

Review

A Review of Surrogate Safety Measures Uses in Historical Crash Investigations

Dimitrios Nikolaou , Apostolos Ziakopoulos  and George Yannis 

Department of Transportation Planning and Engineering, National Technical University of Athens, 5 Heroon Polytechniou Str., GR-15773 Athens, Greece; apziak@central.ntua.gr (A.Z.); geyannis@central.ntua.gr (G.Y.)

* Correspondence: dnikolaou@mail.ntua.gr; Tel.: +30-210-772-1155

Abstract: Historical road crash data are the main indicator for measuring road safety outcomes. Over the past few decades, significant efforts have been made in obtaining and exploiting Surrogate Safety Measures (SSMs). SSMs have the potential to provide excellent sustainable road safety indicators and proxy measurements which can complement traditional historical crash analyses or even substitute them. By using SSMs, crash data collection demands can be bypassed and areas can be investigated before crashes occur. Due to such advantages, the objective of the present research is to provide a review of the scientific literature regarding studies exploiting SSMs for historical crash record investigations. Specifically, 34 studies were examined, providing insights on the different types of SSMs collected under real road environment conditions, the way they are collected, their connection with specific road crash types, and the type of the developed statistical models are examined and discussed. Particular focus is also placed on the temporal dimension of the collection period of both SSMs and road crashes. Finally, the overall trends deriving from the reviewed studies are summarized and future research directions are provided.

Keywords: road safety; surrogate safety measures; road crashes; study characteristics; data collection periods



Citation: Nikolaou, D.; Ziakopoulos, A.; Yannis, G. A Review of Surrogate Safety Measures Uses in Historical Crash Investigations. *Sustainability* **2023**, *15*, 7580. <https://doi.org/10.3390/su15097580>

Academic Editor: Juneyoung Park

Received: 31 March 2023

Revised: 25 April 2023

Accepted: 3 May 2023

Published: 5 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Road crashes and their related casualties constitute a major societal and public health problem as it is estimated that more than 1.35 million people are killed in road crashes and tens of millions are seriously injured annually [1]. Improving road safety is also included as a key component of the United Nations' Agenda, as manifested by Sustainable Development Goals (SDGs) 3.6 and 11.2, which aim to reduce road fatalities and injuries by half and provide sustainable and safe transport for road users of all age groups respectively [2]. Until now, the main indicator for measuring road safety outcomes has been historical crash data, considered to be hard evidence for the measurements of road safety performance. Even if it is natural to rely on road crash historical records for the assessment of the road safety level of an examined area or road, specific drawbacks of road safety analyses based on historical crash records have been determined as well.

In particular, a long period of time is typically required to collect a sufficient sample of road crash data that could allow for reliable estimates of the road safety level as road crashes are rare events by nature [3]. When examining large geographical areas, road crashes also face the typical issues inherent in all point data such as spatial dependence and spatial heterogeneity [4]. Moreover, any before-and-after study based on historical crash records for the evaluation of the implementation of a road safety measure may be biased by the regression-to-the-mean phenomena [5]. In addition, significant discrepancies are found between the non-fatal road crash injury data provided by various data sources. This problem is known as under-reporting and several studies indicate that the Police Departments do not report an appreciable proportion of road crash injuries, whereas the

extent of under-reporting may vary depending on the severity of the injuries or the road user types [6,7]. Apart from the aforementioned, it can be perceived that road safety analyses based on historical crash records are a reactive approach that forces road safety analysts to wait for road crashes to occur in order to examine measures that could prevent them and should rely on valid crash data, including accurate location data, which is not always the case [8].

Therefore, over the past few years, significant efforts have been made in utilizing Surrogate Safety Measures (SSMs) in order to address this issue [9]. SSMs include all measures, parameters, or quantities, which do not stem directly from or rely on crash data. Such approaches are a sustainable way of gauging road safety and may be more preferable as they allow for road safety analyses before the physical occurrence of road crashes. According to Tarko [10], the use of SSMs in the field of road safety aids in the detection of road crashes' excessive risk, the knowledge improvement of crash-leading conditions, and the effectiveness estimation of various countermeasures. Wang et al. [9] provide a comprehensive review of important SSMs and divide them into two key categories: (i) SSMs and (ii) SSM-based models. The first category includes key time-based, deceleration-based, and energy-based SSMs. These subcategories include predominant SSMs that use predefined thresholds for traffic conflicts' identification and are used widely across studies in the road safety literature such as Time-to-Collision (TTC), Post Encroachment Time (PET), Time-to-Crash/Accident (TC/TA) and Deceleration Rate to Avoid the Crash (DRAC) [11]. On the other hand, the second category aims to directly associate each traffic conflict with either a crash or non-crash outcome, by estimating its crash probability [12,13].

Initially, data collection of SSMs was based on roadside observation techniques [14]. As it can be intuitively perceived, such approaches were not accurate as they were based on subjective criteria [15]. In order to reduce such biases, video-based measurements were introduced many years ago [16] and have been improving significantly since then. Recent, technological advancements have led to more advanced techniques that reduce human interventions and deploy computer vision and sensor techniques [17–19]. Moreover, several simulation-based analyses have been conducted aiming to derive SSMs from traffic simulation models [20,21]. The rapid technological development in naturalistic driver recording has also brought about an increasing availability of data from sensors in vehicles and smartphones that can be used to extract various SSMs such as TTC, harsh braking events, and harsh acceleration events [22–24]. All in all, SSMs can either be an alternative to road safety analyses or even complement analyses that are based on historical crash records [25].

Within this framework, the aim of the present paper is to provide a review of the scientific literature of studies exploiting SSMs in historical crash record investigations. More specifically, this paper focuses on studies that attempt either (i) to investigate the correlation of SSMs and historical crash records or (ii) to predict the number of expected road crashes through SSMs and then compare them with the historical crash records. The different types of SSMs, the manner in which they are collected, their connection with specific road crash types, and the type of the developed statistical models are examined and discussed. Particular emphasis is placed on the temporal periods dedicated to data collection for both the SSMs and road crash data, as uncertainties in the length of the data collection periods are a problem typically investigated in driver recording [26]. In order to achieve this aim, published scientific studies that are authored in English are critically examined. It should be mentioned that this study only includes relevant papers that concern SSMs collected under real road environment conditions, as opposed to studies that are based on traffic simulation and driver simulators.

During the review process, studies dealing with the use of traffic conflict techniques for use in-road safety assessments were also identified. Arun et al. [27] focused on mapping the concepts and methods related to surrogate safety assessment using traffic conflicts. Their study deals with specific topics such as the concept of crash surrogacy, the definition and identification of traffic conflicts, and the specification of the relationship between

crashes and conflicts. In other studies, Arun et al. [28] assessed the different traffic conflict safety thresholds among various road environments and applications, while Zheng et al. [29] discussed various conceptual and methodological issues related to traffic conflict modeling. However, the current study presents novelty in different areas. Specifically, it (i) exclusively investigates studies that use both SSMs and historical crash records, (ii) extends beyond measures with predefined thresholds for traffic conflicts' identification to SSMs that can be extracted from smartphone sensors and instrumented vehicles related to harsh driving behavior events, and (iii) sheds light on the temporal periods dedicated to data collection for both SSMs and crashes.

Following this Introduction, the paper is organized as follows. Section 2 describes the methodological framework of the current review paper, including the Preferred Reporting Items for Systematic Reviews and the Meta-Analyses (PRISMA) approach that was adopted. Section 3 showcases the main review findings in terms of the different types of SSMs and crashes, various modeling approaches, and the temporal dimension of the data used in the examined studies. Subsequently, a discussion of overall findings and trends from the reviewed studies and some future research directions are provided in Section 4, while Section 5 includes the main concluding remarks of the current research.

2. Review Methodology

The current review was carried out during June 2022 and adhered to the PRISMA guidelines [30]. The search was undertaken in the Scopus, TRID and Web of Science databases; Figure 1 depicts the search terms and the study selection process. It should be noted that there was no specific search restriction on the publication date of the examined articles. Moreover, articles had to be peer-reviewed before publication and authored in English which is the predominant written language in the global scientific literature. Emphasis should be placed on the fact that the present paper aims to provide a review of the scientific literature regarding studies exploiting SSMs towards historical crash record investigations and thus includes only studies that were conducted under real road environment conditions (as opposed to simulators).

After the exclusion of some papers based on their titles and abstracts, a total of 52 articles were selected for full-text review. After the full-paper review, 18 studies were excluded for not meeting the inclusion criteria (e.g., absence of historical crash data or SSMs, separate statistical models for SSMs and road crashes, crash data available but not used in statistical modeling, etc.) Finally, 34 articles were identified and reviewed. The literature review findings are presented and discussed in detail in the following sections of the present paper.

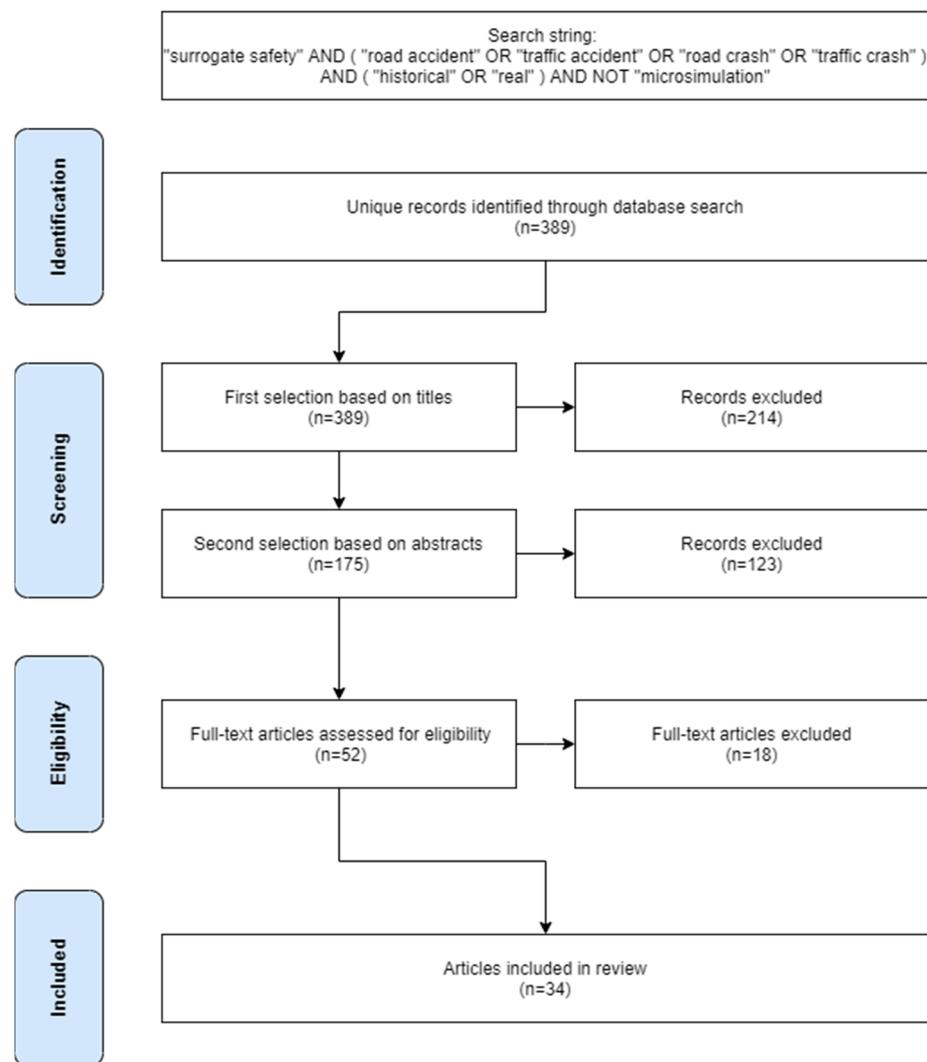


Figure 1. PRISMA flow diagram.

3. Review Findings

3.1. Types of SSMs and Historical Crash Data

As already pointed out in the introductory section of this paper, SSMs can be leveraged in road safety analyses in two ways. On one hand, they can provide an alternative to road safety analyses when road crash data are not available as a proactive approach. On the other hand, SSMs complement analyses based on historical crash records, which is also the main subject of the current review paper. The key information about the SSMs and historical crash records (types and temporal dimension), modeling approaches, the scale of analysis, and other considered variables used in the reviewed studies are summarized in Table A1 of Appendix A, sorted by means of collection for SSMs. It should be noted that the column named “Temporal Ratio” of Table A1 has been calculated due to the observed discrepancies in data collection period lengths for crashes and SSMs. The values of this column are dimensionless numbers as they have been calculated by converting the crash and SSMs data collection periods into the same time units.

Technological improvements during recent decades have led to the development of a wide array of sophisticated tools that provide more rich and rapid data acquisition in terms of various aspects of driving performance [31]. As can be observed from Table A1, during the last five years, the use of smartphone data has also begun to gain significant ground in studies featuring SSMs [32–39]. Exploiting smartphone sensors such as accelerometers, digital compasses, gyroscopes, and GPS allows the extraction of various driver performance

metrics and SSMs through an inexpensive and rapid way, even without requiring user engagement [40].

The SSMs collected via smartphone sensors in the examined studies concern harsh driving behavior events such as harsh braking and harsh acceleration. Harsh braking events are generated by drivers as a reaction to various possibly dangerous situations in order to avoid a near miss or even a road crash [24]. Moreover, harsh braking events are a critical element for the assessment of driving risk [41], as they are innately associated with crash occurrence probability [42]. However, harsh acceleration events are different phenomena than harsh braking events, as they are mainly affected by drivers' levels of anger, frustration, and anxiety [43]. Based on previous studies, it is noted that the levels of deceleration and acceleration that define harsh braking and harsh acceleration events respectively may vary across different studies and transport modes [44,45]. A frequent barrier encountered in studies exploiting harsh events is that they do not provide their specific thresholds and calculation methods for commercial reasons [39,46,47].

As can be observed from Table A1, naturalistic driving experiments using instrumented vehicles are another frequently selected option for collecting SSMs. These experiments are a quite similar alternative to smartphone data but much more expensive as there are significant costs that depend on the equipment used [48] and the duration of the experiment [49]. The majority of the SSMs collected through instrumented vehicles range in a similar concept to the data collected by smartphones and concern harsh driving behavior events [44,45,50–58]. Apart from these studies that focus on harsh driving behavior events, traffic conflicts and related measures for rating their severity have also been examined in other naturalistic driving experiments using instrumented vehicles [59,60].

The term traffic conflict denotes an observable event that would end in a road crash unless one of the involved road users slows down, changes lane, or accelerates to avoid a collision [61]. Based on Table A1, it is demonstrated that the collection of traffic conflict-related SSMs under real road conditions in the majority of the examined studies is based on video recordings [62–68]. Conflict surveys through field observations are another option for collecting such data [69]. When real vehicle trajectories and speeds are not available, simulation models are widely used [20,70]. However, simulation studies fall outside the scope of the presented research and are not discussed further.

Among the different traffic conflict-related SSMs used in the reviewed studies, it can be observed that PET, TTC, and DRAC are the most widely used. According to Gettman and Head [20], PET is defined as the time elapsed between the encroachment's end of the turning vehicle and the time that the trough vehicle reaches the potential point of the crash, while TTC corresponds to the expected time for two vehicles to collide if they maintain their present speed and path. Various modifications of the TTC have been used in the examined studies such as the minimum TTC (mTTC) [64,66], which corresponds to the TTC's lowest values obtained, and the modified TTC (MTTC) proposed by Ozbay et al. [71] that takes into account relative position, relative speed and relative acceleration of the conflicting vehicles [63,68]. Lastly, DRAC corresponds to the minimum deceleration rate required by the following vehicle to come to a timely stop (or match the leading vehicle's speed) and hence to avoid a crash [72]. However, a frequent issue encountered in such studies and also identified by a relevant study is that the safety thresholds of conflicts vary by traffic environment type and the application purposes of conflict measures [27].

According to Lu et al. [73], connected vehicles are the key to the evolution of next-generation intelligent transportation systems. In addition, they are expected to bring multiple benefits to driving behavior monitoring tools as well [31]. Table A1 reveals that, when utilized, connected vehicles are an additional emerging option for studies exploiting SSMs for historical crash record investigations and can be a standardized, streamlined, and seamless collection source of both harsh event and traffic conflict data [74–76].

Regardless of how SSMs are collected, in most of the studies reviewed, the type of historical road safety data used is either the absolute number of total crashes or the number of total road crashes divided by a risk exposure indicator such as the number of vehicles

or vehicle kilometers traveled [53,58,59]. Furthermore, the severity of road crashes is not taken into account in the majority of the studies included in Table A1. However, there are certain studies that focus on serious or fatal road crashes [62,66]. Several studies attempt to correlate SSMs with specific road crash types such as rear-end, angle and single-vehicle crashes [52,56,64,76]. Other research studies focus on specific road crash characteristics such as the weather conditions, and the time or the day of the crash, which usually correspond to the conditions of SSM collection [33,68]. Moreover, the historical crash records of some other studies target specific road user types such as vulnerable road users [32,57,66] and drivers of various transport modes [38].

Lastly, in addition to the SSMs and historical crash data, most of the examined studies in Table A1 include some supplementary variables that are mainly related to road infrastructure and traffic. Among these variables, road length and road class prevail for infrastructure, while traffic volume and speed prevail for traffic parameters.

3.2. Modelling Approaches

This section of the paper gives a brief overview of the various modeling approaches implemented in the reviewed studies that are presented in Table A1 and exploit SSMs for historical crash records. Initially, it can be observed that some studies are only limited to different correlation methods, such as Pearson or Spearman correlation, which aim to measure the strength of association between SSMs and road crashes [32,35,50,74]. Certainly, correlation matrices are also included in other studies as a preliminary step before the development of more advanced statistical models [38,54,56,57].

Generalized Linear Models (GLMs) have been implemented widely in the road safety literature for many years, as they assume that crashes are independent, random, and sporadic countable events [77]. Based on Table A1, it is observed that Poisson [56,65,78] and Negative Binomial (NB) models [38,45,52,53,60,66,69] are the most common forms of GLMs among studies exploiting SSMs for historical crash record investigations, with NB models being more prevalent than Poisson models. The key difference between these two GLM forms has to do with the fact that NB models relax the equal mean and variance assumption of the Poisson model, which can account for overdispersion resulting from unobserved heterogeneity and temporal dependency [79]. Specific research documents among the reviewed studies have also introduced random effects to GLMs in order to extend them to Generalized Linear Mixed Models (GLMMs) and account for unobserved heterogeneity [44,51].

Several of the reviewed studies have also attempted to incorporate into their analyses the effects of various road safety indicators' spatial characteristics. Bayesian approaches are widely used to consider the spatial correlation for modeling crash frequencies. In that context, Li et al. [57] developed a Bayesian NB model with conditional autoregression (CAR) prior to accounting for spatial correlation between neighboring bus stops. The results of this research indicated the necessity of considering spatial autocorrelation during the crash frequency model process as the developed Bayesian NB-CAR model outperformed the Bayesian model in terms of various model evaluation metrics. In another study, both the spatial and temporal dependence of crash observation were taken into account in a multivariate conditional autoregressive (MVCAR) model in the full Bayesian framework [37].

Yang et al. [76] proposed a new safety measure termed Risk Status, which was modeled as a latent variable in a Structural Equation Model (SEM) in the Bayesian framework that could account for both spatial autocorrelation through CAR spatial effect and unobserved heterogeneity through road segments random parameters (i.e., SEM-CAR-RP). Overall, SEM is a powerful multivariate tool for jointly modeling interrelationships among observed and latent variables [80]. However, the proposed approach of SEM-CAR-RP extends the methodological frontier of SEM applications in the field of road safety as it was found to be superior compared to more traditional alternatives of SEMs that did not take into account CAR spatial effect and unobserved heterogeneity. This finding demonstrates that various fundamental methodological issues of crash data modeling such as spatial autocorrelation,

unobserved heterogeneity, etc. need to be investigated when exploring data from new data sources similar to those that were presented in Section 3.1. Paleti et al. [33] developed a random parameter Generalized Ordered Response Probit (GORP) model which is a type of model that can easily handle over or under-representation of multiple count outcomes at the same time without demanding a hurdle or zero-inflated model. The outcomes of this research revealed that the best-performing model was one including measurement error, random parameter heterogeneity, and spatial dependency.

In a more straightforward approach, Li et al. [58] utilized a line-constrained clustering method that combines DBSCAN with spatial selection functions in order to identify individual-specific risky road segments. Latent Gaussian Models (LGMs) are a subcategory of structure additive models, in which the dependent variable for each subject follows a distribution from the exponential family and can introduce temporal or spatial dependence [81]. This spatial modeling approach using the Integrated Nested Laplace Approximation (INLA) technique has been chosen as an appropriate tool for road network screening [34,36,54]. The INLA approach was introduced by Rue et al. [82] as a computationally efficient alternative to Markov chain Monte Carlo methods. INLA can be combined with the Stochastic Partial Differential Equation (SPDE) approach proposed by Lindgren et al. [83] in order to implement spatial and spatio-temporal models for point-reference data [84].

Extreme Value Theory (EVT) is a statistical approach that enables extrapolation from observed levels to unobserved levels [85], which is in alignment with the goal of predicting less frequent road crashes from more frequent traffic conflicts. EVT Models are becoming increasingly popular with substantial developments achieved recently. These models are mainly used to estimate the number of road crashes and then compare them to the observed historical crash records. Among studies presented in Table A1, bivariate EVT models have been proposed and it was found that this approach generated more accurate crash estimates than univariate models [63,64]. In a more recent study, Fu and Sayed [67] developed a Bayesian hierarchical extreme value model, which had three layers: the data layer, the process layer, and the prior layer. However, as also mentioned for different other model types and highlighted by Zheng et al. [29], one important issue while developing such models is accounting for the unobserved heterogeneity across different observation locations. In order to deal with the issue, Fu and Sayed [68] propose a random parameters Bayesian hierarchical extreme value modeling approach.

As can be observed in Table A1, traditional modeling approaches such as linear or logistic regression models have been used in a few studies exploiting SSMS for historical crash record investigations, but are less preferred [39,55,59,62]. This is partly also due to the emergence of Machine Learning (ML) and Deep Learning (DL) approaches as powerful tools that are gaining more ground for road safety analyses due to their ability to handle large volumes of data, their heightened predictive capabilities, and the complex, non-linear relationships they can disclose. Indicatively, the random forest algorithm is a data-mining tool that has been used to determine the importance of the variables and includes in the statistical models the variables with the strongest impacts on road crashes [39,45]. Furthermore, Hu et al. [75] exploited SSMS derived from connected vehicles' data such as harsh braking, harsh acceleration, and wait time in order to predict the crash risk at intersections using DL approaches. Their analyses revealed that the performance of two black-box DL models, Multi-Layer Perceptron (MLP) and convolutional neural network (CNN) was slightly better than the Decision Tree Model. However, in the context of the examined studies it can be perceived that ML/DL approaches are not among the most prevalent methods at present.

In summary, various modeling approaches have been implemented in the reviewed studies. However, the selection of an appropriate modeling framework depends highly on the research questions being asked, the available data, and the specific context of each study. Specifically, the type of crash data being analyzed (e.g., count data, rates such as crashes divided by an exposure parameter, categorical/binary data, etc.), the level of spatial and temporal dependence, and the existence of unobserved heterogeneity are

some factors that should be taken into consideration towards the selection of a suitable modeling methodology. While there are many different modeling approaches available in the literature, they should be treated as starting points for road safety practitioners, rather than definitive guides.

3.3. Temporal Dimension

When examining Table A1, no clear pattern can be observed with regard to the time periods of historical road crash data and SSMs collection. This is a constant topic, and researchers have to anticipate and plan accordingly in the study design process. Therefore, in this section, the authors attempt to shed light on this issue and identify potential hidden patterns through the visualization of the respective data in Table A1. As already mentioned in previous parts of the current research, there are different ways that can be used to extract SSMs. It is observed that in studies using smartphones, instrumented vehicles, or connected vehicles the time period for which the SSMs were collected can vary from a few days [45,51] to several months [33,44,86]. On the other hand, SSMs collected through video recordings or conflict surveys are collected for a few hours [62,64]. As per the aforementioned, this discrepancy was also one of the main incentives for calculating the ‘Temporal Ratio’ column of Table A1. The difference in time periods between the collection of historical road crash data and SSMs is mainly attributed to the emergence of new technologies, which allow for the rapid collection of SSMs data and the conduction of analyses with shorter time periods. The ‘Temporal Ratio’ column could be interpreted as by how much more time is needed to collect an equivalent sample of SSMs with road crash data. For this reason, as well as for readability reasons, two different graphs have been produced. Specifically, Figure 2 demonstrates the time periods of historical road crash data and SSMs collected through smartphones, instrumented vehicles, and connected vehicles, while Figure 3 presents the respective values for the studies that used video records or conflict surveys for the extraction of SSMs.

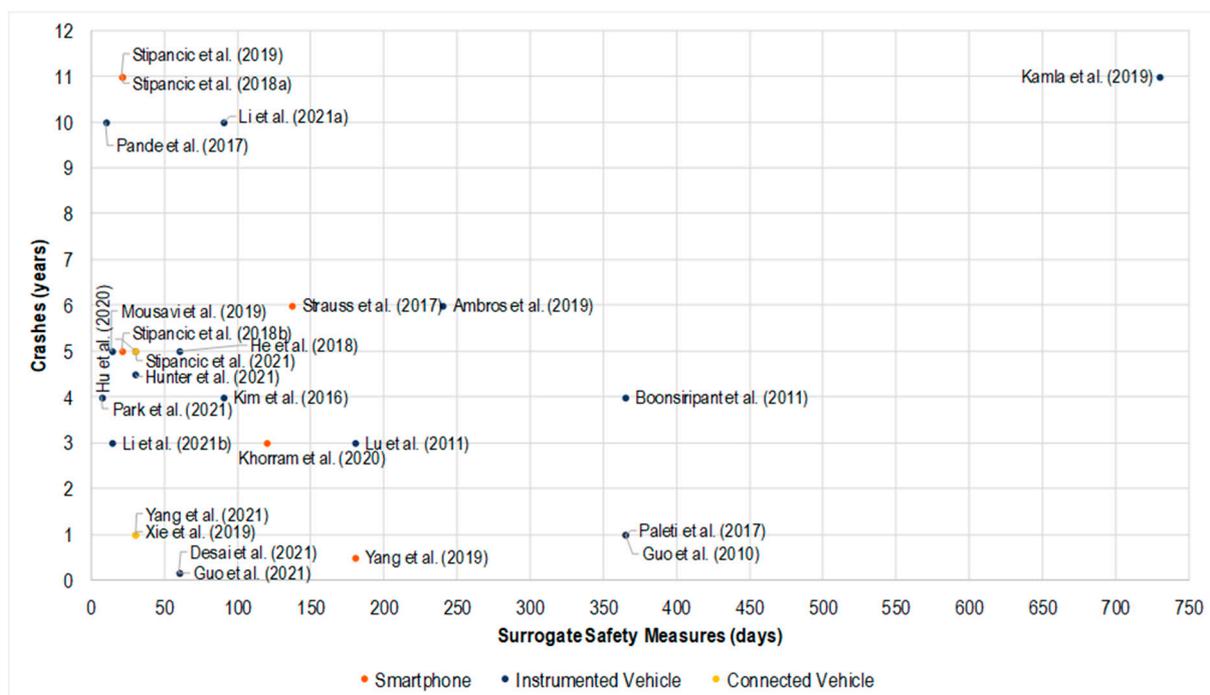


Figure 2. Time periods of historical road crash data and SSMs collected through smartphones, instrumented vehicles, and connected vehicles [32–39,44,45,50–60,74–76,78,86].



Figure 3. Time periods of historical road crash data and SSMs collected through video records and conflict surveys [62–69].

Based on Figures 2 and 3, a general trend that can be observed is that among all the examined studies the time period of road crash data is always greater than or equal to the time period of collection of SSMs, as expected from the increased usability that SSMs provide. Furthermore, regardless of the manner in which SSMs are collected, it is observed that in the majority of the examined studies (21 out of 34), historical road crash data used correspond to periods of three to six years.

Only five of the examined studies, use exactly the same time periods of historical crash data and SSMs. These studies exploit smartphones [33,37,39] and instrumented vehicles [55,78] for the extraction of SSMs. It can be observed that they are concentrated in the low spectrum of the Y-axis of Figure 2 as the crash data that they include in their analyses do not exceed one year. The highest ratio of road crash data time period to the time period of SSMs corresponds to the studies presented in the upper left part of Figure 2 [34,36,45,51]. In particular, in these studies, the road crash data time period is calculated to be between 191 and 365 times longer (mean: 239, st.dev: 84.4) than the SSM time periods. The vast majority of the studies presented in Figure 2 are concentrated in the middle level of the Y-axis and towards the left side of the X-axis. In these studies, the time period of road crashes is estimated to be between 12 and 130 times longer (mean: 50, st.dev: 36.3) than that of the SSMs. In addition, there are also some studies located in the central and upper right part of Figure 2 for which the time period of road crashes is 4–9 times longer than that of SSMs (mean: 7, st.dev: 2.3) [38,44,52,59,86].

Lastly, the comparison between Figures 2 and 3 reveals that the ratio of road crash data time period to the time period of SSMs is much higher in the studies that collect SSMs through video records or conflict surveys compared to the other studies. This is due to the fact that the collection of SSMs through video recordings or conflict surveys requires only a few hours and the historical crash records correspond to time periods of at least three years, lending further credence to the utility of SSMs due to their rapid data collection.

4. Discussion

4.1. Overall Findings and Trends from Reviewed Studies

SSMs are steadily gaining ground in the road safety literature as they are a sustainable way of gauging road safety and allow the conduction of analyses without necessarily requiring historical road crash records. Moreover, the rapid and continuous progress in the field of technology makes it increasingly easier to collect such indicators. However, SSMs

can also be combined with data from historical road crash records in order to complement and provide additional information to relevant road safety analyses. The present research focused on studies that exploit real-condition SSMs for historical crash record investigations.

The examination of the studies in the framework of this research has revealed some noteworthy conclusions for road safety analyses that combine SSMs and historical crash data. It appears that the technological development in recent years has significantly contributed to making smartphones a key choice for collecting SSMs [32–39]. The indicators collected through smartphones' sensors can be quite similar to those collected by instrumented vehicles [52,54,56]. However, the cost of collecting SSMs via smartphones is significantly lower compared to that of instrumented vehicles. A fact that is also reflected by the increase in the use of smartphones in the relevant studies during the last five years.

The majority of SSMs collected through either smartphones or instrumented vehicles involve harsh driving behavior events. Through these studies, it becomes clear that the most commonly exploited harsh driving behavior events such as harsh braking and harsh acceleration events are positively correlated with various types of road crash counts [34,36,37,51–53,55–57] and road crash risk [39]. As this relationship is verified by several studies, it can be deduced that harsh events could be used as dependent variables in statistical models as a proactive approach that does not require the collection of historical road crash data. Another approach used to collect SSMs is based on traffic conflicts. As for real road conditions, the collection of relevant indicators is mainly carried out through the analyses of video recordings [62–68]. As with the SSMs collected through smartphones or instrumented vehicles, the reviewed studies based on traffic conflict indicators aimed either to investigate the relationship between the produced SSMs and historical crash counts or to predict the number of road crashes and then compare it with the observed crash counts.

Regarding the type of statistical analyses used in studies that combine SSMs and historical road crash data, GLMs including their various modifications dominate. There are also several studies that choose more specialized approaches to take into account unobserved heterogeneity and spatial dependence as they are among the most prevalent methodological issues typically faced when dealing with crash data modeling. Another common approach chosen by the reviewed studies concerns the different variants of EVT. Finally, it can be observed that ML techniques are not often used in the reviewed studies. Overall, the research questions, data type, and specific contextual factors of each study are critical to the choice of the respectively developed modeling framework.

Finally, a key finding of the present research that could be also highlighted as its most significant contribution relates to the time periods for which both the historical road crash data and the SSMs are collected. Until recently, it was not clear if there was any particular pattern. This research sheds light on this topic by revealing that in most studies that collect SSMs via smartphones and instrumented or connected vehicles, road crash data correspond on average to time periods that are 50 times longer than the collection periods of the SSMs. In cases of collection of the alternative indicators through video recordings, the time period of crash data is significantly higher than the respective period of collection of SSMs.

4.2. Future Research Directions

This section outlines research directions that do not appear to be sufficiently investigated from the present literature of studies exploiting SSMs for historical crash record investigations and can form meaningful upcoming research endeavors. An important aspect of road safety analyses is the level of injury severity of road crashes. However, it is observed that in the majority of the studies, severity has not been adequately investigated as they mainly exploit the total number of all injury road crashes without taking into account the different severity levels. However, there are a small number of studies that focus on serious or fatal road crashes [62,65]. The inclusion of the level of injury severity in similar studies would be highly interesting for the quantification and the comparative assessment of the relationship between SSMs and different crash severity levels. Injury severity estimation using SSMs is also highlighted as a critical research need by Arun et al. [28]. In that

direction, a few recent research studies have attempted to estimate crashes by severity level using different SSMs [87–89].

Furthermore, most of the reviewed studies focus on road crashes involving all road users without separating them. However, there are some specific types of road users such as pedestrians, pedal cyclists, and motorcyclists that are considered vulnerable road users (VRUs), as they are prone to injury in any vehicular collision, primarily because there is little or no external protective device that could absorb the impact of a road crash [90]. It is estimated that VRUs account for more than half of all road fatalities globally [1]. Moreover, the noteworthy increase in the use of new micromobility transport modes such as e-scooters in many cities around the globe has raised particular concerns for the safety of these emerging types of VRUs [91]. Therefore, more research is needed on the manner in which various SSMs could be exploited to enhance the safety of VRUs. Towards this direction, Ali et al. developed a Bayesian Generalized EVT model in order to estimate real-time pedestrian crash risks at signalized intersections using Artificial Intelligence (AI)-based video analytics [92].

Regarding the spatial scale of the analyses, it appears that the examined studies focus on the microscopic level as they mainly investigate road segments and intersections. Another promising research direction would be the application of analyses at a more macroscopic level such as regional areas (cities, metropolitan areas, local administrative units, etc.). In such cases, apart from different SSMs and road crash rates, various demographic, socioeconomic, and traffic exposure factors of the examined areas could be taken into consideration in the analyses. However, it is important to note that as the size of the examined area increases, capturing unobserved heterogeneity becomes more challenging [93]. Apart from demographic and socioeconomic factors, key road safety performance indicators reflecting the safety of road users (seatbelt and helmet use, speeding, driving under the influence of alcohol, distraction), infrastructure, vehicles, and post-crash response in the examined regional areas could be also taken into account.

Over the last years, ML models have been proven to be very efficient prediction tools, making them also particularly popular in road safety analyses. ML and DL approaches have come to challenge the hitherto dominance of traditional modeling approaches by being implemented alongside or instead of them. Based on the results of the present literature review research, it appears that these approaches have not found frequent application in studies that exploit SSMs for historical crash record investigations. This could be attributed to the major challenge of interpreting the results generated by the respective algorithms accurately. However, this issue could be tackled by using model agnostic methods such as the SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) that would explain the interpretation of the model regardless of the model type. Furthermore, hybrid modeling approaches integrating both statistical and ML techniques could be considered in future research studies, as this framework represents a methodological advancement in traffic conflict-based crash estimation models [94].

Lastly, the aforementioned future research directions can all be further augmented by the constant improvements in the technological field such as the further exploitation of smartphone data that can provide a vast amount of driving big data under real road conditions and connected vehicles that can be used for a more connected traffic environment. The rollout of fifth-generation networks (5G) provides a unique opportunity for creating and exploiting innovative solutions to improve communication between all transport system components and reduce road crash casualties. The application of 5G in traffic environments could be a game changer over the next years as it enhances direct communication capabilities with very low latency such as Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I) and Vehicle-to-Everything (V2X) [95]. This framework could assist in the collection of a wealth of real-time data that can be also used for the extraction of various SSMs that could be integrated into traditional road safety analysis.

5. Conclusions

The main indicator used both to identify road safety problems and to evaluate the effectiveness of interventions is historical road crash data. However, a typically long time period, measuring in spans of years, is required to gather an adequate sample of road crash data that could lead to reliable road safety assessments. As a consequence, noteworthy efforts have been made in addressing this issue by creating sustainable metrics for assessing road safety such as SSMs, which are not based on historical road crash records. In the road safety literature, SSMs are either an alternative to road safety analyses or complement analyses based on historical crash records.

This paper has provided a review of the current literature on studies that exploit SSMs for historical crash record investigations. Particular focus has been placed on the different types of SSMs that are correlated with road crashes, their means of collection, the different modeling approaches used in the reviewed studies, and the temporal dimension of the collection period of both SSMs and road crashes. It was determined that constant technological advancements have highlighted smartphones as a rapidly emerging option for collecting SSMs under real road conditions. Moreover, a noteworthy novel conclusion of the current paper is that in the majority of studies that collect SSMs via smartphones and instrumented or connected vehicles, road crash data correspond on average to time periods that are 50 times longer than the collection periods of the SSMs. In all cases, the time period of road crash data is always greater than or equal to the respective period of SSM collection. Regarding the modeling approaches followed in the examined studies, it is highlighted that different fundamental methodological issues of crash data modeling such as spatial autocorrelation and unobserved heterogeneity should be taken into account in the selected statistical models.

For further research, there are some factors that need to be examined further such as the relationship between SSMs and road crashes with different levels of injury severity and different types of road users. It is also argued that the rapid improvements in the field of technology can further assist in the collection of a wealth of driving behavior data and related SSMs through the exploitation of smartphones and connected vehicles. Finally, other research directions have also been provided, such as the implementation of interpretable ML algorithms for investigating the relationship between SSMs and road crashes.

Author Contributions: Conceptualization, D.N.; methodology, D.N.; writing—original draft preparation, D.N. and A.Z.; writing—review and editing, D.N. and A.Z.; supervision, G.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Studies exploiting SSMs in historical crash record investigations.

Reference	Surrogate Safety Measures				Other Variables			Historical Crash Data		Temporal Ratio (Crashes Period/ SSMs Period)	Modelling Approach	Scale of Analysis
	Type	Sample	Collection	Period	Infrastructure	Traffic	Other	Period	Type			
Khorram et al. [38]	harsh braking	176 bus drivers	smartphone	4 months	length	deceleration	driver age and experience	3 years	Bus driver at-fault	9	Pearson correlation, GLM (NB)	2 routes (13 km, 10 km)
Paleti et al. [33]	harsh braking, harsh acceleration	11 drivers, 228 trips, 58 h of driving (4–6 pm)	smartphone	1 year	interchange, surface	traffic volume, mean speed, SD acceleration	-	1 year	4–6 pm weekdays	1	random parameters Generalized Ordered Response Probit (GORP)	513 freeway segments
Stipančić et al. [34]	harsh braking	~22,000 trips, >4000 drivers	smartphone	21 days	length, class	congestion, mean speed, speed variation	-	11 years	Total	191	INLA Full Bayesian Latent Gaussian Model	1000 links and intersections
Stipančić et al. [35]	harsh braking, harsh acceleration	~22,000 trips, >4000 drivers	smartphone	21 days	class	-	-	5 years	Total	87	Spearman correlation and pairwise Kolmogorov-Smirnov test	20,586 links and 10,721 intersections
Stipančić et al. [36]	harsh braking	~22,000 trips, >4000 drivers	smartphone	21 days	length, class	congestion, mean speed, speed variation	-	11 years	Total	191	INLA Full Bayesian Latent Gaussian Model, Fractional Multinomial Logit	4623 links and 4429 intersections
Strauss et al. [32]	harsh braking	over 10,000 trips, ~1000 cyclists	smartphone	137 days	-	traffic volume	-	6 years	Cyclists	16	empirical Bayes (EB) estimates—Spearman correlation	13,279 intersections and 19,837 segments (aggregated also at corridors level)
Yang et al. [37]	harsh braking, harsh acceleration	10,512 events	smartphone	6 months	bus and subway stations, intersections, length	traffic volume, truck flow, speeding	distraction, land use, population, unemployment, income, housing, commuting	6 months	Total	1	MVCAR, UCAR, two-sample Kolmogorov-Smirnov test, Wilcoxon signed-rank test	282 census tracts
Guo et al. [39]	Harsh: braking, acceleration, turn, merge into lane	-	in-vehicle navigation software	2 months	-	traffic volume, congestion, mean speed, speed variation	-	2 months	Total	1	Random Forest, Logistic regression	40 freeway segments
Ambros et al. [52]	harsh braking, harsh acceleration	1172 company vehicles	instrumented vehicle	8 months	curve length and radius	traffic volume, acceleration	-	6 years	Single-vehicle	9	GLM (NB)	30 rural curves
Boonsiripant et al. [86]	stop frequency, variation of stops, 90th percentile count of stops	36,724 trips, 408 drivers	instrumented vehicle	1 year	speed limits	traffic volume, speed variation, V85, V95, V5, acceleration	-	4 years	Daytime, clear weather, motor vehicle	4	Regression tree and GLM	61 urban corridors
Desai et al. [55]	harsh braking	196,215 events	instrumented vehicle	2 months	length	-	-	2 months	Injury and PDO	1	Linear regression	23 construction work zones (150 miles)

Table A1. Cont.

Reference	Surrogate Safety Measures				Other Variables			Historical Crash Data		Temporal Ratio (Crashes Period/ SSMs Period)	Modelling Approach	Scale of Analysis
	Type	Sample	Collection	Period	Infrastructure	Traffic	Other	Period	Type			
Guo et al. [78]	near crash	100 cars, 2 million veh-miles, 43,000 h	instrumented vehicle	1 year	-	-	-	1 year	Total	1	GLM (Poisson)	Northern Virginia/Metro Washington, DC
He et al. [60]	TTC, MTTC, DRAC, brake duration	100 vehicles	instrumented vehicle	2 months	length	mean speed	mean trip duration, extreme trip index	5 years	Rear-end mid-block	30	GLM (NB)	2772 links
Hunter et al. [56]	harsh braking	10,000 events	instrumented vehicle	1 months	-	traffic volume	-	4.5 years	Rear-end	55	Spearman, Pearson and Kendall Cor., Sensitivity Analysis, GLM (Poisson)	8 intersections
Kamla et al. [44]	harsh braking	8000 trucks, 195,297 harsh braking events	instrumented vehicle	2 years	width, inscribed circle diameter	traffic volume, truck traffic	-	11 years	Total	6	GLM (NB) random/fixed-parameters	70 roundabouts
Kim et al. [50]	harsh braking	20 vehicles, 150 k seconds of data, 224 trips	instrumented vehicle	3 months	internal TMC, recurrent bottleneck	speed, acceleration, deceleration	-	4 years	Rear-end /veh-km	16	Correlation, Spatial distribution using GIS	60 segments (63-mile freeway)
Li et al. [57]	harsh braking, harsh acceleration	300 buses, 6.7 million GPS records	instrumented vehicle	3 months	-	-	number of buses	10 years	Pedestrian and bicycle	41	Spearman correlation, Bayesian NB, Bayesian NB-CAR	200 m and 100 m buffer circles
Li et al. [58]	harsh braking	16 participants	instrumented vehicle	2 weeks	length	traffic volume	-	3 years	Total/veh-miles	78	Line-constrained clustering method (combines DBSCAN with spatial selection functions)	156 quarter mile segments of two highways
Lu et al. [59]	conflicts/vehicles	50 taxies, 2.25 million km traveled	instrumented vehicle	6 months	-	-	-	3 years	Total/vehicles	6	Linear regression	city, country
Mousavi et al. [53]	harsh braking	31 participants	instrumented vehicle	2 weeks	curvature	traffic volume	-	5 years	Total/traffic volume	130	GLM (NB)	31 + 21 quarter mile segments of two highways
Pande et al. [51]	harsh braking	33 drivers	instrumented vehicle	10 days	curve (y/n), auxiliary lane (y/n)	traffic volume	-	10 years	Total	365	GLM (NB) random/fixed-parameters	39 freeway segments
Park et al. [45]	Harsh: acceleration, braking, start, stop, lane change, overtaking, turning, U-turn	all commercial vehicles in Korea	instrumented vehicle	1 week	length	speeding	city	4 years	Total	209	Random Forest, GLM (NB)	38 segments in 4 cities
Stipancic et al. [54]	harsh braking	~1.5 million trips	instrumented vehicle	30 days	length, class	congestion, mean speed, speed variation	-	5–11 years	Total	61	INLA Full Bayesian Latent Gaussian Model	123,792 links
Hu et al. [75]	harsh braking, harsh acceleration, wait-time	90 vehicles	connected vehicle	1 month	approaches, traffic light	-	traffic volume, speed, acceleration, deceleration	5 years	Total	61	Multi-layer perceptron (MLP), Convolutional Neural Network (CNN), Decision Tree	774 intersections

Table A1. Cont.

Reference	Surrogate Safety Measures				Other Variables			Historical Crash Data		Temporal Ratio (Crashes Period/ SSMs Period)	Modelling Approach	Scale of Analysis
	Type	Sample	Collection	Period	Infrastructure	Traffic	Other	Period	Type			
Xie et al. [74]	TTC, DRAC, TTCD	90 vehicles, 15.7 million GPS points	connected vehicle	1 month	-	traffic volume	-	1 year	Rear-end/traffic volume	12	Pearson correlation	75 highway segments
Yang et al. [76]	TTC, DRAC, TTCD	2.7 million trajectory points	connected vehicle	1 month	class, speed limit, lanes	traffic volume	GPS points	1 year	Rear-end	12	SEM-CAR-RP	220 road segments
Alhajyaseen [62]	kinetic energy, PET	-	video records	3 h	-	-	-	6 years	Severe	17,520	Sensitivity Analysis, Exponential Relationships	5 urban intersections
Fu and Sayed [67]	DRAC	2202 events	video records	15 h	-	-	-	3 years	Rear-end, daytime	1752	Bayesian hierarchical extreme value model	4 signalized intersections
Fu and Sayed [68]	TTC, MTTC, PET, DRAC	7998 conflicts	video records	24 h	-	traffic volume, shock wave area, platoon ration	-	3 years	Rear-end, daytime, good weather	1095	Random Parameters Bayesian hierarchical extreme value model	4 signalized intersections
Johnsson et al. [66]	mTTC, PET	-	video records	24 h	-	traffic volume	country	7 years	Between cyclists and motor vehicles	2555	GLM (NB)	9 signalized intersections
Mukherjee and Mitra [65]	PET	187,174 crossing behaviors	video records	6 h	pavement marking, night visibility street light	traffic volume, pedestrian traffic, overtaking tendency, speed	land use, zebra cross. following, cross/wait time, cross difficulty, population, attraction zone, residential area	6 years	Fatal Pedestrian	8760	GLM (NB), GLM (Poisson)	110 intersections and 54 midblock segments
Wang et al. [64]	TA, PET, mTTC, MaxD	-	video records (UAV)	4 h × 10 inters.	-	-	-	5 years	Angle, Rear-end	1095	Bivariate extreme value model	10 urban signalized intersections
Zheng et al. [63]	TTC, MTTC, PET, DRAC	-	video records	2 h × 4 inters.	-	-	-	3 years	Rear-end, daytime	3285	Bivariate extreme value model	4 signalized intersections
El-Basyouny and Sayed [69]	TTC	-	conflict survey	8 h × 2 days	class, right turn	traffic volume	-	3 years	Total	1643	Two-phase model: Lognormal (conflicts)—GLM (NB) (crashes)	51 signalized intersections

References

1. World Health Organization. *Global Status Report on Road Safety 2018*; WHO: Geneva, Switzerland, 2018.
2. United Nations. Available online: <https://www.undp.org/sustainable-development-goals> (accessed on 18 December 2022).
3. Theofilatos, A.; Yannis, G.; Kopelias, P.; Papadimitriou, F. Impact of Real-Time Traffic Characteristics on Crash Occurrence: Preliminary Results of the Case of Rare Events. *Accid. Anal. Prev.* **2019**, *130*, 151–159. [[CrossRef](#)]
4. Ziakopoulos, A.; Yannis, G. A review of spatial approaches in road safety. *Accid. Anal. Prev.* **2020**, *135*, 105323. [[CrossRef](#)] [[PubMed](#)]
5. Elvik, R. The predictive validity of empirical Bayes estimates of road safety. *Accid. Anal. Prev.* **2008**, *40*, 1964–1969. [[CrossRef](#)]
6. Yannis, G.; Papadimitriou, E.; Chaziris, A.; Broughton, J. Modeling road accident injury under-reporting in Europe. *Eur. Transp. Res. Rev.* **2014**, *6*, 425–438. [[CrossRef](#)]
7. Janstrup, K.H.; Kaplan, S.; Hels, T.; Lauritsen, J.; Prato, C.G. Understanding traffic crash under-reporting: Linking police and medical records to individual and crash characteristics. *Traffic Inj. Prev.* **2016**, *17*, 580–584. [[CrossRef](#)] [[PubMed](#)]
8. Imprialou, M.; Quddus, M. Crash data quality for road safety research: Current state and future directions. *Accid. Anal. Prev.* **2019**, *130*, 84–90. [[CrossRef](#)] [[PubMed](#)]
9. Wang, C.; Xie, Y.; Huang, H.; Liu, P. A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. *Accid. Anal. Prev.* **2021**, *157*, 106157. [[CrossRef](#)]
10. Tarko, A.P. *Surrogate Measures of Safety, in Safe Mobility: Challenges, Methodology and Solutions*; Emerald Publishing Limited: Bingley, UK, 2018.
11. Bonela, S.R.; Kadali, B.R. Review of Traffic Safety Evaluation at T-intersections Using Surrogate Safety Measures in Developing Countries Context. *IATSS Res.* **2022**, *3*, 307–321. [[CrossRef](#)]
12. Songchitruksa, P.; Tarko, A.P. The extreme value theory approach to safety estimation. *Accid. Anal. Prev.* **2006**, *38*, 811–822. [[CrossRef](#)]
13. Wang, C.; Stamatiadis, N. Evaluation of a simulation-based surrogate safety metric. *Accid. Anal. Prev.* **2014**, *71*, 82–92. [[CrossRef](#)]
14. Sayed, T.; Zein, S. Traffic conflict standards for intersections. *Transp. Plan. Technol.* **1999**, *22*, 309–323. [[CrossRef](#)]
15. Shinar, D. The Traffic Conflict Technique: A Subjective vs. Objective Approach. *J. Safety Res.* **1984**, *15*, 153–157. [[CrossRef](#)]
16. Hydén, C. *The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique Front. Cover*; Lund Institute of Technology Department of Traffic Planning and Engineering: Lund, Sweden, 1987.
17. Chen, P.; Zeng, W.; Yu, G.; Wang, Y. Surrogate Safety Analysis of Pedestrian-Vehicle Conflict at Intersections Using Unmanned Aerial Vehicle Videos. *J. Adv. Transp.* **2017**, *2017*, 5202150. [[CrossRef](#)]
18. Laureshyn, A.; De Ceunynck, T.; Karlsson, C.; Svensson, Å.; Daniels, S. In search of the severity dimension of traffic events: Extended Delta-V as a traffic conflict indicator. *Accid. Anal. Prev.* **2017**, *98*, 46–56. [[CrossRef](#)]
19. Wu, J.Q.; Xu, H.; Zheng, Y.C.; Tian, Z. A novel method of vehicle-pedestrian near-crash identification with roadside LiDAR data. *Accid. Anal. Prev.* **2018**, *121*, 238–249. [[CrossRef](#)] [[PubMed](#)]
20. Gettman, D.; Head, L. Surrogate Safety Measures from Traffic Simulation Models. *Transp. Res. Rec. J. Transportation Res. Board* **2003**, *1840*, 104–115. [[CrossRef](#)]
21. Mahmud, S.S.; Ferreira, L.; Hoque, M.S.; Tavassoli, A. Micro-simulation modelling for traffic safety: A review and potential application to heterogeneous traffic environment. *IATSS Res.* **2018**, *43*, 27–36. [[CrossRef](#)]
22. Guido, G.; Vitale, A.; Astarita, V.; Saccomanno, F.; Giofrè, V.P.; Gallelli, V. Estimation of Safety Performance Measures from Smartphone Sensors. *Procedia Soc. Behav. Sci.* **2012**, *54*, 1095–1103. [[CrossRef](#)]
23. Fazeen, M.; Gozick, B.; Dantu, R.; Bhukhiya, M.; González, M.C. Safe Driving Using Mobile Phones. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 1462–1468. [[CrossRef](#)]
24. Ziakopoulos, A.; Vlahogianni, E.; Antoniou, C.; Yannis, G. Spatial Predictions of Harsh Driving Events Using Statistical and Machine Learning Methods. *Saf. Sci.* **2022**, *150*, 105722. [[CrossRef](#)]
25. Johnsson, C.; Laureshyn, A.; De Ceunynck, T. In search of surrogate safety indicators for vulnerable road users: A review of surrogate safety indicators. *Transp. Rev.* **2018**, *38*, 765–785. [[CrossRef](#)]
26. Stavrakaki, A.-M.; Tselentis, D.I.; Barmponakis, E.; Vlahogianni, E.I.; Yannis, G. Estimating the Necessary Amount of Driving Data for Assessing Driving Behavior. *Sensors* **2020**, *20*, 2600. [[CrossRef](#)]
27. Arun, A.; Haque, M.M.; Washington, S.; Sayed, T.; Mannering, F. A systematic review of traffic conflict-based safety measures with a focus on application context. *Anal. Methods Accid. Res.* **2021**, *32*, 100185. [[CrossRef](#)]
28. Arun, A.; Haque, M.M.; Bhaskar, A.; Washington, S.; Sayed, T. A systematic mapping review of surrogate safety assessment using traffic conflict techniques. *Accid. Anal. Prev.* **2021**, *153*, 10601. [[CrossRef](#)] [[PubMed](#)]
29. Zheng, L.; Sayed, T.; Mannering, F. Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions. *Anal. Methods Accid. Res.* **2021**, *29*, 100142. [[CrossRef](#)]
30. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097. [[CrossRef](#)] [[PubMed](#)]
31. Ziakopoulos, A.; Tselentis, D.; Kontaxi, A.; Yannis, G. A critical overview of driver recording tools. *J. Saf. Res.* **2020**, *72*, 203–212. [[CrossRef](#)] [[PubMed](#)]
32. Strauss, J.; Zangenehpour, S.; Miranda-Moreno, L.; Saunier, N. Cyclist deceleration rate as surrogate safety measure in Montreal using smartphone GPS data. *Accid. Anal. Prev.* **2017**, *99*, 287–296. [[CrossRef](#)]

33. Paleti, R.; Sahin, O.; Cetin, M. Modeling the impact of latent driving patterns on traffic safety using mobile sensor data. *Accid. Anal. Prev.* **2017**, *107*, 92–101. [[CrossRef](#)]
34. Stipancic, J.; Miranda-Moreno, L.; Saunier, N.; Labbe, A. Surrogate safety and network screening: Modelling crash frequency using GPS travel data and latent Gaussian Spatial Models. *Accid. Anal. Prev.* **2018**, *120*, 174–187. [[CrossRef](#)]
35. Stipancic, J.; Miranda-Moreno, L.; Saunier, N. Vehicle manoeuvres as surrogate safety measures: Extracting data from the gps-enabled smartphones of regular drivers. *Accid. Anal. Prev.* **2018**, *115*, 160–169. [[CrossRef](#)]
36. Stipancic, J.; Miranda-Moreno, L.; Saunier, N.; Labbe, A. Network screening for large urban road networks: Using GPS data and surrogate measures to model crash frequency and severity. *Accid. Anal. Prev.* **2019**, *125*, 290–301. [[CrossRef](#)] [[PubMed](#)]
37. Yang, D.; Xie, K.; Ozbay, K.; Yang, H.; Budnick, N. Modelling of time-dependent safety performance using anonymized and aggregated smartphone-based dangerous driving event data. *Accid. Anal. Prev.* **2019**, *132*, 105286. [[CrossRef](#)] [[PubMed](#)]
38. Khorram, B.; Af Wählberg, A.; Tavakoli Kashani, A. Longitudinal jerk and celeration as measures of safety in bus rapid transit drivers in Tehran. *Theor. Issues Ergon. Sci.* **2020**, *21*, 577–594. [[CrossRef](#)]
39. Guo, M.; Zhao, X.; Yao, Y.; Yan, P.; Su, Y.; Bi, C.; Wu, D. A study of freeway crash risk prediction and interpretation based on risky driving behavior and traffic flow data. *Accid. Anal. Prev.* **2021**, *160*, 106328. [[CrossRef](#)] [[PubMed](#)]
40. Mantouka, E.G.; Barmponakis, E.N.; Vlahogianni, E.I. Mobile sensing and machine learning for identifying driving safety profiles. In Proceedings of the Transportation Research Board 97th Annual Meeting, Washington, DC, USA, 1–7 January 2018.
41. Gündüz, G.; Yaman, Ç.; Peker, A.U.; Acarman, T. Prediction of Risk Generated by Different Driving Patterns and Their Conflict Redistribution. *IEEE Trans. Intell. Veh.* **2017**, *3*, 71–80. [[CrossRef](#)]
42. Tselentis, D.I.; Yannis, G.; Vlahogianni, E.I. Innovative motor insurance schemes: A review of current practices and emerging challenges. *Accid. Anal. Prev.* **2017**, *98*, 139–148. [[CrossRef](#)] [[PubMed](#)]
43. Stephens, A.N.; Groeger, J.A. Situational specificity of trait influences on drivers' evaluations and driving behaviour. *Transp. Res. Part F Traffic Psychol. Behav.* **2009**, *12*, 29–39. [[CrossRef](#)]
44. Kamla, J.; Parry, T.; Dawson, A. Analysing truck harsh braking incidents to study roundabout accident risk. *Accid. Anal. Prev.* **2019**, *122*, 365–377. [[CrossRef](#)]
45. Park, S.; Son, S.O.; Park, J.; Oh, C.; Hong, S. Using vehicle data as a surrogate for highway accident data. *Proc. Inst. Civ. Eng.—Munic. Eng.* **2021**, *174*, 67–74. [[CrossRef](#)]
46. Kontaxi, A.; Ziakopoulos, A.; Yannis, G. Trip Characteristics Impact on the Frequency of Harsh Events Recorded via Smartphone Sensors. *IATSS Res.* **2021**, *45*, 574–583. [[CrossRef](#)]
47. Zhao, X.; Yang, H.; Yao, Y.; Qi, H.; Guo, M.; Su, Y. Factors Affecting Traffic Risks on Bridge Sections of Freeways Based on Partial Dependence Plots. *Phys. A Stat. Mech. Its Appl.* **2022**, *598*, 127343. [[CrossRef](#)]
48. Ball, K.; Ackerman, M. The older driver (training and assessment: Knowledge, skills and attitudes). In *Handbook of Driving Simulation for Engineering, Medicine and Psychology*; CRC Press: Boca Raton, FL, USA, 2011.
49. Regan, M.; Williamson, A.; Grzebieta, R.; Tao, L. Naturalistic driving studies: Literature review and planning for the Australian naturalistic driving study. In Proceedings of the Australasian College of Road Safety Conference 2012, Sydney, Australia, 9–10 August 2012.
50. Kim, S.K.; Song, T.J.; Roupail, N.M.; Aghdashi, S.; Amaro, A.; Gonçalves, G. Exploring the association of rear-end crash propensity and micro-scale driver behavior. *Saf. Sci.* **2016**, *89*, 45–54. [[CrossRef](#)]
51. Pande, A.; Chand, S.; Saxena, N.; Dixit, V.; Loy, J.; Wolshon, B.; Kent, J.D. A preliminary investigation of the relationships between historical crash and naturalistic driving. *Accid. Anal. Prev.* **2017**, *101*, 107–116. [[CrossRef](#)]
52. Ambros, J.; Altmann, J.; Jurewicz, C.; Chevalier, A. Proactive Assessment of Road Curve Safety Using Floating Car Data: An Exploratory Study. *Arch. Transp.* **2019**, *50*, 7–15. [[CrossRef](#)]
53. Mousavi, S.M.; Zhang, Z.; Parr, S.A.; Pande, A.; Wolshon, B. Identifying High Crash Risk Highway Segments Using Jerk-Cluster Analysis. In Proceedings of the International Conference on Transportation and Development 2019: Smarter and Safer Mobility and Cities, Alexandria, VA, USA, 9–12 June 2019.
54. Stipancic, J.; Racine, E.B.; Labbe, A.; Saunier, N.; Miranda-Moreno, L. Massive GNSS data for road safety analysis: Comparing crash models for several Canadian cities and data sources. *Accid. Anal. Prev.* **2021**, *159*, 106232. [[CrossRef](#)] [[PubMed](#)]
55. Desai, J.; Li, H.; Mathew, J.K.; Cheng, Y.-T.; Habib, A.; Bullock, D.M. Correlating Hard-Braking Activity with Crash Occurrences on Interstate Construction Projects in Indiana. *J. Big Data Anal. Transp.* **2021**, *3*, 27–41. [[CrossRef](#)]
56. Hunter, M.; Saldivar-Carranza, E.; Desai, J.; Mathew, J.K.; Li, H.; Bullock, D.M. A Proactive Approach to Evaluating Intersection Safety Using Hard-Braking Data. *J. Big Data Anal. Transp.* **2021**, *3*, 81–94. [[CrossRef](#)]
57. Li, P.; Abdel-Aty, M.; Yuan, J. Using bus critical driving events as surrogate safety measures for pedestrian and bicycle crashes based on GPS trajectory data. *Accid. Anal. Prev.* **2021**, *150*, 105924. [[CrossRef](#)]
58. Li, X.; Mousavi, S.M.; Dadashova, B.; Lord, D.; Wolshon, B. Toward a Crowdsourcing Solution to Identify High-Risk Highway Segments through Mining Driving Jerks. *Accid. Anal. Prev.* **2021**, *155*, 106101. [[CrossRef](#)]
59. Lu, G.; Cheng, B.; Kuzumaki, S.; Mei, B. Relationship between Road Traffic Accidents and Conflicts Recorded by Drive Recorders. *Traffic Inj. Prev.* **2011**, *12*, 320–326. [[CrossRef](#)] [[PubMed](#)]
60. He, Z.; Qin, X.; Liu, P.; Sayed, M.A. Assessing Surrogate Safety Measures Using a Safety Pilot Model Deployment Dataset. *Transp. Res. Rec.* **2018**, *2672*, 1–11. [[CrossRef](#)]
61. Risser, R. Behavior in traffic conflict situations. *Accid. Anal. Prev.* **1985**, *17*, 179–197. [[CrossRef](#)] [[PubMed](#)]

62. Alhajyaseen, W.K. The integration of conflict probability and severity for the safety assessment of intersections. *Arab. J. Sci. Eng.* **2015**, *40*, 421–430. [[CrossRef](#)]
63. Zheng, L.; Sayed, T.; Essa, M. Validating the Bivariate Extreme Value Modeling Approach for Road Safety Estimation with Different Traffic Conflict Indicators. *Accid. Anal. Prev.* **2019**, *123*, 314–323. [[CrossRef](#)]
64. Wang, C.; Xu, C.; Dai, Y. A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data. *Accid. Anal. Prev.* **2019**, *123*, 365–373. [[CrossRef](#)]
65. Mukherjee, D.; Mitra, S. Comprehensive study of risk factors for fatal pedestrian crashes in urban setup in a developing country. *Transp. Res. Rec.* **2020**, *2674*, 100–118. [[CrossRef](#)]
66. Johnsson, C.; Laureshyn, A.; Dágostino, C. Validation of Surrogate Measures of Safety with a Focus on Bicyclist–Motor Vehicle Interactions. *Accid. Anal. Prev.* **2021**, *153*, 106037. [[CrossRef](#)]
67. Fu, C.; Sayed, T. Comparison of threshold determination methods for the deceleration rate to avoid a crash (DRAC)-based crash estimation. *Accid. Anal. Prev.* **2021**, *153*, 106051. [[CrossRef](#)]
68. Fu, C.; Sayed, T. Random parameters Bayesian hierarchical modeling of traffic conflict extremes for crash estimation. *Accid. Anal. Prev.* **2021**, *157*, 106159. [[CrossRef](#)]
69. El-Basyouny, K.; Sayed, T. Safety performance functions using traffic conflicts. *Saf. Sci.* **2013**, *51*, 160–164. [[CrossRef](#)]
70. Saccomanno, F.F.; Cunto, F.; Guido, G.; Vitale, A. Comparing safety at signalized intersections and roundabouts using simulated rear-end conflicts. *Transp. Res. Rec.* **2008**, *2078*, 90–95. [[CrossRef](#)]
71. Ozbay, K.; Martin, B.; Yang, H. Derivation and Validation of New Simulation-Based Surrogate Safety Measure. *Transp. Res. Rec. J. Transp. Res. Board* **2008**, *2083*, 105–113. [[CrossRef](#)]
72. Zheng, L.; Sayed, T. Comparison of traffic conflict indicators for crash estimation using peak over threshold approach. *Transp. Res. Rec.* **2019**, *2673*, 493–502. [[CrossRef](#)]
73. Lu, N.; Cheng, N.; Zhang, N.; Shen, X.; Mark, J.W. Connected vehicles: Solutions and challenges. *IEEE Internet Things J.* **2014**, *1*, 289–299. [[CrossRef](#)]
74. Xie, K.; Yang, D.; Ozbay, K.; Yang, H. Use of real-world connected vehicle data in identifying high-risk locations based on a new surrogate safety measure. *Accid. Anal. Prev.* **2019**, *125*, 311–319. [[CrossRef](#)] [[PubMed](#)]
75. Hu, J.; Huang, M.-C.; Yu, X. Efficient mapping of crash risk at intersections with connected vehicle data and deep learning models. *Accid. Anal. Prev.* **2020**, *144*, 105665. [[CrossRef](#)]
76. Yang, D.; Xie, K.; Ozbay, K.; Yang, H. Fusing crash data and surrogate safety measures for safety assessment: Development of a structural equation model with conditional autoregressive spatial effect and random parameters. *Accid. Anal. Prev.* **2021**, *152*, 105971. [[CrossRef](#)]
77. Dobson, A.J.; Barnett, A.G. *An Introduction to Generalized Linear Models*, 4th ed.; CRC Press: London, UK, 2018.
78. Guo, F.; Klauer, S.G.; Hankey, J.M.; Dingus, T.A. Near crashes as crash surrogate for naturalistic driving studies. *Transp. Res. Rec.* **2010**, *2147*, 66–74. [[CrossRef](#)]
79. Lord, D.; Mannering, F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transp. Res. Part A Policy Pract.* **2010**, *44*, 291–305. [[CrossRef](#)]
80. Washington, S.; Karlaftis, M.; Mannering, F.; Anastasopoulos, P. *Statistical and Econometric Methods for Transportation Data Analysis*, 3rd ed.; Chapman and Hall/CRC: London, UK, 2020.
81. Blangiardo, M.; Cameletti, M. *Spatial and Spatio-Temporal Bayesian Models with R-INLA*; John Wiley & Sons: Hoboken, NJ, USA, 2015; ISBN 1-118-32655-5.
82. Rue, H.; Martino, S.; Chopin, N. Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion). *J. R. Stat. Soc. B* **2009**, *71*, 319–392. [[CrossRef](#)]
83. Lindgren, F.; Rue, H.; Lindström, J. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2011**, *73*, 423–498. [[CrossRef](#)]
84. Lindgren, F.; Rue, H. Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* **2015**, *63*, 1–25. [[CrossRef](#)]
85. Coles, S. *An Introduction to Statistical Modeling of Extreme Values*; Springer: London, UK, 2001.
86. Boonsiripant, S.; Rodgers, M.O.; Hunter, M.P. Speed profile variation as a road network screening tool. *Transp. Res. Rec.* **2011**, *2236*, 83–91. [[CrossRef](#)]
87. Arun, A.; Haque, M.M.; Bhaskar, A.; Washington, S.; Sayed, T. A bivariate extreme value model for estimating crash frequency by severity using traffic conflicts. *Anal. Methods Accid. Res.* **2021**, *32*, 100180. [[CrossRef](#)]
88. Arun, A.; Haque, M.M.; Bhaskar, A.; Washington, S. Transferability of Multivariate Extreme Value Models for Safety Assessment by Applying Artificial Intelligence-Based Video Analytics. *Accid. Anal. Prev.* **2022**, *170*, 106644. [[CrossRef](#)]
89. Goyani, J.; Paul Aninda, B.; Gore, N.; Arkatkar, S.; Joshi, G. Investigation of crossing conflicts by vehicle type at unsignalized t-intersections under varying roadway and traffic conditions in India. *J. Transp. Eng. A Syst.* **2021**, *147*, 05020011. [[CrossRef](#)]
90. Yannis, G.; Nikolaou, D.; Laiou, A.; Stürmer, Y.A.; Buttler, I.; Jankowska-Karpa, D. Vulnerable road users: Cross-cultural perspectives on performance and attitudes. *IATSS Res.* **2020**, *44*, 220–229. [[CrossRef](#)]
91. Karpinski, E.; Bayles, E.; Sanders, T. Safety Analysis for Micromobility: Recommendations on Risk Metrics and Data Collection. *Transp. Res. Rec.* **2022**, *2676*, 420–435. [[CrossRef](#)]
92. Ali, Y.; Haque, M.M.; Mannering, F. A Bayesian generalised extreme value model to estimate real-time pedestrian crash risks at signalised intersections using Artificial Intelligence-based Video Analytics. *Anal. Methods Accid. Res.* **2022**, *38*, 100264. [[CrossRef](#)]

93. Wang, X.; Yang, J.; Lee, C.; Ji, Z.; You, S. Macro-level safety analysis of pedestrian crashes in Shanghai, China. *Accid. Anal. Prev.* **2016**, *96*, 12–21. [[CrossRef](#)] [[PubMed](#)]
94. Hussain, F.; Li, Y.; Arun, A.; Haque, M.M. A Hybrid Modelling Framework of Machine Learning and Extreme Value Theory for Crash Risk Estimation Using Traffic Conflicts. *Anal. Methods Accid. Res.* **2022**, *36*, 100248. [[CrossRef](#)]
95. Hussein, H.; Radwan, M.H.; Elsayed, H.A.; Abd El-Kader, S.M. Depth-First-Search-Tree Based D2D Power Allocation Algorithms for V2I/V2V Shared 5G Network Resources. *Wirel. Netw.* **2021**, *27*, 3179–3193. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.