceleration and yaw rate. As mentioned in the literature, several algorithms have been developed from researchers to automatically identify events of interest (Benmimoun et al., 2011; Dozza and González, 2013; Montgomery et al., 2014; 2014b; Satzoda and Trivedi, 2014). In the current research, three algorithms were developed and applied in a reduced dataset covering the 30s prior to impact to investigate driver braking and steering behaviour before the crash and near-crash events. The first two extract deceleration and steering events accordingly, that the drivers have before the final critical event. The third one determines the sequence of the driver manoeuvres.

In the framework of this research, a braking event occurs when the deceleration is more than three standard deviations for at least 10 timepoints (1s) and a steering event when yaw rate exceeds the three standard deviation threshold for 7 timepoints (0.7s). A personalised driver threshold of three standard deviations was used as this way values that represent extreme ends of deceleration or yaw rate distributions can be captured (0.3% are below these values). The decision for this threshold differs from other previous studies that utilised one single threshold to investigate rapid deceleration events or capture events of interest e.g. 0.75g for 5s (Chevalier et al., 2016), 0.4g (Foss et al., 2014) and 0.27g for 0.5s (Wang et al., 2005; Scanlon et al., 2015). The duration was set empirically and based on the perception that a steering manoeuvre is quicker than the braking one.

It has been reported in the literature that drivers start evasive braking manoeuvres on average 2.3 seconds before the impact point while in considerable percentage (40%) of the events, drivers initiate braking only the last second prior the proximity impact point (Haus and Sherony, 2018). Moreover, the median time of braking that has been reported is 0.5 to 1.5s prior to impact, the corresponding steering time is 0.5 to 0.9s prior to impact while there is a percentage (19.9%) of drivers that do not perform any evasive manoeuvre (Scanlon et al., 2015). In accordance to that, this research revealed that in 19.3% of the trips, drivers did not implement any evasive braking manoeuvre before the event; this could mean that they were not aware of the imminent hazardous situation or they did not have time to react once noticed. In 44.3% of the trips, drivers had just one braking event, while in 36.4% of them, drivers performed more than 1 braking manoeuvre before the critical event (up to 6 for some drivers). Regarding the steering behaviour, the percentage of the trips where the drivers did not react is higher up to almost 57%, but this can be explained from the fact that in some events the drivers only braked and did not steer. For the rest of the trips, in most cases drivers performed one steering manoeuvre, there were enough (23%) that they performed 2 to 4 and 4% that implement more than 6 evasive steering manoeuvres before the final event.

The number of evasive manoeuvres could be an indicator of the driver awareness of the imminent situation/event and this has been stated in the literature before. More specifically, Pande et al., (2017) developed a model that suggested that "the frequency and rates of sudden deceleration events on a freeway segment can also be used as a surrogate safety measure in addition to other measures".

Regarding the duration of the final deceleration events, the median value is 2s while for the corresponding yaw rate events the median value is 0.9s. The duration of deceleration events differs between crash and near-crash events, and across different incident types (95% confidence level, median test, $p<0.0001$) but concerning the gender and the age, no statistical differences were detected. In terms of the duration of the yaw rate events, this was found to differ significantly (median test, $p<0.0001$) in near-crash and crash incidents, but not throughout the age groups, gender and incident type.

For the deceleration and yaw rate values that detected during the events, the median average deceleration value was -0.38g and the corresponding yaw rate value was 8.9 °/s. The deceleration value is lower than in previous studies where values of -0.52g (Kusano and Gabler, 2011; 2013b) and 0.58g (Scanlon et al., 2015) have been reported. Furthermore, in 11.4% of the trips, there was a braking and a steering manoeuvre simultaneously. However, in all the cases, the drivers first started with evasive braking, then started steering simultaneously with braking and finally, finished their evasive action by braking. In a previous study by Mazzae et al., (1999) where they also explored the timing of rapid manoeuvring, it was found that 46% of the drivers braked before steered while 52% steered before braked.

✓ *Objective 6 "To model the evolution of TTC values during crash sequence development".*

🕮 *(§5.6)*

TTC 5th percentile indicators did not reveal any explicit pattern and TTC values during crash sequence were further investigated by employing multilevel regression modelling. This way, factors that affect TTC values during the whole crash development, from a normal driving situation to a safety critical event, were examined to gain insights for the evolution of the TTC during the crash or near-crash sequence. Due to the nested nature of the data (multiple observations within trips), there was a trip effect in TTC values that was addressed by utilising multilevel modelling. According to the model, longitudinal acceleration, speed, timestamp and timestamp squared affect the TTC values in a statistically significant way while their effects (slope coefficients) vary across the trip-files. Vehicle type proved also to be significant factor.

Surprisingly, the factors associated with the human characteristics, age and gender, were not found statistically significant, in contradiction to previous research (Papazikou et al., 2017) that found differences related to TTC and normal driving behaviours across gender and age groups. However, it should be noted that the aforementioned study investigated only normal driving conditions, while in the current study the whole event sequence is examined. In the study by Farah et al., (2009), driver characteristics had also a significant effect on TTC. Therefore, this needs further investigation in the future to check if the situation remains the same regarding the other indicators or using a bigger sample.

Speed and longitudinal acceleration have a mixed effect in TTC values with 62.2% and 26.4% of the coefficients correspondingly affecting them in a negative way. This percentage for speed (affecting negatively TTC) is remarkably high supporting studies that correlate high speed with higher crash occurrence probability (Elvik, 2005; Aarts and Schagen, 2006). The usefulness of the speed as collision probability predictor is questionable though as it depends also on the road geometry and this is a fact that should be considered here, as well. Nevertheless, it should be noted that in general the impact of speed on TTC values is very significant according to the model and this is consistent with the findings of previous studies (Farah et al., 2009; Kusano et al., 2015; Chen et al., 2016) that explored the relationship of TTC with speed.

Meanwhile, the longitudinal acceleration proved to be mostly (74.6%) positively associated with TTC values. This can be justified as reasonably higher accelerations occur when the lead vehicle is in a long distance or absent, so the crash risk could be lower. The driving style though can affect the acceleration behaviour and correspondingly the TTC values (i.e. aggressive drivers, tailgating).

Most importantly, time within crash or near crash sequence, expressed as timestamp and timestamp squared, proved to be affecting TTC values significantly. This required further investigation to find out the point where a change in TTC values is detectable. It was found that at timestamp 526, 1.62 minutes before the event, the TTC values started to drop. This seems to be very early in the event sequence, but there are several factors that could potentially explain it. Firstly, TTC is not a pure variable (it is calculated by the division of the distance between two following cars in collision course and their relative speed) and there have been restrictions to the calculation and manipulation of the specific indicator. TTC presents extremely high values, meaningless from the safety aspect, so they have been replaced by the 95th percentile TTC values threshold. It should be noted that still the dataset contains some high values that could be problematic towards the modelling process. Moreover, in terms of individual traces when events are being investigated separately, low TTC events seem to happen very regularly, complicating further the interpretation of the real-world event. Certainly, in-depth investigation of individual behaviours and exploration of TTC in combination with other variables as longitudinal, lateral acceleration and yaw rate can enlighten the pre-crash conditions employing, in a way, humans as sensors for risk perception. Lastly, but possibly most importantly, the traffic conditions of the relative trips that used for the model development could have played a significant role in the low (early) critical timestamp, but in the framework of this analysis the Road Infrastructure Data was not available to be matched with NDS, so the traffic conditions and the road geometry were not incorporated.

Although several methods (e.g. extreme value theory, cluster analysis, multivariate Poisson log-normal model) have been employed for NDS data investigation over the years (Jonasson & Rootzén, 2014; Zheng et al., 2014; Wu et al., 2014), HLM has not yet been widely utilised. In this study, given the specific structure of the data, proved to be extremely useful and reliable statistical technique as it gives to the researcher the chance to control for the thousands of observations derived from the same driver or the same trip.

Attention should be also drawn to the representativeness of the sample, as SHRP2 NDS data is biased towards young drivers. Therefore, generalisation of the results should be considered restricted and treated with great caution.

Objective 7 "To recommend potential thresholds for safety indicators for designing safe and trusted ADAS"

🕮 *(§5.3, §5.5)*

In this research, normal driving conditions from SHRP2 NDS data were investigated and thresholds for normal driving were proposed (Table 5.5). Functional equations were also employed to provide more dynamic normal driving thresholds in relation to speed utilising as input the 1st or 99th percentile values of longitudinal acceleration, lateral acceleration, yaw rate and TTC, during twelve different speed bins (Table 5.7). Furthermore, investigating pre-event braking and steering driver behaviour, events duration and mean values were extracted from the dataset covering 30s prior to the events. Exploiting the median values of duration and the mean values of the safety indicators' during the events, across event type and severity level, thresholds for detecting emerging situations have been recommended (Table 5.18, Table 5.25). Figure 6.1 presents a summary of the thresholds set in the frame of this research for normal driving and emerging situations. It should be noted that that some thresholds are lower for emerging situations than for normal driving, but the concept of duration has been introduced. The measurements regarding the normal driving thresholds are instantaneous, not capturing progression while, for instance, the deceleration should be less than -0.28g for more than 2 seconds to trigger an emerging situation alarm. Moreover, as the aim is to detect early deviations from normal driving, the thresholds chosen from the pre-event driver behaviour analysis are conservative (the values represent the 50th percentile for both event duration and event mean)”.

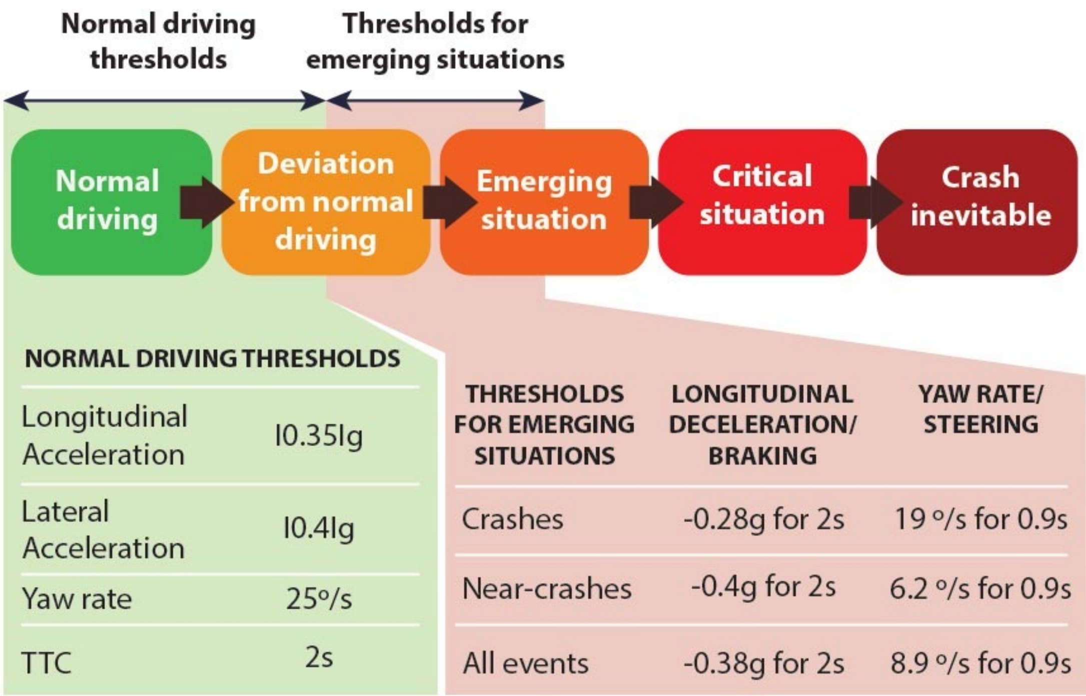


Figure 6.1 Summary of thresholds

Overall, as addressed in this thesis, normal driving is this combination of factors-both human and vehicle kinematics-that produce a typical mode or pattern of driving, relying on the principle of road safety (safety of the driver and people on the road) and ensuring compliance to the driving regulations. Normal driving is uneventful as it allows adequate time for a safe reaction on any event or any activity of other road users. Deviation from normal driving is any departing from this established standard, any change in this driving pattern that can compromise safety, while the emerging critical conditions stand in a next stage where after a deviation, the normal driving was not regained. However, the current research revealed that the stages of crash development cannot be strictly defined, they differ across incident types, driver behaviour and driving styles. The transition from normal driving to a safety critical event usually lasts less than 10 seconds, and in occasions there is no way to detect deviations as drivers do not react to an imminent danger and stages are omitted.

6.2 Implications of research

In this section, the implications of research are discussed structured and presented by research objective. Within Objective 1, a framework produced that helped in the comprehension of the literature and structured the theory for the model development while Objective 2 aid to the choice of NDS as the approach to fulfil the aim of this research. Thus, the implications are mostly with regards this research per se. In the following subsections, the implications of research that concerns the Objectives 3, 4, 5, 6 and 7 are discussed.

✓ *Objective 3 "To quantify safety indicators during normal driving*

According to the results, the design of collision avoidance systems should take into account the vehicle speed and set the warning threshold in relation to this in order to be more effective and meet the drivers' needs. Examining extreme values of accelerations, yaw rate and TTC distributions quantified rarely experienced braking or steering events during normal driving that can be exploited to inform ADAS with new improved thresholds for events detection. The functional equations provide new dynamic thresholds that can better determine intervention time and facilitate driver acceptance by adjusting thresholds and warning timing to be more harmonised with natural driving patterns and therefore have a positive effect regarding the number of false-positive warnings.

This study used SHRP2 data in order to quantify normal driving and the methods and results of this research could be utilised to promote the understanding of natural driving characteristics and can be implemented in other datasets to investigate normal driving distributions.

✓ *Objective 4 "To develop safety indicators for detecting deviations from normal driving"*

Investigating the whole crash sequence, from a normal driving situation until the event, gave insights in the understanding of the relationship between vehicle kinematics indicators and crash risk and this knowledge gained can be exploited to inform existing warning systems or to develop new ADAS, able to efficiently support the drivers and prevent more crashes. The existing systems are set to function in the last stages of the crash sequence and mainly when the driver cannot react anymore (automated braking, ESC, etc.), but there may also be early indications of deviations that could possibly predict the crash. Systems that could detect these inconsistences in driving that are associated with crash risk will be beneficial to road safety. The type and the magnitude of the intervention depends on the stage that the deviation is detected. The earlier the deviation can be detected, the less severe and more successful the intervention. This research revealed how far we can go back from the event to investigate and hopefully prevent it using acceleration, yaw rate and TTC as indicators and set the stage for further investigation of the indicators that can be proved reliable in crash risk prediction.

- ✓ *Objective 5 "To examine safety indicators during safety critical driving scenarios"*

The opportunity to determine risky driver behaviour before actual crash or near-crash involvement could facilitate the development of new safety indicators and measures and could aid in the assessment of the effectiveness of the existing ones. The percentage of the drivers in the present study that did not react suggests that they were not aware of the imminent safety critical situation or they had less than 1s to brake and less than 0.7s to steer. Future development of ADAS can focus on providing better risk awareness for this group of drivers. Whether or not the drivers performed any evasive manoeuvres could serve as a metric for determining driver awareness. Investigating pre-event driver behaviour is important not only because researchers can explore the drivers' awareness of imminent events, but also, they can identify characteristics of pre-event manoeuvres facilitating driver behaviour models and better comprehend notably risky driving patterns in order to be able to detect them and prevent hazardous situations. Moreover, results from the driver behaviour investigation could be useful as input to traffic simulation modelling.

- ✓ *Objective 6 "To investigate the evolution of TTC values during crash sequence development".*

The result of the modelling process revealed once again and confirmed that a single TTC threshold for detecting deviations should not be employed as TTC depends on a range of factors that need to be taken into account. TTC is the most widely adopted metric for forward collision warning (FCW) systems and calibration of traffic simulation models due to the correlation with car following driver behaviour and the relative ease of computation. However, industry should be more sceptical regarding the accuracy and efficiency of TTC and consider utilising it along with other indicators, for the more trusted design and evaluation of new systems.

✓ *Objective 7 "To recommend potential thresholds for safety indicators for designing safe and trusted ADAS"*

The implications regarding this last objective have been separately discussed before in previous sections. In summation, the thresholds and the relevant analysis proposed in the framework of this thesis can be exploited in various ways:

i. Evaluation of collision avoidance or warning systems
ii. Designing of ADAS systems that will capture deviations from normal driving considering more than a single threshold
iii. Use of safety indicators distributions to inform activation thresholds and provide customisable use settings for different driver groups, e.g. for older and young drivers.
iv. Real-time or post-ride performance feedback to drivers
v. Monitoring of driver behaviour (e.g. identifying aggressive drivers) from insurance companies and evaluation of risks with data recorder devices
vi. Development and validation of traffic simulation models.

6.3 Contribution to knowledge

This work has produced new methodological, qualitative and quantitative outcomes which could facilitate future driver behaviour and safety analyses. The main contributions to knowledge of this research are:

SHR2 NDS data investigation and analysis: SHRP2 NDS is a relatively new dataset that offers the opportunity to investigate driver behaviour through hundreds of variables. It is the largest and most comprehensive existing dataset and researchers will be exploring it for years to exploit it in all the possible magnitude. In this thesis, the investigation and analysis of this new dataset checked its potentials regarding the insights it can give to the driver behaviour and safety research in general. It also identified inconsistences, erroneous measurements, problems and limitations of the dataset, as well as, the practicalities and issues that can occur and the way they can be handled.

A naturalistic driving data pre-processing framework: A framework for NDS data cleaning and transformation has been produced which can be adopted by other researchers who are willing to use the data. Furthermore, learning from others' experience with dealing with the data can save time to researchers and facilitates their research.

A method to quantify normal driving: many studies have attempted to set thresholds for safety critical events detection and use them to extract events of interest from big datasets, but only a few have investigated normal driving per se. Exploring and quantifying normal driving could be the key to identify driving patterns and deepen into driver behaviour. In this thesis dynamic thresholds for four indicators, TTC, longitudinal and lateral acceleration and yaw rate, have been employed to quantify normal, uneventful driving in relation to speed. These refer to the first stage of crash development that is the baseline driving and can be exploited in timely detecting deviations.

A method to derive indicators for investigation of crash sequence and detecting abnormalities in measurements of driving behaviour: A data driven approach to create indicators was followed in the framework of this thesis in order to determine compromises in driver safety during crash sequence. This gave the opportunity to check the feasibility of specific indicators, TTC, lateral and longitudinal acceleration and yaw rate, in characterising driving and provided the knowledge for more efficient future use of them in various applications, e.g. warning threshold settings.

A method to detect the point when TTC values progressively reduce towards a safety critical event: The empirical process of detecting the point when TTC values start to decrease can be applied to other datasets or analyses to identify the change in pattern. This method revealed how far we should go backwards in the crash sequence to determine compromises in driver safety.

Manoeuvre extraction algorithms that can facilitate driver behaviour and safety analysis: new algorithms for braking and steering event data extraction, as well as, an algorithm for exploring the sequence of the manoeuvres occurred during the crash sequence have been developed and applied in SHRP2 NDS data. These algorithms aid to pre-event driver behaviour examination and also contribute in setting thresholds for detection of deviations from normal driving.

Thresholds recommendation for ADAS: investigating the pre-event braking and steering driver behaviour provided thresholds for longitudinal acceleration and yaw rate that could be exploited to design new or inform existing ADAS. Thresholds for normal driving were also proposed in the framework of this research.

6.4 Limitations

The research presented in this thesis includes limitations, the most important of which are outlined below:

Sample:

- Initially all the crash and near-crash events were required from VTTI. Apart from them data from drivers that did not have any involvement in safety related events were asked. However, according to VTTI there were no drivers that have not been involved in any crash or near-crash event, therefore in terms of the normal driving analysis, it was not possible to control for "crash prone" drivers. Drivers that have had an event could present different driving behaviour than others that did not experience one; they can either be more cautious or it can be the case that because they drive carelessly or aggressively (tailgating, abrupt braking, etc.) are involved in events.
- Thousands of events were firstly received from VTTI. Although the dataset was very comprehensive, data manipulation and cleaning were conducted to finalise it for analysis. Moreover, hundreds of variables were available, but TTC was not provided. The calculation of TTC in the most appropriate way indicated a specific data manipulation and transformation that reduced the final dataset considerably. Therefore, the final dataset includes 774 events (127 crashes, 647 near-crashes) and involves 553 drivers.

No video data:

Naturalistic driving studies can provide insights for the driver behaviour with the video recordings of the driver's reactions or of the surroundings. The investigation of video data though adds greatly to the time and effort especially considering the magnitude of the data and in the framework of this research, video data was decided to not be investigated. Driver behaviour is only explored by vehicle kinematics.

Road Infrastructure Data not included:

The Road Infrastructure Data was not available to be matched with the NDS, so the traffic conditions and the road geometry could not be taken into account in the modelling process and in other analyses. The exclusion of these variables could have potentially lead to erroneous estimations for the included variables (i.e.omitted-variable bias).

Event time inconsistency:

The actual near-crashes and crashes were acquired from VTTI along with 2 minutes data prior the events. Every event was released as a 30 second time series dataset and the exact moment of the incident was not apparently consistent between the events. This could have affected the result of the modelling and played a role in the visualisation of the data (i.e. indicators plots).

6.5 Extensions and suggestions for future research

The research conducted in the framework of this thesis, examined the transition from a normal driving condition to a safety critical scenario through measurements of driving behaviour in Naturalistic driving studies. Considering the limitations of this study as mentioned in 6.3, there are improvements and extensions that can be made for future research.

Dynamic normal driving thresholds were set for deceleration, lateral acceleration, yaw rate and TTC employing bivariate linear regression models with speed as explanatory variable. An interesting extension would be to employ multivariate analysis to incorporate gender and age as the results of the research showed that the indicators distributions differ across driver demographics. Hence, more customisable thresholds can be developed to quantify normal driving and detect more efficiently deviations.

TTC was investigated by employing hierarchical linear modelling to determine factors affecting its evolution during crash sequence. Traffic and road information were not taken into account for the model development though. Matching the Road infrastructure Data with the NDS data will render possible to incorporate in the model factors that can be linked to TTC patterns including traffic conditions, speed limits, road configuration and other. An obvious question that would be raised from this addition to the model relates to the moment that TTC pattern changes and starts to reduce. Potentially, with the enhancement of the model, the pre-event conditions will be more accurately represented, and this could provide more valid estimations on the onset of TTC declining course. This improvement will inform researchers about the timing in the crash sequence or how far backwards from an event they should step to investigate the factors contributed to the change in pattern of normal driving conditions.

In the frame of this thesis, exploratory analysis and statistical modelling were used to investigate the transition from normal driving to safety critical events due to greater explanation power. As the data is annotated, a supervised approach can be also employed to identify factors affecting this transition through a multivariate learning procedure. For example, fuzzy logic technique appears to be suitable for deviation detection based on simultaneous measurements of indicators.

Regarding the investigation of driver pre-event manoeuvres, there are also interesting extensions for potential follow-up research that can be planned on the basis of driver behaviour analysis. The frequency, the magnitude and the timing of the manoeuvres should be combined with driver demographics and characteristics or personality traits (SHRP2 NDS data provides rich information from several kinds of questionnaires) and facilitate the development of pre-event driving profiles. These can classify driver's pre-event behaviour into categories or inform in-vehicle systems for post-ride feedback that would be valuable for insurance companies to capture for instance, aggressive driving. Driver behaviour research can also be benefited by the exploration of other factors that influence the number and the timing of the driver manoeuvres, e.g. road and traffic related elements.

Another extension concerning the algorithms developed for the event duration and manoeuvre sequence detection would be a real-world driving experiment where the drivers will be asked to perform braking and steering manoeuvres accordingly. The data produced could be used to validate the algorithms and verify that they successfully detect the pre-event driver manoeuvres. Validation and verification of the algorithms can be also achieved by applying them in a dataset with video data available.

Exploiting the comprehensive SHRP2 NDS dataset in the future, normal driving profiles should also be developed on a basis of cluster analysis employed for the drivers' vehicle kinematics, demographics, personality traits and summary statistics. This would devise normal driving styles patterns and thresholds and deliver insights in normal driving behaviour which can benefit future advanced systems to provide personalised and adjustable use settings, e.g. for older or younger drivers, but also insurance companies to formulate profiles of risky drivers and monitor aggressive driving.

Finally, future research should consider SHRP2 NDS data in relation to other European Naturalistic driving studies datasets that are available, for instance, U-

DRIVE or EuroFOT. Comparisons considering driving behaviour across the different datasets, with various research tools and methods depending on the research focus, will potentially provide answers to whether the factors that contribute to aggressive or notably risky driving are and up to which point similar, whether deviations from normal driving can be detecting using uniform thresholds and other. Examining driving styles and behaviour in different datasets will promote the understanding of driving pattern differentiations between regions (Europe, USA) and enhance road safety by aiding to the design of, more integrated and adapted to specific driver groups, ADAS.

6.6 Conclusions

This thesis analyses the development of the crash sequence and examines the transition from normal driving conditions to safety critical scenarios. The methodology developed is applied across three time segments within the crash sequence - the first stage of normal driving, the whole crash sequence development and the last stage during the development of the safety critical scenarios. The analysis has examined the first stages of the integrated safety chain with the characterisation of safety indicators during normal driving, provided insights for the feasibility of detecting deviations with specific indicators and proposed kinematic thresholds for identifying emerging situations. The outputs of this research could be very useful to insurance companies as a basis to create new or update their drivers' profiles in order to efficiently recognise risky driving. Furthermore, considering the progressive introduction of vehicle automation, the outputs of this thesis lead to an enhanced design of ADAS that can provide tailored assistance for different groups of drivers, e.g. older drivers, but most importantly, can intervene in a timely manner before a deviation from normal driving culminates to a crash scenario.

7 References

ALMAHASNEH, H., CHOOI, W.T., KAMEL, N. and MALIK, A.S., 2014. Deep in thought while driving: An EEG study on drivers' cognitive distraction. *Transportation research part F: traffic psychology and behaviour,* **26**(1)*,* 218-226.

AERON-THOMAS, A. and JACOBS, G. 2011. 14 Road crashes and low-income cities: impacts and options. In, DIMITRIOU, H.T, and GAKENHEIMER, R, ed., *Urban Transport in the Developing World: A Handbook of Policy and Practice.* 1st Ed. Londom, UK: Edward Elgar Publishing Ltd, 2011, pp. 386.

ABUALI, N and ABOU-ZEID, H, 2016. Behavior Modeling: Developments and Future Directions,. *International Journal of Vehicular Technology*.

ALBERTI, E., and BELLI, G., 1978. Contributions to the Boltzmann-like approach for traffic flow: A model for concentration dependent driving programs. *Transportation Research,* **12**(33-42).

ALLEN, R.W., et al., 2005. Driver assessment with measures of continuous control behavior. In: Proceedings of the third international driving symposium on human factors in driver assessment, training and vehicle design, 27–30 June, Rockport, ME, 165–172.

ALLEN, B.L., SHIN, B.T., and COOPER, P.J., 1977. *Analysis of Traffic Conflicts and Collisions*. Dept. Of Civil Engineering, McMaster University: Hamilton.

ARA VAN DER HORST, M DE GOEDE, S DE HAIR-BUIJSSEN, and R METHORST, 2014. Traffic conflicts on bicycle paths: A systematic observation of behaviour from video. *Accident Analysis & Prevention* **62**, 358-368.

ARCHER, J., 2001. Traffic Conflict Technique. Historical to current State-of-the-Art. Stockholm: Institutionen för Infrastruktur KTH, Stockholm, pp.2-3.

ARCHER, J., 2004. *Methods for the assessment and prediction of traffic safety at urban intersections and their application in micro-simulation modelling.* PhD Thesis, Royal institute of technology: Stockholm, Sweden.

ARCHER, J., 2005. *Indicators for Traffic Safety Assessment and Prediction and Their Application in Micro-Simulation Modelling: A Study of Urban and Suburban Intersections*. PhD Thesis, Royal Institute of Technology, Stockholm.

BAGDADI, O., 2012. *The development of methods for detection and assessment of safety critical events in car driving*. Published thesis (PhD), Lund University, Lund, Sweden.

BAGDADI, O., and VARHELYI, A, 2013. Development of a method for detecting jerks in safety critical events. *Accident Analysis and Prevention*, **50**, 83-91.

BAKER, J. S., FRICKE L., 1986. The Traffic Accident Investigation Manual, Northwestern University Traffic Institute. 9th ed. United States of America.

BAKER, J. S., FRICKE L., 1986. The Traffic Accident Investigation Manual. 9th edition. Northwestern University: Northwestern University Traffic Institute.

BÄRGMAN, J., BODA, C., and DOZZA, M., 2017. Counterfactual simulations applied to SHRP2 crashes: The effect of driver behavior models on safety benefit estimations of intelligent safety systems. *Accident Analysis & Prevention* **102**, 165–180. 10.1016/j.aap.2017.03.003.

BEKEY, G.A., BURNHAM, G.O., and SEO, J., 1977. Control theoretic models of human drivers in car following. *Human Factors*, **19** (4), 399–413.

BENMIMOUN, M., FAHRENKROG, F., ZLOCKI, A., ECKSTEIN, L., 2011. 'Incident detection based on vehicle CAN-data within the large scale field operational test "euroFOT"'. In: *Proceedings of the 22nd International Technical Conference on the Enhanced Safety of Vehicles (ESV 2011),* Washington, D.C: Mira. BISHOP, R. 2005.*Intelligent Vehicle Technology and Trends*. Norwood, MA: USA Artech House.

BENMIMOUN, M., FAHRENKROG, F., ZLOCKI, A. and ECKSTEIN, L., 2011. Incident detection based on vehicle CAN-data within the large scale field operational test "euroFOT". In: *22nd Enhanced Safety of Vehicles Conference (ESV 2011), Washington, DC/USA*.

BLANCO, M., HICKMAN, J. S., OLSON, R. L., BOCANEGRA, J. L., HANOWSKI, R. J., NAKATA, A., GREENING, M., MADISON, P., HOLBROOK, G. T., BOWMAN, D.,

2008. *Investigating critical incidents, driver restart period, sleep quantity, and crash countermeasures in commercial vehicle operations using naturalistic data collection: Final report.* Washington, D.C: Federal Motor Carrier Safety Administration, 2008. Report No. FMCSA-RRR-13-017.

BOER, E.R., 1999. Car following from the driver's perspective. *Transportation Research Part F*, **2** (4), 201–206.

BOER, E.R., WARD, N.J., MANSER, M.P., YAMAMURA, T. AND KUGE, N., 2005. Driver performance assessment with a car following model. In: *Proceedings of the third international driving symposium on human factors in driver assessment, training and vehicle design , June 27–30, Rockport, Maine, USA: , 433–440.*

BRACKSTONE, M. and MCDONALD, M., 1999. Car-following: a historical review. *Transportation Research Part F,* **2** (4), 181–196.

BUCCHI, A., SANGIORGI, C. and VIGNALI, V., 2012. Traffic psychology and driver behavior. *Procedia-social and behavioral sciences*, **53**, 972-979.

CACCIABUE, P.C., 2007. *Modelling driver behaviour in automotive environments: critical issues in driver interactions with intelligent transport systems*. 1St ed. New York, USA: Springer-Verlag.

CAMPBELL, K., 2012.The SHRP 2 Naturalistic Driving Study. Addressing Driver Performance and Behavior in Traffic Safety. *TR News,* 09/10, pp 30-35.

CARSTEN, O., 2007. From driver models to modelling the driver: what do we really need to know about the driver? In: P.C. Cacciabue, ed. *Modelling driver behaviour in automotive environments*. London: Springer-Verlag, pp.105–120.

CARSTEN, O., FOWKES, M., LAI, F., CHORLTON, K., JAMSON, S., TATE, F., SIMPKIN, B., 2008. *ISA – UK Intelligent Speed Adaptation*. Leeds: University of Leeds. Report no. 01385800CHAPMAN, A.L., 1954. An Epidemiological Approach to Traffic Safety. *Public Health Reports*, **69**(8), 773–775.

CHEN S.H, JAKEMAN A.J., and NORTON JP, 2008. Artificial intelligence techniques: an introduction to their use for modelling environmental systems. *Mathematics and Computers in Simulation* **78**(2), 379–400.

CHEN, R., SHERONY, R., AND GABLER, H., 2016 *Comparison of Time to Collision and Enhanced Time to Collision at Brake Application during Normal Driving SAE Technical Paper*. USA: Virginia Tech.

CHEVALIER, A., CHEVALIER, A.J., CLARKE, E., COXON, K., BROWN, J., ROGERS, K., BOUFOUS, S., IVERS, R. and KEAY, L., 2016. Naturalistic rapid deceleration data: Drivers aged 75 years and older. *Data in brief*, **9**, 909-916.

CHIN, H.C., and QUEK, S.T., 1997. Measurement of traffic conflicts. *Safety Science,* **26**(3), 169–185.

CIRILLO J.A., 1968. Interstate system crash research; study II, interim report II, Public Roads, **35** (3), pp. 71-76.

CONGER, J.J., GASKILL, H.S., GLADD, D.D., HASSELL, L., RAINEY, R.V. and SAWREY, D.W.L., 1959. Psychological and psychophysical factors in motor vehicle accidents. *Journal of the American Medical Association* **169**, 1581-1587.

CUERDEN R. and MCCARTHY M., 2016, The methodology and initial findings for the Road Accident In Depth Studies (RAIDS) Programme, RAIDS Phase 1 Report, Department for Transport, International Vehicle Standards Division.

CUERDEN, R. et al., 2008. The UK on the spot accident data collection study phase II report. London: Department for Transport.

CUNTO, F., 2008. *Assessing Safety Performance of Transportation Systems using Microscopic Simulation*. PhD Thesis, University of Waterloo, Ontario, Canada.

DATTA, T.K., 1979. Accident surrogates for use in analysing highway safety hazards. In: *Proceedings of the Second International Traffic Conflict Technique Workshop, January 1979.* Paris, Transport Research Laboratory, pp. 4-20.

DAVIS, G.A., and SWENSON, T., 2006. Collective responsibility for freeway rear-ending accidents: an application of probabilistic causal models. *Accident Analysis and Prevention,* **38** (4), 728–736.

DAVIS, G.A., HOURDOS, J., and XIONG, H., 2008. Outline of a causal theory of traffic conflicts and collisions. In: *Proceedings of the 87th Transportation Research Board Annual Meeting, Washington DC* (CD-ROM). Washington DC: National Research Council, TRB, pp 10-15.

DE WINTER, J.C.F. and HAPPEE, R., 2012. Modelling driver behaviour: a rationale for multivariate statistics. *Theoretical Issues in Ergonomics Science*, **13**(5), 528-545.

DI FLUMERI G, BORGHINI G, ARICÒ P, SCIARAFFA N, LANZI P, POZZI S, VIGNALI V, LANTIERI C, BICHICCHI A, SIMONE A and BABILONI F , 2018. EEG-Based Mental Workload Neurometric to Evaluate the Impact of Different Traffic and Road Conditions in Real Driving Settings. *Frontiers in human neuroscience*, **12**, 509.

DINGUS, T. A., KLAUER, S. G., NEALE, V. L., PETERSEN, A., LEE, S. E., SUDWEEKS, J., PEREZ, M. A., HANKEY, J., RAMSEY, D., GUPTA, S., BUCHER, C., DOERZAPH, Z. R., JERMELAND, J., and KNIPLING, R.R., 2006. *The 100-Car Naturalistic Driving Study: Phase II – Results of the 100-Car Field Experiment*. Washington, D.C: National Highway Traffic Safety Administration, 2006. Report No. DOT HS 810 593.

DOZZA, M., and GONZÁLEZ, N. P., 2013. Recognising safety critical events: can automatic video processing improve naturalistic data analyses? *Accident Analysis and Prevention*, **60**, 298–304.

DOZZA, M., BÄRGMAN, J., and LEE, J. D, 2013. Chunking: a procedure to improve naturalistic data analysis. *Accident Analysis and Prevention*, **58**, 309–17.EMBERGER, M, 1993. The contribution of PROMETHEUS to European traffic safety. In: M. PARKES, A and FRANZEN, S. *Driving future Vehicles*. London: Taylor & Francis, P. 30.

EBOLI, L., MAZZULLA, G. and PUNGILLO, G., 2017. How to define the accident risk level of car drivers by combining objective and subjective measures of driving style. *Transportation research part F: traffic psychology and behaviour*, **49**, 29-38.

EBOLI, L., MAZZULLA, G. and PUNGILLO, G., 2017. The influence of physical and emotional factors on driving style of car drivers: A survey design. *Travel Behaviour and Society*, **7**, 43-51.

EDIE, L.C., and FOOTE, R.S., 1960. Effect of shock waves on tunnel traffic flow. In: *Proceedings of the Thirty-Ninth Annual Meeting of the Highway Research Board, Washington, D.C., January 11-15, 1960*. Washington, USA: Highway Research Board, pp. 492 – 505.

ELVIK, R., CHRISTENSEN, P. and AMUNDSEN, A., 2004. *Speed and road accidents. An evaluation of the Power Model*. TØI report 740/2004. Institute of Transport Economics TOI, Oslo.

ESKANDARIAN A., 2012. *Handbook of Intelligent Vehicles*. 2nd ed. London: Springer-Verlag.

European Road Safety Observatory, 2015. Fatigue.

EVANS, L, 2004. *Traffic safety*. Bloomfield Hills, MI: Science Serving Society.

EVANS, L., 2006. The Dramatic Failure of U.S. Safety Policy. *TR News 242,* 1st Jan, pp. 28–31. EuroFOT-Consortium. EuroFOT – *Bringing Intelligent Vehicles on the Road*. EU: EuroFOT Consortium.

FABER, F., JONKERS, E., LJUNG AUST, M., BENMIMOUN, M., REGAN, M., JAMSON, S., and DOBBER- STAIN, J., 2012. EuroFOT *D6.2 – Analysis Methods for user related aspects and impact assessment on traffic safety, traffic efficiency and environment (PU)*.

FAIRCLOUGH S., and GRAHAM, R., 1999. Impairment of driving performance caused by sleep deprivation or alcohol: a comparative study. *Human Factors*, **41**(1), 118-28.

FANCHER, P., ERVIN, R., SAYER, J., HAGAN, M., BOGARD, S., BAREKET, Z., MEFFORD, M., and HAUGEN, J. *Intelligent cruise control field operational test, Final Report, Volume I: Technical Report*. Ann Arbor: MI University of Michigan Transportation Research Institute, 1998.

FELL, J.C, 1976. A motor vehicle accident causal system: The human element. *Human Factors*, **18**(2), 85-94.

FILTNESS, A.J., THOMAS, P., TALBOT, R., QUIGLEY, C., PAPADIMITRIOU, E., YANNIS, G., THEOFILATOS, A., AIGNER-BREUSS, E., KAISER, S., MACHATA, K. and WEIJERMARS, W., 2016. *The application of systems approach for road safety policy making, Deliverable 8.1 of the H2020 project SafetyCube.* Loughborough University: SafetyCube.

FLEISHMAN, E.A, 1967. Performance assessment based on an empirically derived task taxonomy. *Human Factors*, **9**(2), 349-366.

FLEISHMAN, E.A, 1975. Toward a taxonomy of human performance. *American Psychologist,* **30**, 1127-1149.
FOSS, R.D. and GOODWIN, A.H., 2014. Distracted driver behaviors and distracting conditions among adolescent drivers: Findings from a naturalistic driving study. Journal of Adolescent Health, **54**(5), 50-60.

FRICKE, L., 1990. Traffic Accident Reconstruction: Traffic Accident Investigation Manual. Vol.2. Evanston, Illinois: The Traffic Institute, Northwestern University.

FRIDMA, LEX., Brown, D., Glazer, M., Angell, W., DOOD, S., JENIK, B., TERWILLIGER, J., KINDELSBERGER, J., DING, L., SEAMAN, S., ABRAHAM, H., MEHLER A., SIPPERLEY A., PETTINATO, A., SEPPELT, B., ANGELL, L, MEHLER, R., and REIMER., 2018. *MIT Autonomous Vehicle Technology Study: Large-Scale Deep Learning Based Analysis of Driver Behavior and Interaction with Automation*. MIT: Arvix..

FULLER, R, 2005. Towards a general theory of driver behaviour. *Accident Analysis and Prevention*, **37**(3), 461–472.

FULLER, R. A, 1984. Conceptualisation of driving behaviour as threat avoidance. *Ergonomics,* **27**(2), 1139-1155.

GERMAN IN-DEPTH ACCIDENT STUDY (GIDAS), "About GIDAS", [viewed 01/07/2019], Available from: https://www.gidas.org/en/about-gidas/ueberblick-ueber-gidas/

GETTMAN, D. and HEAD, I., 2003, *Surrogate Safety Measures from Traffic Simulation Models. Final Report*. Tucson: Transportation Research Record. FHWA-RD-03-050.

GIBSON, J.J. and CROOKS, L.E., 1938. A theoretical field-analysis of automobile-driving. *The American Journal of Psychology*, **51**(3), 453–471.

GORDON, T., GREEN, P., KOSTYNIUK, L., BARNES, M., BLANKESPOOR, A., BOGARD, S., and BLOWER, D., 2009. A multivariate analysis of crash and naturalistic event data in relation to highway factors using the GIS framework. In: *Proceedings of the 4th SHRP2 Safety Research Symposium*. *Washington, DC*: SHRP.GRAYSON, G.B., and HAKKERT, A.S., 1987. Accident analysis and conflict behaviour. In ROTHERNGATTER, J. and DE BRUINE, R.A. (Eds). *Road-user and Traffic Safety.* Assen: Van Gorcum p. 27–59.

GREEN, M., 2000. How long does it take to stop? Methodological analysis of driver perception-brake times. *Transportation Human Factors*, **2** (3), 195-216.

GREEN, P., 1995. *Measures and Methods Used to Assess the Safety and Usability of Driver Information Systems.* McLean, VA: U.S. Department of Transportation, Federal Highway Administration. FHWA-RD-94-088.

GREEN, P., CULLINANE, B., ZYLSTRA, B., and SMITH, D. , 1995. *Typical Values for Driving Performance with Emphasis on the Standard Deviation of Lane Position: A Summary of the Literature*. University of Michigan: UMTRI, 2004.GULLY, S.M., WHITNEY, D.J., and VANOSDALL, F.E. Prediction of police officers traffic accident involvement using behavioral observations. *Accident Analysis and Prevention*, **27** (3), 355–362.

GREENBERG, H., 1959. An analysis of traffic flow. *Operations research,* **7**(1), 79-85.
GUO, F., KLAUER, K.G., HANKEY, J.M., and DINGUS, T.A., 2010. Near crashes as crash surrogate for naturalistic driving studies. *Transportation Research Record*, **2147**(1), 66–74.

HANCOCK, P.A., and DESMOND, P.A. 2001. Stress, workload and fatigue. Mahwah, New Jersey: Lawrence Erlbaum Associates.

GUO, K and GUAN, H., 1993. Modelling of driver/vehicle directional control system. *Vehicle System Dynamics*, **22**(3), 141–184.

HALE, A. R., STOOP, J., and HOMMELS, J, 1990. Human error models as predictors of accident scenarios for designers in road transport systems. *Ergonomics,* **33**, 1377-1388.

HANCOCK, P.A., 1999. IS car following the real question – are equations the answer? *Transportation Research Part F,* **2**(4), 197–199.

HANKEY, J., PEREZ, M., and MCCLAFFERTY, J. , 2008. *Description of the SHRP 2 naturalistic database and the crash, near-crash, and baseline data sets, Task report*. Blacksburg, VA: Virginia Tech Transportation Institute, 2016.

HANOWSKI, R. J., BLANCO, M., NAKATA, A., HICKMAN, J. S., SCHAUDT, W. A., FUMERO, M. C., OLSON, R. L., JERMELAND, J., GREENING, M., HOLBROOK, T., KNIPLING, R. R., and MADISON, P. *The Drowsy Driver Warning System Field Operational Test: Data Collection Methods, Final Report*. Blacksburg, VA: Virginia Tech Transportation Institute. HS 810 035.

HANKEY, J.M., PEREZ, M.A. AND MCCLAFFERTY, J.A., 2016. *Description of the SHRP 2 naturalistic database and the crash, near-crash, and baseline data sets*. Virgina Tech, USA: Virginia Tech Transportation Institute.

HARRIS, C.R. and JENKINS, M., 2006. Gender differences in risk assessment: why do women take fewer risks than men? *Judgement and Decision Making*, 1(1), 48-63.

HATAKKA, M., KESKINEN, E., GREGERSEN, N.P., GLAD, A. and HERNETKOSKI, K., 2002. From control of the vehicle to personal self-control; broadening the perspectives to driver education. *Transportation Research Part F: Traffic Psychology and Behaviour*, **5**(3), 201-215.

HAUER, E., 1982. Traffic conflicts and exposure. *Accident Analysis and Prevention*, **14** (5), 359–364.

HAUER, E., 1999. Safety in Geometric Design Standards. White Paper.

HERMAN, R., MONTROLL, E., POTTS, R., and ROTHERY, R., 1959. Traffic dynamics: Analysis of stability in car following. *Operations Research*, **7**(1), 1-139.

HEVELKE A., and NIDA-RUMELIN J., 2014. Responsibility for crashes of autonomous vehicles: An ethical analysis. *Science and engineering ethics,* **21**(3), 619–630.

HILL, J. et al., 2001. The methodology of on the spot accident investigation in the UK. IN: Proceedings of the 17th International Technical Conference on the Enhanced Safety of Vehicles (ESV 17), Amsterdam. The Netherlands, 4 7 June, 10

HILLENBRAND, J., SPIEKER, A.M., and KROSCHEL, K., 2006. A multilevel collision mitigation approach: situation assessment, decision making, and performance trade-offs. *IEEE Transactions on Intelligent Transportation Systems,* **7**(4), 528–540.

HJÄLMDAHL, M., 2004. *In-vehicle speed adaptation – on the effectiveness of a voluntary system*. PhD thesis, Lund University.

HOLLNAGEL, E., NÅBO, A. and LAU, I.V., 2003. A systemic model for driver-in-control. In: *Proceedings of the second international driving symposium on human factors in driver assessment, training and vehicle design, 21–24 July, Park City, UT*. Iowa City: University of Iowa, pp. 86–91.

HONG, S., MIN, B., DOI, S. and SUZUKI, K., 2016. Approaching and stopping behaviors to the intersections of aged drivers compared with young drivers. *International Journal of Industrial Ergonomics*, **54**(6), 32-41.

HOWARD, D, 2013. *Robots on the road : The moral imperative of the driverless car.* [online]. Science Matters. [Viewed 27/06/2019]. Available from http://donhoward-blog.nd.edu/2013/11/07/robots-on-the-road-the-moral-imperative-of-the-driverless-car/#.VcjZZvlmHgY.

HOYOS, C.G, 1984. Opinion on a discussion on the theory of risk-co-compensation. *Journal of Traffic Safety*, **30**, 61-62.

HUGUENIN, R.D., 1982 On the problem of risk homeostasis theories in traffic psychology. *Journal of Traffic Safety*, **28**, 180-187.

HYDEN, C., 1987. *The Development of Method for Traffic Safety Evaluation: The Swedish Traffic Conflict Technique*. Lund Institute of Technology, Lund.

HYDÉN, C., 1996. *The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique*. Lund University, Sweden: Department of Traffic Planning and Engineering, 1987.HYDÉN, C. Traffic conflicts technique: state-of-the-art. In: TOPP, H.H., ed,, *Traffic Safety Work with Video-processing*. University Kaiserslautern, Germany: Transportation Department: pp. 3-14

JAGACINSKI, R.J. and FLACH, J.M., 2003. *Control theory for humans: quantitative approaches to modelling performance*. Mahwah, NJ: Lawrence Erlbaum Associates.

JONAH, B.A., THIESSEN, R., and AU-YEUNG, E., 2001. Sensation seeking, risky driving and behavioral adaptation. *Accident Analysis and Prevention*, **33** (5), 679–684.

JONASSON, J. K., and ROOTZÉN, H. 2014. Internal validation of near-crashes in naturalistic driving studies: a continuous and multivariate approach. *Accident Analysis and Prevention*, **62**, 102–9.

JOVANIS, P.P., SHANKAR, V., AGUERO-VALVERDE, J., WU, K., and GREENSTEIN, A. *Analysis of Existing Data: Prospective Views on Methodological Paradigms, Draft Final Report*. Larson Transportation Institute, Pennsylvania State University: Transportation Research Board of The National Academies.

KARRER, K., VOHRINGER-KUHNT, T., BAUMGARTEN, T., and BRIEST, S., 2004. The role of individual differences in driver fatigue prediction. In: *3rd International Conference on Traffic & Transport Psychology*, Nottingham, UK.

KIDD, E.A and LAUGHERY, K.R., 1964. *A computer model of driving behavior: The highway intersection situation*. Buffalo, NY: Cornell Aeronautical Laboratories. Report no. VI-1843-V-I.

KLEBELSBERG, D, 1977. The model of subjective and objective security. *Swiss Journal of Psychology and its Applications,* ***36***(4), 285-294.

KLAUER, S.G., DINGUS, T.A., NEALE, V.L., SUDWEEKS, J.D., and RAMSEY, D.J., 2006. *The impact of driver inattention on near-crash/crash risk: an analysis using the 100-car naturalistic driving study data*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration, 2006. HS 810 594.

KLEBELSBERG, D., 1971. Subjective and objective safety in road traffic as a task for traffic safety work. *Series of the German Road Patrol*, **51**, 3-12.

KLUGER, R., SMITH, B. L., PARK, H., and DAILEY, D. J., 2016. Identification of safety-critical events using kinematic vehicle data and the discrete fourier transform. *Accident Analysis and Prevention*, **96**, 162–168.

KOBAYASHI, Y., 2007. The emosign – analyzing the emotion signature in human motion. In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Montréal, Canada, 7-10 October 2007*. Montreal, Canada: IEEE, 2096–2101.

KUSANO, K.D., CHEN, R., MONTGOMERY, J. and GABLER, H.C., 2015. Population distributions of time to collision at brake application during car following from naturalistic driving data. *Journal of safety research*, 54 (29), 95-104

KUSANO, K.D., MONTGOMERY, J. and GABLER, H.C., 2014. Methodology for identifying car following events from naturalistic data. In: *IEEE Intelligent Vehicles Symposium Proceedings*. Dearborn, MI, USA, 2014. USA: IEEE, pp. 281-285).

LAURESHYN, A., ARDÖ, H., JONSSON, T., and SVENSSON, Å, 2009. Application of automated video analysis for behavioural studies: concept and experience. *IET Intelligent Transport Systems* **3** (3), 345–357.

LAURESHYN, A., ÅSTRÖM, K., and BRUNDELL-FREIJ, K., 2009b. From speed profile data to analysis of behaviour: classification by pattern recognition techniques. *IATSS Research,* **33** (2), 88–98.

LAURESHYN, A., SVENSSON, Å., and HYDÉN, C., 2010. Evaluation of traffic safety based on micro-level behavioural data: theoretical framework and first implementation. *Accident Analysis and Prevention,* **42** (6), 1637–1646.

LEBLANC, D., SAYER, J., WINKLER, C., ERVIN, R., BOGARD, S., DEVONSHIRE, J., MEFFORD, M., HAGAN, M., BAREKET, Z., GOODSELL, R., GORDON, T., 2006. Road departure crash warning system field operational test: methodology and results. Report UMTRI- 2006-9-1. University of Michigan Transportation Research Institute, Ann Arbor, MI.

LEE J. D., 2008. Fifty Years of Driving Safety Research. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **50** (3), 521-528.

LEE, J.D., 2006. Driving safety. In NICKERSON, R. S. (Ed.). *Reviews of human factors and ergonomics.* Santa Monica, CA: Human Factors and Ergonomics Society, pp. 173-218.

LEE, S.E., SIMONS-MORTON, B.G., KLAUER, S.E., OUIMET, M.C., AND DINGUS, T.A., 2011. Naturalistic assessment of novice teenage crash experience. *Accident Analysis and Prevention,* **43** (4), 1472–1479.

LENARD, J., BADEA-ROMERO, A. and DANTON, R., 2014. Typical pedestrian accident scenarios for the development of autonomous emergency braking test protocols. Accident analysis and prevention, 73. 73-80

LERNER, N., JENNESS, J., SINGER, J., KLAUER, S., LEE, S., DONATH, M., MANSER, M., and WARD, N., 2010. *An Exploration of Vehicle-Based Monitoring of Novice Teen Drivers: Final Report*. Washington, D.C: National Highway Traffic Safety Administration, 2010.

LIANG, Y., LEE, J. D., and YEKHSHATYAN, L., 2012. How dangerous is looking away from the road? Algorithms predict crash risk from glance patterns. *Human Factors*, **54**(6), 1104–1116.

MACADAM, C.C., 2003. Understanding and modelling the human driver. *Vehicle System Dynamics*, **40** (1), 101–134.

MAMDANI, E.H., AND ASSILIAN, S, 1975. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, **7** (1), 1-13.

MANN, H. B., AND WHITNEY, D.R, 1947. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, **18**(1), 50-60.

MARKKULA, G., 2015. *Driver behavior models for evaluating automotive active safety: from neural dynamics to vehicle dynamics*. PhD Thesis, Chalmers University of Technology, Göteborg, Sweden.

MARTENS, M. and BROUWER, R., 2011. Linking Behavioural Indicators to Safety: What is Safe and What is Not? In: *Proceedings of the 3rd International Conference on Road Safety and Simulation.* Indianapolis, USA: Transportation Research Board, pp.11-14.

MAZZAE, E.N., BARICKMAN, F., BALDWIN, G.S. and FORKENBROCK, G., 1999. Driver crash avoidance behavior with ABS in an intersection incursion scenario on dry versus wet pavement. *SAE transactions*, **108**(6), 2340-2352.

MCCLAFFERTY, J., PEREZ, M., and HANKEY, J., 2008. *The Drowsy Driver Warning System Field Operational Test: Data Collection Methods, Final Report*. Blacksburg, VA: Virginia Tech Transportation Institute. HS 810 035.

MCGEHEE, D. V., RABY, M., CARNEY, C., LEE, J. D., and REYES, M. L., 2007. 'Extending parental mentoring using an event-triggered video intervention in rural teen drivers'. *Journal of Safety Research,* **38** (2), 215–227.

MCGEHEE, D.V., BOYLE, L.N., HALLMARK, S.J., LEE, J.D., NEYENS, D.M., and WARD, N.J., 2010. *Integration of Analysis Methods and Development of Analysis Plan Phase II Report*. Washington, D.C: Transportation Research Board of the National Academies.

MCKENNA, F., 1982. The human factor in driving accidents: An overview of approaches and problems. *Ergonomics* **25**(10), 867-877.

MCKNIGHT, A.J., and ADAMS, B.B., 1970. *Driver education task analysis. Volume II: Task analysis methods*. Alexandria, VA: Human Resources Research Organization, 1970.

MCKNIGHT, A.J., and HUNDT, A.G., 1971. *Driver education task analysis. Volume III: Instructional Objectives.* Alexandria, VA: Human Resources Research Organization.

MCLAUGHLIN, S.B., HANKEY, J.M. and DINGUS, T.A., 2008. A method for evaluating collision avoidance systems using naturalistic driving data. *Accident Analysis & Prevention*, **40**(1), 8-16.

MCLAUGHLIN., S.B. and HANKEY J.M., 2015. *Naturalistic Driving Study: Linking the Study Data to the Roadway Information Database. SHRP 2 Report S2-S31-RW-3.* Washington DC, USA: Transport Research Board.

MCRUER, D.T. and JEX, H.R., 1967. A review of quasi-linear pilot models. *IEEE Transactions on Human Factors in Electronics*, **8** (3), 231–249.

MCRUER, D.T. and JEX, H.R., 1969. Theory of manual vehicular control. *Ergonomics*, **12**, 599-633.

MCRUER, D.T., ALLEN, R.W., WEIR, D.H. and KLEIN, R.H., 1977. New results in driver steering control models. Human factors, **19**(4), 381-397.

MICHON, J. A., 1985. A critical view of driver behavior models. What do we know, what should we do? In: EVANS, L., and SCHWING, R., eds. *Human behavior and traffic safety*. New York: Plenum Press, pp. 381-397.

MIGLETZ, D.J., GLAUZ, W.D. and BAUER, K.M. *Relationships Between Traffic Conflicts and Accidents: Final technical report (Vol. 2*). US Department of Transportation, Federal Highway Administration, 1985.

MINDERHOUD, M., and BOVY, P., 2001. Extended time-to-collision measures for road traffic safety assessment. *Accident Analysis and Prevention,* **33**(1), 89–97.

MOLINERO, A., EVDORIDES, H., NAING, C., KIRK, A., TECL, J., BARRIOS, J.M., SIMON, M.C., PHAN, V., and HERMITTE, T., 2009. *Accident causation and pre-accidental driving situations. Part 2. In-depth accident causation analysis*. Loughborough University, UK: TRACE. Project No. 027763.

MONTGOMERY, J., KUSANO, K.D. and GABLER, H.C., 2014. Age and gender differences in time to collision at braking from the 100-car naturalistic driving study. *Traffic injury prevention*, **15**(1), 15-20.

MORRIS, A., HASSAN, A., MACKAY, M. and HILL, J., 1995. Head injuries in lateral impact collisions. Accident analysis and prevention, **27**(6), 749-756

MUHLRAD, N., 1993. Traffic Conflict Techniques and Other Forms of Behavioural Analysis, In: *Proceedings of the 6th ICTCT Workshop, Salzburg*.

MURONGA, K. and RUXWANA, N., 2017. NATURALISTIC DRIVING STUDIES: THE EFFECTIVENESS OF THE METHODOLOGY IN MONITORING DRIVER BEHAVIOUR. In: *36th Southern African Transport Conference. Pretoria, South Africa*.

NAATANEN, R., and SUMMALA, H. 1976. *Road user behavior and traffic accidents*. 1st ed. Amsterdam: North Holland Publishing Company.

NAJI, H. A. H., LYU, N., WU, C. and ZHANG., H. 2011. Examining contributing factors on driving risk of naturalistic driving using K-means clustering and ordered logit regression. In, YAN, X, and LINGONG-DAXUE, W, *4th International Conference on Transportation Information and Safety (ICTIS), Banff, Canada, June, 2017*. Canada: *IEEE*, pp. 1189-1195.

NAJM, W.G., STEARNS, M.D., HOWARTH, H., KOOPMANN, J., and HITZ, J., 2006. *Evaluation of an Automotive Rear-end Collision Avoidance System.* Volpe National Transportation Systems Centre, USA: U.S. Department of Transportation.

NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION, 2016. *Motor Vehicle Crashes: Overview*. Washington, DC: U.S. National Highway Traffic Safety Administration.

NEALE, V.L., DINGUS, T.A., KLAUER, S.G., SUDWEEKS, and J., GOODMAN, M., 2005. AN overview of the 100-car naturalistic study and findings. In: *Proceedings – 19th International Technical Conference on the Enhanced Safety of Vehicles, Washington, DC*, June 6–9, 2005. Washington, DC: NHTSA.

NGUYEN H.T, and WALKER E.A, 1999. *A first course in fuzzy logic*, 2nd ed. Boca Raton, USA: Chapman & Hall.

NIEZGODA, M., KAMINSKI, T., and KRUSZEWSKI, M., 2012. Measuring driver behaviour indicators for traffic. *Journal of KONES powertrain and transport*, **19** (4), 504–511.

NILSSON, D., LINDMAN, M., VICTOR, T. and DOZZA, M., 2018. Definition of run-off-road crash clusters- For safety benefit estimation and driver assistance development. Accident analysis and prevention. 113, 97-105.

NITSCHE, P., THOMAS, P., STUETZ, R. and WELSH, R., 2017. Pre-Crach Scenarios at road junctions: A clustering method for car crash data. Accident Analysis and Prevention107. 137-151.

NORMAN, D. A., and BOBROW, D. G, 1975. On data-limited and resource-limited processes. *Cognitive Psychology,* ***7***(2), 44-64.

NYGÅRD, M., 1999. *A Method for Analysing Traffic Safety with Help of Speed Profiles*. M.Sc., Tampere University of Technology, Espoo.

O'NEILL, B. and WILLIAMS, A., 2004. Risk homeostasis hypothesis: a rebuttal. *Injury Prevention*, **4** (2), 92–93.

OECD, 2008. *Towards Zero, Ambitious Road Safety Targets and the Safe System Approach*. Transport Research Centre, USA: International Transport Forum.

OLSON, R.L., HANOWSKI, R.J., HICKMAN, J.S., and BOCANEGRA, J., 2009. *Driver distraction in commercial vehicle operations*. Washington, DC: U.S. Department of Transportation, 2009. Report No. FMCSA-RRR-09-042.

OLTEDAL, S. and RUNDMO, T., 2006. The effects of personality and gender on risky driving behaviour and accident involvement. *Safety science*, 44(7), 621-628

PANDE, A., CHAND, S., SAXENA, N., DIXIT, V., LOY, J., WOLSHON, B. and KENT, J.D., 2017. A preliminary investigation of the relationships between historical crash and naturalistic driving. *Accident Analysis & Prevention*, **101(3)**, 107-116.

PANOU, M., BEKIARIS, E., AND PAPAKOSTOPOULOS, V., 2007. Modelling driver behaviour in European Union and International Projects. In: P.C. CACCIABUE, ed. *Modelling driver behaviour in automotive environments*. London: Springer-Verlag, 2007, pp. 3–25.

PAPAZIKOU, E, QUDDUS, M, THOMAS, P, and KIDD, D., 2019. What came before the crash? An investigation through SHRP2 NDS data. *Safety Science*, (*in Press*). DOI: 10.1016/j.ssci.2019.03.010.

PAPAZIKOU, E, QUDDUS, M, THOMAS, P., 2017. Detecting Deviation from Normal Driving Using SHRP2 NDS Data. In *Transportation Research Board 2017*, Washington, US, January 2017.

PARKES, A., and FRANZEN, S.,1993. *Driving Future Vehicles*. 1St Ed. London: Taylor & Francis, pp 32-34.

PERKINS, S.R., and HARRIS, J.L. *Traffic Conflict Characteristics: Accident Potential at Intersections*. Warren, MI: General Motors Corporation, 1967.

PATTERSON, H.D. and THOMPSON, R., 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**(3), 545-554.

PERCHONOK, K., 1972. *Accident cause analysis. Report prepared for the US Department of Transportation*. Buffalo, NY: Cornell Aeronautical Laboratory.

PILGERSTORFER, M., RUNDA, K., BRANDSTÄTTER, C., CHRISTOPH, M., HAKKERT, S., TOLEDO, T., ISHAQ, R., and GATSCHA, M. *DaCoTA, 2012 - Deliverable 6.3, Report on Small Scale Naturalistic Driving Pilot*. Loughborough University, UK: Austrian Road Safety Board.

PLOCHL, M. and EDELMANN, J., 2007. Driver models in automobile dynamics application. *Vehicle System Dynamics*, **45**(7), 699–741.

PREYSS, A.E., 1968. *A theory and model of human learning behavior in a manual control task*. Cambridge, MA: MIT.

QUDDUS, M.A., 2006. *High Integrity Map Matching Algorithms for Advanced Transport Telematics Applications*. PhD Thesis, Imperial College, London.

QUIMBY, A.R., 1998. *In-car observation of unsafe driving actions*. Vermont, South Victoria: Australian Road Research Board. Research Report ARR No. 153.

RANNEY, T.A., 1994. Models of driving behavior: a review of their evolution. *Accident Analysis & Prevention*, **26**(6), 733–750.

RANNEY, T.A., 1999. Psychological factors that influence car-following and car-following model development. *Transportation Research Part F*, **2**(4), 213–219.

RANNEY, T.A., MAZZAI, E., GARROTT, R., and GOODMAN, M.J. *NHTSA, 2000. Driver distraction research: Past, present, and future*. Washington, DC: National Highway Traffic Safety Administration, USDOT.

RASMUSSEN, J., 1983. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, **13**(3), 257–266.

RASMUSSEN, J., 1987. The definition of human error and a taxonomy for technical system design. In: RASMUSSEN, J., DUNCAN, K., and LEPLAT, J., eds. *New technology and human error*. Chichester, U.K: Wiley, 1987, pp.32-42.

REAGAN, M.A., TRIGGS, T.J., YOUNG, K.L., TOMASEVIC, N., MITSOPOULOS, E., STEPHAN, K., and TINGVALL, C., 2006. *On-road Evaluation of Intelligent Speed Adaptation, Following Distance Warning and seatbelt Reminding System: Final Results of the Tac Safecar Project.* Victoria, Australia: Monash University Accident Research Centre.

REASON, J., 1990. *Human error.* New York: Cambridge University Press.RISSER, R., 1993. A study of the behaviour of drivers of electric cars. In: *Proceedings of the 26th International Symposium on Automotive Technology and Automation, Aachen, Germany, September, 1993.* Germany: Automotive Automation Limited, pp.529-538.

REASON, J., MANSTEAD, A., STRADLING, S., BAXTER, J. AND CAMPBELL, K., 1990. Errors and violations on the roads: a real distinction?. *Ergonomics*, 33(10), 1315-1332.

REED, S. et al., 2018. Identification of key risk factors related to serious road injuries and their health impacts, deliverable 7.4 of the H2020 project SafetyCube. Loughborough. Loughborough University.

REID, L. D., 1983. A survey of recent driver steering behavior models suited to accident studies. *Accident Analysis and Prevention*, **15**(1), 23–40.

RICHARDSON, M., BARBER, P., KING, P., HOARE, E., and COOPER, D., 1997. Longitudinal driver support systems. In: Proceedings of Autotech '97, 87-97, 4-6 Nov 1997, Bury St Edmunds, UK.

RISSER, R., 1985. Behavior in traffic conflict situations. *Accident Analysis and Prevention*, **17** (2), 179–197.

ROYAL, D., 2006. *National Survey of Speeding and Unsafe Driving Attitudes and Behaviour.* Washington DC, USA: National Highway Traffic Safety Administration.

RTI, 1970. Speed and accidents. Vols. I & II. Research Triangle Institute, RTI, North Carolina.

SAMANTHA H. HAUS, RINI SHERONY, H. C. G., 2018. Feasibility of using Naturalistic Driving Data to Characterise Vehicle-Pedestrian Crashes and Near-Crashes. *IRC-18-27 IRCOBI conference 2018*. 182–183.

SANDER, U. and LUBBE, N., 2018. The potential of clustering methods to define intersection test scenarios: Assessing real-life performance of AEB. Accident analysis and prevention. 113, 1-11.

SATZODA, R.K. and TRIVEDI, M.M., 2014. Drive analysis using vehicle dynamics and vision-based lane semantics. *IEEE Transactions on Intelligent Transportation Systems*, **16**(1), 9-18.

SAXBY, D.J., MATTHEWS, G., and HITCHCOCK, T., 2007. Fatigue states are multidimensional: Evidence from studies of simulated driving. In: *Proceedings of the Driving Simulation Conference North America, Iowa City, IA: University of Iowa, October, 2007*. Iowa: University of Iowa.

SAYED, T. and ZEIN, S., 1999. Traffic conflict standards for intersections. *Transportation Planning and Technology*, **22**(4), 309-323.

SAYER, J., LEBLANC, D., BOGARD, S., HAGAN, M., SARDAR, H., BUONAROSA, M.L., and BARNES, M. *Integrated Vehicle-based Safety Systems Preliminary Field Operational Test Plan*. The University of Michigan, Michigan: Transportation Research Institute, 2008.

SCANLON, J. M., KUSANO, K. D., and GABLER, H. C, 2015. Analysis of Driver Evasive Manoeuvring Prior to Intersection Crashes Using Event Data Recorders. *Traffic Injury Prevention*, **16**, 182–189.

SHANKAR, V., JOVANIS, P.P., AGUERO, J., and GROSS, F., 2008. Analysis of naturalistic driving data: a prospective view on methodological paradigms. *Transportation Research Record*, 2061(1), 1–8.

SHAW, L., and SICHEL, H., 1971. *Accident proneness*. 1st ed. Oxford: Pergamon Press.

SHERIDAN, T.B., 2004. Driver distraction from a control theory perspective. *Human Factors*, **46** (4), 587–599.

SHINAR, D. 2007. *Traffic Safety and Human Behavior*. Amsterdam: Elsevier.

SHINAR, D., 1984. The traffic conflict technique: A subjective vs. objective approach. *Journal of Safety Research*, **15**(4), 153-157.

SMILEY, A., 2005. What is driver distraction? In: *Proceedings of the International Conference on Distracted Driving. Toronto, Ontario, October 2005*. Canada: TIRF.

SOLOMON, D., 1964. Crashes on main rural highways related to speed, driver and vehicle. Bureau of Public Roads, U.S. Department of Commerce. United States Government Printing Office, Washington, D.C.

STAUBACH, M., 2009. Factors correlated with traffic accidents as a basis for evaluating Advanced Driver Assistance Systems. *Accident Analysis and Prevention*, **41**(5), 1025–33.

STRAHAN, C., WATSON, B., and LENNON, A., 2008. Can organisational safety climate and occupational stress predict work-related driver fatigue? *Transportation Research Part F*, 11(6), 418-426.

STUTTS, J., FEAGANES, J., REINFURT, D., RODGMAN, E., HAMLETT, C., GISH, K., and STAPLIN, L., 2005. Driver's exposure to distractions in their natural driving environment. *Accident Analysis and Prevention*, 37(6), 1093-1101.

SUGENO, M., 1985. *Industrial applications of fuzzy control*. New York: Elsevier.

SUMMALA, H., 1985. Modeling driver behavior: A pessimistic prediction? In: EVANS, L.., and SCHWING, R. C., eds. *Human behavior and traffic safety*. New York: Plenum Press, pp. 43-65.

SUMMALA, H., 1988. Risk control is not risk adjustment: The zero-risk theory of driver behavior and its implications. *Ergonomics* **31**(4), 491-506.

SUMMALA, H., 1996. Accident risk and driver behaviour. *Safety Science*, **22**(3), 103–117.

SUZDALEVA, E. and NAGY, I., 2018. An online estimation of driving style using data-dependent pointer model. *Transportation research part C: emerging technologies*, **86**, 23-36.

SVENSSON, Å., 1992. *Further development and validation of the Swedish traffic conflict technique*. PhD thesis, Lund University, Institute of Technology, Dept. of Traffic Planning & Engineering.

SVENSSON, Å., 1998. *A Method for Analyzing the Traffic Process in a Safety Perspective.* PhD thesis., Dept. of Traffic Planning and Engineering, Lund University, Lund, Sweden.

TALEBPOUR, A., MAHMASSANI, H.S., METE, F. and HAMDAR, S.H., 2014. Near-crash identification in a connected vehicle environment. *Transportation Research Record*, **2424** (1), 20-28.

TAMKE, A., DANG, T., and BREUEL, G., 2011. A flexible method for criticality assessment in driver assistance systems. In: *Proceedings of the IEEE Intelligent Vehicles Symposium. Baden-Baden, Germany.* New Jersey: IEEE, pp 697–702.

TAOKA, G.T., 1989. Brake Reaction Times of Unalerted Drivers. *ITE Journal*, **59** (3), 19–21.

TARKO, A.P., and SONGCHITRUKSA, P., 2005. Estimating frequency of crashes as extreme traffic events. In: *Proceedings of the 84th Transportation Research Board Annual Meeting*, *National Research Council, TRB, Washington, DC, 9th -13th January, 2005.* Washington, USA: Transportation Research Board.

TARKO, A.P., DAVIS, G, SAUNIER, N., SAYED, T., WASHINGTON, S., 2009. Surrogate Measures of Safety, White Paper.

TAYLOR, D.H., 1964. Drivers' galvanic skin response and the risk of accident. *Ergonomics*, **7**, 439-451.

THEEUWES, J., 2001. The effects of road design on driving. In: BARJONET, P.E., ed. *Traffic Psychology Today.* Boston: Kluwer Academic, pp. 241–264.

TIJERINA, L. ANGELL, L., AUSTRIA, A., TAN, A., AND KOCHHAR, D, 2003. *Driver Workload Metrics Literature Review (technical report)*. Washington, D.C: U.S. Department of Transportation.

TOPP, H.H, 1990. *Traffic safety work with video processing*. University Kaiserslautern: Transportation Department.

TOPP, H.H., 1998. Traffic Safety Work with Video-Processing. University Kaiserslautern, Transportation Department, Green Series No. 43, Kaiserslautern, Germany.

TREAT, J.R., TUMBAS, N.S., MCDONALD, S.T., SHINAR, D., HUME, R.D., MAYER, R.E., STANSIFER, R.L., AND CASTELLAN, N.J. 1979. Tri-level study of the causes of traffic accidents. Final report, vol. I. Causal factor tabulations and assessments. Indiana University Institute for Research in Public Safety.

TRIMPOP, R.M., 1996. Risk homeostasis theory: problems of the past and promises for the future. *Safety Science*, **22** (1), 119–130.

VAA, T., 2007. Modelling driver behaviour on basis of emotions and feelings: intelligent transport systems and behavioural adaptations. In: *P.C. Cacciabue, ed. Modelling driver behaviour in automotive environments*. London: Springer-Verlag, pp 208–232.

VAN DER HORST, A.R.A., 1990. *A Time-based Analysis of Road User Behaviour in Normal and Critical Encounters.* Ph.D. Thesis, Delft University of Technology, Delft.

VAN DER MOLEN H.H., and BOTTICHER, A. M. T, 1988. A hierarchical risk model for traffic participants. *Ergonomics,* **31**(3), 537-555.

VAN DER MOLEN H.H., AND BOTTICHER, A. M. T., 1987. Risk models for traffic participants: A concerted effort for theoretical operationalizations. In: Rothengatter, J. A.., and de Bruin, R. A., eds. Road users and traffic safety. Assen, The Netherlands: Van Gorcum.

VANAREM, B., and DEVOS, A.P., 1997. *The Effect of a Special Lane for Intelligent Vehicles on Traffic Flows. TNO-INRO Report 1997-02a*. TU Delft: The Netherlands.

VÁRHELYI, A., 1996. Dynamic Speed Adaptation Based on Information Technology—A Theoretical Background, Bulletin 142. Dept. of Traffic Planning and Engineering, Lund University, Lund, Sweden.

VÁRHELYI, A., 1996. *Dynamic Speed Adaptation Based on Information Technology—A Theoretical Background*. Phd Thesis, Lund University, Lund, Sweden.

VARHELYI, A., 1998. Drivers' speed behaviour at a zebra crossing: a case study. *Accident Analysis & Prevention*, **30**(6), 731-743.

VICTOR, T., 2005. *Keeping eye and mind on the road*. PhD Thesis, Uppsala University, Sweden.

VICTOR, T., BÄRGMAN, J., HJÄLMDAHL, M., KIRCHER, K., SVANBERG, E., HURTIG, S., GELLERMAN, H., and MOESCHLIN, F., 2010. *Sweden-Michigan* *naturalistic field operational test (semifot) phase 1: Final report.* Chalmers University, Sweden: SAFER Vehicle and Traffic Safety Centre.

VLAHOGIANNI, E.I., YANNIS, G. and GOLIAS, J.C., 2014. Detecting Powered-Two-Wheeler incidents from high resolution naturalistic data. Transportation Research Part F. 22, 86-95.

VOGEL, K., 2003. A comparison of headway and time to collision as safety indicators. *Accident Analysis and Prevention,* **35** (3), 427–433.

WANG, J., DIXON, K.K., LI, H. and OGLE, J., 2005. Normal deceleration behavior of passenger vehicles at stop sign–controlled intersections evaluated with in-vehicle Global Positioning System data. *Transportation research record*, **1937**(1), 120-127.

WANG, J-S., KNIPLING, R.R., and GOODMAN, M.J., 1996. The role of driver inattention in crashes: New statistics from the 1995 Crashworthiness Data System. In: *Conference proceedings of the 40th annual meeting of the Association for the Advancement of Automotive Medicine. Vancouver, British Columbia, 1996.* Des Plaines, USA: AAAM, pp 377-392.

WANG, X., WANG, T., TARKO, A., and TREMONT, P. J, 2015. The influence of combined alignments on lateral acceleration on mountainous freeways: A driving simulator study. *Accident Analysis and Prevention*, **76**, 110–117.

WEGMAN, F., 2017. The future of road safety: A worldwide perspective. *IATSS Research*, 40 (2) 66–71.

WEGMAN, F., BERG, H.Y., CAMERON, I., THOMPSON, C., SIEGRIST, S. and WEIJERMARS, W., 2015. Evidence-based and data-driven road safety management. IATSS research, **39**(1), 19-25.

WEIR, D.H., AND MCRUER, D.T., 1968. A theory for driver steering control of motor vehicles. In: *Road user characteristics, Highway Research Record*. Washington DC: Highway Research Board, pp. 7-28.

WELSH, R., MORRIS, A., HASSAN, A. and CHARLTON, J., 2006. Crash characteristics and injury outcomes for older passenger car occupants. Transportation Research Part F, **9**, 322-334.

WIERWILLE, W.W., and GAGNE, G.A., 1966. Non-linear and time varying dynamical models of human operators in manual control systems. *Human Factors*, **8**, 97-120.

WICKENS, C., HOLLANDS, J.G., BANBURY, S. and PARASUAMAN, R., 2016. *Engineering Psychology and Human Performance*. 4th edition. New York, NY: Routledge.

WILCOXON, F, 1945. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, **1** (6), 80-83.

WILDE G.J.S., 1984. Evidence refuting the theory of risk homeostasis? a rejoinder to Frank P. McKenna. *Ergonomics*, **27**, 297-304.

WILDE, G. J. S., 1988. Risk homeostasis theory and traffic accidents: Propositions, deductions and discussion of dissension in recent reactions. *Ergonomics* **31,** 441-468.

WILDE, G.J.S, 1982. The theory of risk homeostasis: Implications for safety and health. *Risk Analysis*, **2**, 209-225.

WILDE, G.J.S. and KUNKEL, E., 1984. Die begriffliche and empirische Problematik der Risikokompensation. Z. Verkehrssicherheit, 30: 52-62.

WILDE, G.J.S., 1978. *The risk compensation of accident causation and its practical implications for accident paralysis*. *Folders for accident medicine*, **130**, 134-156.

WILDE, G.J.S., and MURDOCH, P., 1982. Incentive systems for accident-free and violence-free driving in the general population. *Ergonomics*, **25**, 879-890.

WILDE, G.J.S., ROBERTSON, L.S., and PLESS, I.B., 2002. For and against: does risk homoeostasis theory have implications for road safety. *British Medical Journal*, **324**, 1149–1152.

WILLIAMS, M.J., 1981. Validity of the traffic conflicts technique. *Accident Analysis and Prevention*, **13** (2), 133–135.

WINSUM, W., 1999. The human element in car following models. *Transportation Research Part F*, **2**(4), 207–211.

WOLF, J.D., and BARRETT, M.F, 1979. *Driver-vehicle effectiveness model; volume Appendices*. Washington, DC: Department of Transportation.

WORLD HEALTH ORGANIZATION, 2015. World health statistics 2015. World Health Organization.

WU, K. F., and LIN, Y. J., 2019. Exploring the effects of critical driving situations on driver perception time (PT) using SHRP2 naturalistic driving study data. Accident Analysis and Prevention, **128**(4), 94–102.

WU, K., AGUERO-VALVERDE, J., and JOVANIS, P.P., 2014. Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. *Accident Analysis & Prevention*, **72**, 210-218.

WU, K., AND THOR, C. P., 2015. METHOD for Using Naturalistic Driving Study Data to Analyze Rear-end Crash Sequences. In: *Proceedings of TRB 2015 annual meeting. Washington DC, January 11th-15th, 2015.* Washington DC: Transportation Research Record, pp 27-36.

WU, K.-F., and JOVANIS, P. P., 2012. Crashes and crash-surrogate events: exploratory modelling with naturalistic driving data. *Accident Analysis and Prevention*, **45**, 507–16.

WU, K.F., AND JOVANIS, P. P. 2013. Defining and screening crash surrogate events using naturalistic driving data. *Accident Analysis and Prevention*, **61**, 10–22.

YAO, J., YANG, J. and OTTE, D., 2007. Head injuries in child pedestrian accidents-In-depth case analysis and reconstructions. Traffic Injury Prevention, 8:1, 94-100.

YIN, R.K, 2009. *Case Study Research: Design and Methods*. 4th edition. London, UK: SAGE Publications.

YOUNG, K., LEE, J.D., REGAN, M.A.. 2008. *Driver distraction: Theory, effects and mitigation*. New York: CRC Press.

YOUNG, L.R., 1969. On adaptive manual control. *Ergonomics*, **12**(4), 635-674.

YOUNG, M. S., BIRRELL, S. A., & STANTON, N. A., 2011. Safe driving in a green world: A review of driver performance benchmarks and technologies to support ‘smart’ driving. *Applied Ergonomics,* **42**(4), 533-539.

ZADEH, L.A., 1965. Fuzzy sets, *Information and Control*, **8**, 338-353.

ZADEH, L.A., 1973. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics*, **3** (1), 28-44.

ZADEH, L.A., 1989. Knowledge representation in fuzzy logic. *IEEE Transactions on Knowledge and Data Engineering*, **1** (83), 28-32.Zaidel, D., & Noy, I., 1997. Automatic Versus Interactive Vehicle Navigation Aids. In: Ian Noy. Ergonomics and Safety of Intelligent Driver Interfaces. Mahwah, New Jersey. Lawrence Erlbaum associates, Publishers, 291- 292.

ZADEH, L.A., 1999. From computing with numbers to computing with words. From manipulation of measurements to manipulation of perceptions, Circuits and Systems I: Fundamental Theory and Applications. *IEEE Transactions*, **46** (1), 105-119.

ZAIDEL, D., 1992. *Quality of driving with route guidance assistance. An evaluation methodology.* Canada: Transport Canada Report.

ZAIDEL, D., AND NOY, I., 1997. Automatic Versus Interactive Vehicle Navigation Aids. In: Ian Noy, ed., *Ergonomics and Safety of Intelligent Driver Interfaces*. Mahwah, New Jersey: Lawrence Erlbaum associates, pp.291- 292.

ZHAO, Y., 1997. *Vehicle Location and Navigation System*. 1st edition. MA: Artech House, Inc.

ZHOU, J., PENG, H. and GORDON, T.J., 2008. Characterization of the lateral control performance by human drivers on highways. *SAE International Journal of Passenger Cars-Mechanical Systems*, **1**(5), 450-458.

Appendix A

Pilot study questionnaire & Adult participant information sheet

Pilot study Questionnaire

1.Do you think you have **deviated** from normal driving during this route?

A.Yes **B.**No

2.**Where** do you think this happened and **why**?

- **A.** Road characteristics
- **B.** Traffic
- **C.** Weather
- **D.** Distractions
- **E.** Fatigue
- **F.** A combination of them
- **G.** Other …….Please specify……………………….

3. In a scale from 1- 6 (1 not at all - 6 very very tired), how much tired do you feel?

4. Have you felt distracted anytime during the route? What was the reason?

A.Yes **B.**No

5.What was the **most challenging** about this route? **Why**?

6.How many years have you been driving?

7. Age:

Waiting Time in a junction for the suitable gap:

Gap acceptance behaviour:

Car following behaviour:

Glances off the road:

Eyes and head movement behaviour:

General comments about the driver/participant:

Adult Participant Information Sheet

Project title: Driving study in order to investigate methods to detect deviations from normal driving

Main investigator: Evita Papazikou- Safe and smart mobility cluster (LDS) &

Transport Studies Group (CBE), Loughborough Design School, LDS.1.25, Loughborough University, LE11 3TU

Email: A.Papazikou@lboro.ac.uk

Supervisor: Professor Pete Thomas- Loughborough Design School, LDS.1.10, Loughborough University, LE11 3TU

Email: P.D.Thomas@lboro.ac.uk

What is the purpose of the study?

The purpose of the study is to gather data in order to validate the methods to collect and analyse naturalistic driving behaviour data, explain the deviation from normal driving and find methods to detect it.

Who is doing this research and why?

The study is part of a student PhD Research project supported by Loughborough University. It is being undertaken by Evita Papazikou and supervised by Prof Pete Thomas. It is investigating the deviation from normal driving and the methods to detect it.

Are there any exclusion criteria?

Yes, only university registered drivers.

What will I asked to do?

Participants will be asked to drive a test route using the fully instrumented university vehicle with the researcher as passenger. The test will start at Design School and will end at Design School. The participants will be asked questions about the driving task whilst driving and a questionnaire will be filled in afterwards in a form of a small interview.

Prior to driving the test route participants will have approximately 5 minutes to drive around campus with the researcher to familiarise themselves with the test car.

The safety of participants, researcher and the other road users will be of paramount importance at all times.

Once I take part, can I change my mind?

Yes, of course! After you have read this information and asked any questions you may have, you can withdraw from the study at any time, for any reason and you will not be asked to explain your reasons for it.

How long will it take?

The whole process will take approximately one hour.

What personal information will be required?

No specific personal information is required.

Are there any risks in participating?

There are no specific risks in participating.

Will my taking part in this study be kept confidential and what will happen to the results?

Yes, your taking part in the study will be kept confidential. No personal details that identify you will be held and you will not be named in the study. The video recordings will be kept in a secure place and destroyed within 10 years. Storage of data will comply with the Data Protection Act 1998. The results will form part of the researcher's dissertation.

I have some more questions; who should I contact?

The researcher Evita Papazikou or Supervisor Pete Thomas.

What if I am not happy with how the research was conducted?

If you are not happy with how the research was conducted, please contact Samantha Porter, responsible for the Ethics Approvals (Human participants) Sub-Committee,

LDS 1.17, tel: +44 (0)1509 222782, **Email:** c.s.porter@lboro.ac.uk

Date:

Participant signature:

Appendix B

VTTI COST PROPOSAL

VTTI COST PROPOSAL (DATA REQUEST)

SHRP 2 Data Support for Loughborough University

Submitted to:

Evita Papazikou
Loughborough University
Tel: (307) 766-5550

Prepared by:

Miguel Perez
Virginia Tech Transportation Institute
3500 Transportation Research Plaza (0536)
Blacksburg, VA 24061
(540) 231-1537

February 3, 2016

General Statement of Work, Deliverable Schedule, and Budget

Objective

The purpose of this effort is to support the use of various components of the SHRP 2 database by personnel from Loughborough University. The proposed project aims to examine normal driving and the stages of crash/near-crash sequence development using the SHRP 2 naturalistic driving data available on InSight combined with expanded time series data for these epochs.

Approach

The initial activity in this contract will be to assist Loughborough University with the execution of a final data sharing agreement for the access of SHRP 2 NDS data they require to complete this project. For this request, NDS data will be limited to crashes and near-crashes that have been identified, reduced, and published on the InSight website.

The main event dataset that Loughborough University will access is specified as follows:

- Vehicles
  - Anonymous Vehicle ID
  - Vehicle classification
  - Advanced technology Vehicle
  - Model Year
- Events: only events identified as a crash and near-crash in either event severity (1 or 2)
  - Variables from the event table:
    - Event ID
    - Anonymous Participant ID
    - Event severity 1
    - Event start
    - Event end
    - Subject Reaction start
    - Impact or proximity time
    - Pre-incident maneuver
    - Maneuver judgment
    - Precipitating event
    - Event nature 1
    - Incident type 1
    - Crash severity 1
    - V1 evasive maneuver 1
    - V1 post-maneuver control 1
    - Driver behaviour(s)
    - Secondary task(s) start time, end time and outcome
    - Hands on the wheel
    - V1 lane occupied

        - Vehicle contributing factors
        - Visual Obstructions
        - Lighting
        - Traffic control
        - Relation to junction
        - Intersection influence
        - Grade
        - Fault
        - Through travel lanes
        - Contiguous travel lanes
        - Infrastructure
        - Weather
        - Surface condition
        - Traffic density
        - Alignment
        - Locality
        - Final narrative
- Questionnaires: for all participants with a crash, near-crash, or a baseline identified in the study (i.e., participants that had some driving time in the study)
    - Driver Demographic Questionnaire
        - Anonymous Participant ID
        - Gender
        - Age group
        - Driver mileage last year
        - Participant receive license
    - Risk Perception Questionnaire
        - Red light
        - Driving sleepy
        - Illegal turns
        - Yellow light acceleration
        - Drinking after taking drugs and alcohol
        - Driving while talking
        - Checking rearview mirror
        - Not wearing safety belt
        - Sudden lane changes
        - Running stop sign
        - Speeding for thrill
        - Failure to yield
        - Tailgating
        - In a hurry
        - Bad weather
        - Secondary tasks
        - Eyes off road
        - Speeding more than 20MPH over limit
        - Not yielding to pedestrians
        - Risk perception score
    - Risk Taking Questionnaire
        - The same variables with the risk perception questionnaire
        - CARDS Frequency of Risky Behavior Score
    - Driver Behaviour Questionnaire
        - Impatiently pass on the right
        - Tailgating often
        - Wrong switch
        - Run red light

        - Miss pedestrians
        - Road rage
        - Disregards speed limits
        - Miss lead vehicle
        - Driving above alcohol limit
        - Underestimate speed of oncoming
        - Fail to check rearview Mirror
        - Involve in racing
        - Brake aggressively
    - Sleep Habits Questionnaire
        - Sleep schedule
        - Average Sleep needed
        - Average sleep hours when working
        - Sleep duration
        - Quality of sleep
        - Well being while awake
        - Fatigued while awake
        - Sleepiness while awake
- Trips: only for trips that involved crashes or near-crashes included in the event table provided for this project
    - Variables from the trip summary table
        - Trip start local time
        - Trip end local time
        - Trip distance
        - Maximum speed
        - Mean speed
        - Maximum longitudinal acceleration
        - Minimum longitudinal deceleration
        - Maximum lateral acceleration
        - Minimum lateral acceleration
        - Maximum turn rate
        - Minimum turn rate
        - Number of longitudinal accelerations > threshold
        - Number of longitudinal decelerations > threshold
        - Number of lateral accelerations > threshold
        - Number of brake activations
        - Turn signal activations
        - ABS available
        - ABS Activation
        - Time spent at the different speed bins
        - Distance spent at the different speed bins
        - Vehicle model year
        - Lane Tracker Right-side High Quality Time
        - Lane Tracker Left-side High Quality Time
        - Cell phone flag
        - Alcohol flag
        - Speed limit
        - Time where headway = (all)
        - Distance where headway = (all)
        - Minimum Time to Collision (to lead vehicles)
- **Time series data** for crashes or near-crashes included in the event table provided for this project. The time series data will encompass a period of up to two minutes preceding the crash or near-crash event. In some circumstances, a full two minutes may not be available due to proximity to trip

origin. This may also result in latitude and longitude for some near-crash events not being exported at all to avoid disclosing trip origin.

- All variables except for:
    - Occupancy snapshots
- Note: latitude and longitude can only be provided for near-crashes.

These different datasets will be provided as Excel tables or, when this is not possible due to file size, as csv files. Specifically, time series data will be provided as separate csv files, one per event, uniquely identified. The tables and files will be linked to one another by anonymized participant ID, file ID, event ID, and/or anonymized vehicle ID.

VTTI understands that Loughborough University is not currently interested in obtaining any video snippets corresponding to these events as part of this data export.

The expected period of performance for this statement of work is two months.

Deliverables

We will deliver to Loughborough University the datasets and information described in the previous section. We will adhere to the planned period of performance of two months.

Budget

The expected estimated fixed-price cost for the proposed tasks is $11,296.

Appendix D

MATLAB code

Initial data preparation

```
clear all

clc

 load('data.mat')

y=data(:,1);

y1=find(isnan(y));

data(y1,:)=[];

t=find(data(:,1)==2);

data(t,:)=[];


load('d2.mat')

data=[data d2];


x1=data(:,1);

x_log1=diff(x1);

x_log1=[0; x_log1]; %when not 0 event changes

actualevents=find(x_log1~=0);

actualevents=[1; actualevents];

actualevents(:,2)=data(actualevents(:,1),1);

%%%%%%%%%%%%%%%%%%%%%create a copy

data2=data; %for dec_x

 t1=find(isnan(data2(:,4))); %remove nans

data2(t1,:)=[];

 t2=find(isnan(data2(:,173))); %remove nans

data2(t2,:)=[];
```

```
%%%%%%%%%%%%%%%%%%%%%

data6=data; %for gyro_all

 t9=find(isnan(data6(:,6))); %remove nans

data6(t9,:)=[];

 t10=find(isnan(data6(:,198))); %remove nans

data6(t10,:)=[];

 %%%%%%%%%%%%%%%%%%%%%

 %find greater or equal to zero deceleration to delete it

acc=find(data2(:,4)>=0);

data2(acc,:)=[];

 %find smaller or equal to zero deceleration_y to delete it

acc_y_pos=find(data4(:,5)<=0);

data4(acc_y_pos,:)=[];

 %find greater or equal to zero deceleration_y to delete it

acc_y_neg=find(data5(:,5)>=0);

data5(acc_y_neg,:)=[];

 %find smaller or equal to zero yaw to delete it

yaw_y_pos=find(data7(:,6)<=0);

data7(yaw_y_pos,:)=[];

 %find greater or equal to zero yaw rate to delete it

yaw_y_neg=find(data8(:,6)>=0);

data8(yaw_y_neg,:)=[];
```

Development of the algorithm for event extraction

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    %find which ones are greater than 3sd

    dec=data2(:,4);

    dec(:,2)=data2(:,173);

    dec(:,1)=abs(dec(:,1));

    st=find(dec(:,1) < dec(:,2));

    data2(st,:)=[];

    %finding all gyro & find which ones are greater than 3sd

    yaw_yall=data6(:,6); %all

    yaw_yall(:,2)=data6(:,198); %3sd  %%%%%%%%%%%%

    yaw_yall(:,1)=abs(yaw_yall(:,1));

    st_yawall=find(yaw_yall(:,1) < yaw_yall(:,2));

    data6(st_yawall,:)=[];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%% DECELERATION_X%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

     %find events

    x=data2(:,1);

    x_log=diff(x);

    x_log=[0; x_log];

    event=find(x_log~=0);

    event=[1; event];

    event(:,2)=data2(event(:,1),1);

     %create template

    template_dec_x=zeros(774, 33);

    template_dec_x(:,1)=actualevents(:,2);
```

```
%events which have no values

tf = setdiff(actualevents(:,2),event(:,2));

counter6=1;

 for i=1:length(event)

    clear y3 y7 y4 x k1 s b y8 y9 acc_mean acc_min acc_max

    y4=find(event(i,2)==template_dec_x(:,1));

    y3=find(data2(:,1)==event(i,2));

    y7=data2(y3,2);

    y8=data2(y3,168);

    y9=data2(y3,189);

    %find consecutive numbers

    k1 = [true;diff(y7(:))~=1 ];

    s = cumsum(k1);

    x = histc(s,1:s(end));


    %change template size

    b = cumsum(x);

    act=y3(1)+b;

    y3(:,2)=data2((y3(:,1)),4);

    y3(:,3)=data2((y3(:,1)),2);

    counter3=1;

    for k=1:length(b)

        counter4=counter3:1:b(k);

        acc_mean(k)=mean(y3(counter4,2));

        acc_min(k)=min(y3(counter4,2));

        acc_max(k)=max(y3(counter4,2));
```

```
        if counter4==1
            timestamp_start(k)=y3(counter4,3);
            timestamp_end(k)=y3(counter4,3);
        else
            timestamp_start(k)=y3(counter4(1),3);
            timestamp_end(k)=y3(counter4(end),3);
        end
        counter3=b(k)+1;
        accrange{i,k}=y3(counter4,2);
        TTC_dec{i,k}=y8(counter4,1);
        yaw_dec{i,k}=y9(counter4,1);
        timestamp{i,k}=y7(counter4);
    end
    acc_mean=acc_mean';
    acc_min=acc_min';
    acc_max=acc_max';
    counter=1;
    counter1=2;
    counter5=1;
     for j=1:length(x)
        if (x(j)<10) %value to change refers to 1s
            counter=counter+1;
        else
            template_dec_x(y4,counter1)=x(j);
            template_dec_x(y4, counter1+1)=acc_mean(j); %mean
            template_dec_x(y4, counter1+2)=acc_min(j); %min
            template_dec_x(y4, counter1+3)=acc_max(j); %max
```

```
                 template_dec_x(y4,   counter1+4)=timestamp_start(j);
%timestamp_start

                 template_dec_x(y4,     counter1+5)=timestamp_end(j);
%timestamp_end

                 decvalues{y4, counter5}(:,1)=timestamp{counter6,j};

                 decvalues{y4, counter5}(:,2)=accrange{counter6,j};

                 decvalues{y4, counter5}(:,3)=TTC_dec{counter6,j};

                 decvalues{y4, counter5}(:,4)=yaw_dec{counter6,j};

                 yyy{y4, counter5}=x(j);

                 counter1=counter1+6;

                 counter5=counter5+1;

              end

         end

         counter6=counter6+1;

        end

      counterC=0;

     for c=1:length(decvalues)

         clear n n1

         n=decvalues(c,:);

         n=n(~cellfun('isempty',n));

         n1=cell2mat(n');

         if (~isempty(n1))

n2(counterC+1:(counterC+length(n1)),1)=template_dec_x(c,1);

             n2(counterC+1:(counterC+length(n1)),2:5)=n1;

         end

         counterC=length(n2);

     end

      EventTimestampDecTTCYaw=n2;
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%% YAW RATE Absolute values%%%%%%%%%%%%%%%%%%%%%%%%%%%

     %find events

    x5=data6(:,1);

    x_log5=diff(x5);

    x_log5=[0; x_log5]; %when not 0 event changes

    event5=find(x_log5~=0);

    event5=[1; event5];

    event5(:,2)=data6(event5(:,1),1);

     %create template

    template_yaw_all=zeros(774, 43);

    template_yaw_all(:,1)=actualevents(:,2);

     %events which have no values

    tf5 = setdiff(actualevents(:,2),event5(:,2));

    counter6=1;

     for i=1:length(event5)

        clear y3 y7 y4 x k1 s b y8 y9 acc_mean acc_min acc_max
timestamp_start timestamp_end



        y4=find(event5(i,2)==template_yaw_all(:,1));

        y3=find(data6(:,1)==event5(i,2));

        y7=data6(y3,2); %timestamp

        y8=data6(y3,168); %TTC

        y9=data6(y3,4); %decelerationx
```

```
%find consecutive numbers

    k1 = [true;diff(y7(:))~=1 ];

    s = cumsum(k1);

    x = histc(s,1:s(end));

    b = cumsum(x);

    act=y3(1)+b;

    y3(:,2)=data6((y3(:,1)),6);

    y3(:,3)=data6((y3(:,1)),2);


    counter3=1;

    for k=1:length(b)

        counter4=counter3:1:b(k);

        acc_mean(k)=mean(y3(counter4,2));

        acc_min(k)=min(y3(counter4,2));

        acc_max(k)=max(y3(counter4,2));

        if counter4==1

            timestamp_start(k)=y3(counter4,3);

            timestamp_end(k)=y3(counter4,3);

        else

            timestamp_start(k)=y3(counter4(1),3);

            timestamp_end(k)=y3(counter4(end),3);

        end

        counter3=b(k)+1;

        yawrange_all{i,k}=y3(counter4,2);

        TTC_yawall{i,k}=y8(counter4,1);

        dec_yawall{i,k}=y9(counter4,1);

        timestamp_yaw_all{i,k}=y7(counter4);
```

```
        end


        acc_mean=acc_mean';

        acc_min=acc_min';

        acc_max=acc_max';

        counter=1;

        counter1=2;

        counter5=1;

        for j=1:length(x)

            if (x(j)<7) %value to change refers to 0.7s

                counter=counter+1;

            else

                template_yaw_all(y4,counter1)=x(j);

                template_yaw_all(y4, counter1+1)=acc_mean(j); %mean

                template_yaw_all(y4, counter1+2)=acc_min(j); %min

                template_yaw_all(y4, counter1+3)=acc_max(j); %max

                template_yaw_all(y4, counter1+4)=timestamp_start(j);
%timestamp_start

                template_yaw_all(y4,  counter1+5)=timestamp_end(j);
%timestamp_end

yawvalues_all{y4,counter5}(:,1)=timestamp_yaw_all{counter6,j};

yawvalues_all{y4,counter5}(:,2)=yawrange_all{counter6,j};

yawvalues_all{y4,counter5}(:,3)=TTC_yawall{counter6,j};

yawvalues_all{y4,counter5}(:,4)=dec_yawall{counter6,j};

                yyy_yaw_all{y4, counter5}=x(j);
```

```
                 counter1=counter1+6;

                 counter5=counter5+1;

                 end

            counter6=counter6+1;

         end

      counterC=0; clear n2

     for c=1:length(yawvalues_all)

         clear n n1

         n=yawvalues_all(c,:);

         n=n(~cellfun('isempty',n));

         n1=cell2mat(n');

         if (~isempty(n1))

n2(counterC+1:(counterC+length(n1)),1)=template_yaw_all(c,1);

             n2(counterC+1:(counterC+length(n1)),2:5)=n1;

         end

         counterC=length(n2);

     end

      EventTimestampYawTTCDec=n2;
```

Algorithm for the sequence of overlapping manoeuvres

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%YAW RATE%%%%%%%%%%%%%%%%%%%%%%%%%%%

     for i=1:length(event5)

         v11{i}=find(event5(i,2)==EventTimestampYawTTCDec(:,1));

         v22{i}=find(event5(i,2)==EventTimestampDecTTCYaw(:,1));

     end
```

```
    %compare ts
    for j=1:length(v11)
        w11{j}=EventTimestampYawTTCDec(v11{j},2);
        w22{j}=EventTimestampDecTTCYaw(v22{j},2);
    end
    %find overlapping ts
    c2=1;
     for k=1:length(w11)
        clear p11 p22
        p11=w11{k};
        p22=w22{k};
        if (isempty(p11) || isempty(p22))
            c2=c2+1;
        else
            tfpp(k)=isequal(p11, p22);
        end
    end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%DECELERATION_X%%%%%%%%%%%%%%%%%%
    for i=1:length(event)
        v1{i}=find(event(i,2)==EventTimestampYawTTCDec(:,1));
        v2{i}=find(event(i,2)==EventTimestampDecTTCYaw(:,1));
    end
     %compare ts
    for j=1:length(v1)
        w1{j}=EventTimestampYawTTCDec(v1{j},2);
        w2{j}=EventTimestampDecTTCYaw(v2{j},2);
    end
```

```
    %find overlapping ts

    c2=1;

     for k=1:length(w1)

        clear p1 p2

        p1=w1{k};

        p2=w2{k};

        if (isempty(p1) || isempty(p2))

            c2=c2+1;

        else

            tfp(k)=isequal(p1, p2);

        end

    end

%%%%%%%%%%%%%%% OVERLAPPING EVENTS%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    for nf=1:length(actualevents)

            clear b1 c1 b2 c2 c3

        cv=actualevents(nf,2);

        b1=find(EventTimestampDecTTCYaw==cv);

        c1=EventTimestampDecTTCYaw(b1,2);

        b2=find(EventTimestampYawTTCDec==cv);

        c2=EventTimestampYawTTCDec(b2,2);

        c1(:,2)=1; %dec

        c2(:,2)=2; %yaw

        c3=[c1;c2];

        [~,idx] = sort(c3(:,1)); % sort just the first column

        sortedmat{nf} = c3(idx,:);    % sort the who

    end
```

```
     for jj=1:length(sortedmat)
        x100=sortedmat{1,jj}(:,1);
        [U,I]=unique(x100(:,1));
        repeated{1,jj} = setdiff(1:size(x100,1),  I);
    end
     f3=find(~cellfun(@isempty,repeated)); % events where timestamps overlap
    f3eventnumber=actualevents(f3,2);
    countergh=0;
    counterX=0;
    for jj=1:length(sortedmat)
        clear br1 br2 split v3
        br1=diff(repeated{1,jj});
        br2=find(br1>2);
        if (isempty(br1))
            counterX=counterX+1;
        end
        if (~isempty(br1) && isempty(br2))
            if(sortedmat{1,jj}(repeated{1,jj}(1,1),2)==2)
                v3(1,1)=1;
            end
            if(sortedmat{1,jj}(repeated{1,jj}(1,1),2)==1)
                v3(1,1)=2;
            end
            if (sortedmat{1,jj}(repeated{1,jj}(1,end),2)==2)
                v3(1,2)=2;
            end
```

```
            if (sortedmat{1,jj}(repeated{1,jj}(1,end),2)==1)

                v3(1,2)=1;

            end

            v3(1,3)=sortedmat{1,jj}(repeated{1,jj}(1,1),1);

            v3(1,4)=sortedmat{1,jj}(repeated{1,jj}(1,end),1);



            v3=v3';

            v3=v3(:);

            v3=v3';

            v4{jj}=v3;

        end

        if (~isempty(br1) && ~isempty(br2))

            countergh=countergh+1;

            testret{countergh}=jj; %gives where more than one overlap
occurs in the same event

            for i=1:length(br2)+1

                if i==1

                    split{i}=repeated{1,jj}(1:br2(i));

                end

                if i==(length(br2)+1)

                    split{i}=repeated{1,jj}(br2(i-1)+1:end);

                end

                if (i ~=1 && i~=(length(br2)+1))

                    split{i}=repeated{1,jj}(br2(i-1)+1:br2(i));

                end

            end

             for jk=1:length(split)

                if(sortedmat{1,jj}(split{1,jk}(1,1),2)==2)
```

```
                v3(jk,1)=1;
            end
            if(sortedmat{1,jj}(split{1,jk}(1,1),2)==1)
                v3(jk,1)=2;
            end


            if (sortedmat{1,jj}(split{1,jk}(1,end),2)==2)


                v3(jk,2)=sortedmat{1,jj}(split{1,jk}(1,end),2);
            end


            if (sortedmat{1,jj}(split{1,jk}(1,end),2)==1)
                v3(jk,2)=sortedmat{1,jj}(split{1,jk}(1,end),2);
            end
            v3(jk,3)=sortedmat{1,jj}(split{1,jk}(1,1),1);
            v3(jk,4)=sortedmat{1,jj}(split{1,jk}(1,end),1);
        end
        v3=v3';
        v3=v3(:);
        v3=v3';
        v4{jj}=v3;
    end


    %v1=sortedmat{1,jj}(v,2)
    %v2=sortedmat{1,jj}(v,1)-sortedmat{1,jj}(v-1,1);
    %if (v2>5)
end
```

```
 % % %%%%%%%%%write in a template which came first

% % testcia=zeros(774,49);

% % testcia(:,1)=actualevents(:,2);

% % counterCS=0;

% % for i=1:length(actualevents)

% % lt=length(v4{1,i});

% % if lt==0

% % counterCS=counterCS+1;

% % else

% % testcia(i, 2:lt+1)=v4{1,i};

% % end

% % end
```

Percentile values across different speed bins

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%find speed bins and find percentile
for %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%different bins%%%%%%%%%%%%%%%%

    xx=0:10:110;

    for i=1:length(xx)

        if i==length(xx)

            xxx{i,1}=find(data(:,20)>xx(i)); %index

            xxx{i,2}=(data(xxx{i,1},20)); %speed

            xxx{i,3}=(data(xxx{i,1},168)); %TTC

            xxx{i,4}=(data(xxx{i,1},4)); %acc

            xxx{i,5}=(data(xxx{i,1},5)); %acc lat

            xxx{i,6}=(data(xxx{i,1},6)); %yaw


        else
```

```
            xxx{i,1}=find(data(:,20)>xx(i)  &  data(:,20)<=xx(i+1));
%index

            xxx{i,2}=(data(xxx{i,1},20)); %speed

            xxx{i,3}=(data(xxx{i,1},168)); %TTC

            xxx{i,4}=(data(xxx{i,1},4)); %acc

            xxx{i,5}=(data(xxx{i,1},5)); %acc lat

            xxx{i,6}=(data(xxx{i,1},6)); % yaw

            end

    end

     %find percentile table

    for i=1:length(xx)

        clear b1 b2 c1 c2 c3 d1 d2 d3

        Y (i,2)= prctile(xxx{i,3},10); %TTC

        b1=find(xxx{i,4}>0); %find where positive acc is

        b2=find(xxx{i,4}<0); %find where negative acc is

        Y (i,3)= prctile(xxx{i,4}(b1),10); %+ve acc

        Y (i,4)= prctile(xxx{i,4}(b2),10); %-ve acc

        c1=find(xxx{i,5}>0); %find where positive lat acc is

        c2=find(xxx{i,5}<0); %find where negative lat acc is

        c3=abs(xxx{i,5}); %abs lat acc

        Y (i,5)= prctile(xxx{i,5}(c1),10); %+ve acc

        Y (i,6)= prctile(xxx{i,5}(c2),10); %-ve acc

        Y (i,7)= prctile(c3,10); %abs acc

        d1=find(xxx{i,6}>0); %find where positive yaw is

        d2=find(xxx{i,6}<0); %find where neg yaw is

        d3=abs(xxx{i,6}); %find where abs yaw is

        Y (i,8)= prctile(xxx{i,6}(d1),10); %+ve acc

        Y (i,9)= prctile(xxx{i,6}(d2),10); %-ve acc
```

```
        Y (i,10)= prctile(d3,10); %abs acc
        end
```

5th percentile indicators

```
    %find 5th percentile of each timestamp
    ts=1:1:1500;
    for j=1:length(ts)
        Z{j,1}=find(data(:,2)==ts(j));
        Z{j,2}=(data(Z{j,1},168)); %TTC
        Z{j,3}=(data(Z{j,1},4)); %acc
        Z{j,4}=(data(Z{j,1},5)); %acc lat
        Z{j,5}=(data(Z{j,1},6)); %yaw
    end
     %find percentile table
    for j=1:length(ts)
        clear bb1 bb2 cc1 cc2 cc3 dd1 dd2 dd3
        YY (j,2)= prctile(Z{j,2},5); %TTC
            bb1=find(Z{j,3}>0);
        bb2=find(Z{j,3}<0);

        YY (j,3)= prctile(Z{j,3}(bb1),5); %+ve acc
        YY (j,4)= prctile(Z{j,3}(bb2),5); %-ve acc

        cc1=find(Z{j,4}>0);
        cc2=find(Z{j,4}<0);
        cc3=abs(Z{j,4});
```

```
        YY (j,5)= prctile(Z{j,4}(cc1),5); %+ve acc

        YY (j,6)= prctile(Z{j,4}(cc2),5); %-ve acc

        YY (j,7)= prctile(cc3,10); %abs acc


        dd1=find(Z{j,5}>0);

        dd2=find(Z{j,5}<0);

        dd3=abs(Z{j,5});


        YY (j,8)= prctile(Z{j,5}(dd1),5); %+ve acc

        YY (j,9)= prctile(Z{j,5}(dd2),5); %-ve acc

        YY (j,10)= prctile(dd3,5); %abs acc


    end
```

Appendix E

Publications related to this thesis

PAPAZIKOU, E, QUDDUS, M, THOMAS, P, and KIDD, D., 2019. What came before the crash? An investigation through SHRP2 NDS data. *Safety Science*, (*in Press*). DOI: 10.1016/j.ssci.2019.03.010.

PAPAZIKOU, E, QUDDUS, M, THOMAS, P, KIDD, D., 2018. What came before crash? An investigation through SHRP2 NDS data. In *Transportation Research Board 97th Annual Meeting*, Washington DC, US, January 2018.

PAPAZIKOU, E, QUDDUS, M, THOMAS, P, and KIDD, D., 2017. Deriving indicators for understanding the departure from normal driving using SHRP2 NDS data. In *6th International Naturalistic Driving Research Symposium,* The Hague, Netherlands, June 2017.

PAPAZIKOU, E, QUDDUS, M, THOMAS, P., 2017. Detecting Deviation from Normal Driving Using SHRP2 NDS Data. In *Transportation Research Board 2017*, Washington, US, January 2017.