Transportation Research Part C 91 (2018) 156–175

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

# Exploring relationships between driving events identified by in-vehicle data recorders, infrastructure characteristics and road crashes

Victoria Gitelman[a], Shlomo Bekhor[a,*], Etti Doveh[b], Fany Pesahov[a], Roby Carmel[a], Smadar Morik[a]

[a] Transportation Research Institute, Technion – Israel Institute of Technology, Haifa 32000, Israel
[b] The Technion Statistical Laboratory, Technion – Israel Institute of Technology, Haifa 32000, Israel

ARTICLE INFO

*Keywords:*
In-vehicle data recorder
Driving events
Infrastructure characteristics
Interurban roads
Road crashes

ABSTRACT

There is an increasing interest in technology-based solutions that can assist drivers in reducing their risk of involvement in road crashes. Previous studies showed that driving events produced by in-vehicle data recorders (IVDR) are applicable for identification of unsafe driving patterns, while combined examinations of driving events and road infrastructure characteristics are rare. This study explored the relationship between the IVDR-driving events, road characteristics and crashes, to examine a potential of the events for predicting crashes and identification of high-risk locations on the road network. The study database included 3500 segments of the interurban roads in Israel, for which the automatically produced IVDR events were matched with road infrastructure characteristics and crashes. Negative-binomial regression models were adjusted for the relationships between road characteristics and driving events, and subsequently, between events and crashes, given the exposure. Significant impacts were found, yet various event types showed different relations to the infrastructure characteristics and different effects on crashes, on various road types. Better road conditions were associated with a decrease in "braking" events and an increase in the "speed alert" events, where road layout constraints and junction proximity were associated with an opposite effect on events. "Braking" and total events showed better potential for predicting crashes on single-carriageway roads, with a positive link to crashes, where for other road types the "speed alert" events were stronger related to crashes, but with a negative link. The heterogeneity of findings indicates a need in further research of the above relationship, with a particular focus on definitions of driving events produced by the IVDR or other technologies.

## 1. Introduction

It is well known in the road safety literature that human factors contribute to the majority of road crashes, e.g. 93% as estimated by Treat et al. (1979). Given the importance of understanding the driver's interaction with road, vehicle and the environment for preventing crashes, naturalistic driving studies (NDS) are being conducted over the past decade to collect objective data. In a NDS, volunteer participants drive an instrumented vehicle, which continuously records their driving behavior, the vehicle behavior and the

* Corresponding author.
*E-mail addresses:* trivica@technion.ac.il (V. Gitelman), sbekhor@technion.ac.il (S. Bekhor), ierde01@ie.technion.ac.il (E. Doveh), fanyp@technion.ac.il (F. Pesahov), robyc@cv.technion.ac.il (R. Carmel), smadarm@technion.ac.il (S. Morik).

https://doi.org/10.1016/j.trc.2018.04.003
Received 2 April 2016; Received in revised form 20 December 2017; Accepted 6 April 2018
Available online 11 April 2018

behavior of other road users with whom they interact. Examples of such studies are the seminal "100-car naturalistic driving study" (Klauer et al., 2006), which explored factors leading to rear-end crashes, and the most recent the US Strategic Highway Research Program Phase 2 (SHRP 2) NDS, which deployed 2800 vehicles to explore and analyze a wide range of road safety problems (Antin et al., 2011; Hallmark et al., 2013; TRB, 2014) and is currently underway, on the stage of the NDS data analysis. Other large-scale efforts can be found in Japan (Uchida et al., 2010) and in Europe where the European Commission funded UDRIVE project (e.g. Twisk et al., 2012).

Regan et al. (2012) reviewed around 40 studies that used the NDS approach, and noted that most of them have been small-scale studies. Several research issues have been examined, including factors leading to rear-end crashes (e.g. Klauer et al., 2006); skill development in young drivers (Prato et al., 2010); skill loss in older drivers (e.g. Blanchard et al., 2010); young novice driver crash and incident types (e.g. Lee et al., 2011); distraction and inattention (e.g. Olson et al., 2009; Koppel et al., 2011); fatigue (e.g. Hanowski et al., 2009); interactions between light and heavy vehicle drivers (Hanowski et al., 2007); understanding driver interactions with new vehicle safety technologies (e.g. Sayer et al., 2007); and lane changing behavior (e.g. Lee et al., 2004). Recently, a large-scale study on driver performance and behavior crash risk factors was reported (Dingus et al., 2016).

It is noticeable that the majority of the studies examined the interaction between drivers and the surrounding environment and/or driver characteristics, while the examination of driving events and road infrastructure characteristics are still rare. Hallmark et al. (2011) performed a comprehensive review of the roadway, driver, environmental and vehicle data needs to address lane departures using NDS data. Their study identified the following roadway elements as important for addressing the issue: horizontal and vertical curves, roadway cross section, driveway density, roadway lighting, rumble strips, roadway delineation and sighting, and pavement edge drop-off. Cannon and Sudweeks (2011) used NDS data to compare driver behavior between high-crash and low-crash rural road sites and found a few significant differences, without a consistent pattern. Using the "100 cars" NDS data, Guo et al. (2010) explored near-crash versus crash events and demonstrated similarity in factors contributing to both events, including road conditions such as road surface, lighting, road alignment. Further inputs on the topic are expected from the SHRP 2 NDS, based on combined analyses of driver behavior and detailed roadway data (Hallmark et al., 2013).

Jovanis et al. (2011) noted that comparatively few NDS studies have emphasized exposure-based analyses. In this respect, Shankar et al. (2008) formulated exposure-based risk measures that can be derived from NDS data to perform surrogate analyses. They comment that an essential starting point for analyzing NDS data is to establish and define surrogates beyond the conventionally considered events such as near-crashes. Wu and Jovanis (2012, 2013) further explored the definitions suggesting a distinction between safety-related and surrogate events. The former includes events that were traditionally collected and examined by road safety behavioral studies and referred to as near crashes, risky driving or near misses, while the latter includes both crashes and near crashes with common etiologies to crashes. In any case, a surrogate measure is considered as an indirect measure of safety, where the need in the use of surrogate measures is explained by rareness of crash events that limits the possibilities of studying crash contributing factors (Guo et al., 2010; Hallmark et al., 2013; Wu et al., 2014).

It is worth mentioning that examples of using surrogate measures to evaluate traffic safety were provided by traffic conflict techniques (TCT) that were introduced several decades ago (Perkins and Harris, 1968; Hyden, 1987). Studies applying such techniques demonstrated relationships between traffic conflicts and crashes and, particularly, when exploring road infrastructure impacts. For example, Sayed and Zein (1999) validated the relationship between crash frequencies and observed conflicts at signalized intersections; Retting and Greene (1997) used traffic conflicts to examine the impact of traffic signal timing at urban intersections; Hanowski et al. (2004) demonstrated the application of modified TCT (critical incidents) to ascertain the needs in re-design of existing urban road infrastructure. As stated (Guo et al., 2010; Wu et al., 2014), naturalistic driving approach has much in common with TCT, except for that the instrumentation is vehicle-based rather than road site-based, where NDS approach is much more powerful and promising. Yet, it should be noted that both approaches are not well-controlled studies.

### 1.1. In-vehicle data recorders

There has been an increased interest in technology-based solutions that can assist drivers in reducing their risk of involvement in car crashes. One class of solutions that were proposed is the installation of in-vehicle data recorders (IVDRs), which monitor and provide feedback on driver behavior (Toledo et al., 2008; McGehee et al., 2007; Musicant et al., 2014). IVDRs can be viewed as a component to NDS, which focuses on data acquisition and processing related to the vehicle. The IVDR system is typically based on measuring gravitational forces and GPS location of the vehicles. The type of data stored by an IVDR system may range from full acquisition of all vehicle sensor measurements to concise representation of main indices, e.g. speed, event type. Some systems log data continuously, others store data related to undesirable driver events. The key advantages of commercial IVDRs are the low cost and the "ready to use" package of identification, logging and reporting capabilities (Musicant et al., 2014). IVDR does not include video records and thus, unlike most naturalistic driving studies, it does not require substantial data processing enabling large-scale data collection at a reasonable cost (Toledo et al., 2014). However, IVDR pattern recognition procedures are typically proprietary, thus not permitting to focus on detailed data and possibly reducing comparability of findings received based on various systems.

In Israel, an IVDR system referred to as "green box" collected naturalistic data in private vehicles driven by novice young drivers and their family members (Lotan et al., 2010; Farah et al., 2013). This is a g-force based system developed by Green-Road technologies. The system tracks all trips made by the vehicle and records: trip start and end times; vehicle and driver identification, using magnetic keys; vehicle location measured by a GPS receiver; events of excessive maneuvers defined by patterns of g-forces measured in the vehicle. For example, the accelerations of the vehicle, both in the lateral and longitudinal directions, are measured by accelerometers at a sampling rate of 40 measurements per second. Using raw information, the system defines over 20 driving maneuvers that may be classified into five major categories: braking, accelerating, turn handling, lane handling and speeding. Maneuvers are also defined by their relative direction and their level of severity based on the parameters of the detailed trajectory (e.g., maneuver duration, extent of sudden changes in speed and acceleration, and the speed they are performed at). Based on continuously measured speeds, accelerations and steering wheel movements, and consequent processing of these data dangerous driving maneuvers (or driving events) undertaken by the vehicle, during each trip, are identified. The information about trips and events is transferred via wireless network to a designated server where it is logged for further processing. More details on the system can be found in Fleishman et al. (2003), Prato et al. (2010), and Farah et al. (2013).

Recent studies demonstrated usability of the IVDR data for research purposes. Toledo and Lotan (2007), Toledo et al. (2008) showed that the rates of dangerous driving maneuvers recognized by the IVDR could be used as indicators of the risk of crash involvement. Prato et al. (2010) and Farmer et al. (2010) applied the IVDR data for studying the driving behavior of novice drivers. Farah et al. (2013, 2014) used the IVDR data to examine the impact of various feedbacks on young drivers' behavior. Toledo et al. (2014) demonstrated the applicability of the IVDR data for examining temporal and spatial driving patterns. Musicant et al. (2014) used the IVDR-recognizable driving events to explore over-time changes in driver behaviors. Shiftan and Toledo (2015) examined the IVDR events to evaluate changes in driving behavior related to improved safety and reduced fuel consumption.

Note that all studies above that used the IVDR data related between events and personal characteristics, in order to explain differences in driving patterns. In addition, Paefgen et al. (2014) demonstrated the use of the IVDR-data for Modeling exposure-crash relationship while controlling for the situational variables of daytime, weekday, road type and velocity. The originality of the present study is to use the IVDR driving events combined with detailed infrastructure characteristics and crashes, in an attempt to identify possible safety problems on road sections (as indicated by the events). In such context, the IVDR driving events were not analyzed in the past.

In general, the reason for using the IVDR data is similar to that for using surrogate measures for crashes in NDS studies, i.e. because the number of crashes in not large enough for exploring risk-contributing factors. This idea is closely related to Heinrich's triangle (Heinrich, 1959) which showed that less-severe crashes happen more frequently than severe ones and that more severe injuries can be reduced by reducing the frequency of minor injuries. Further development of the pyramid, for surrogate safety measures, was suggested by Hyden (1987) and Songchitruksa and Tarko (2006), where more layers were added to the pyramid from the bottom indicating the way from normal events to potential danger, then near crashes and, finally, crashes (at the top). Following similar lines, the IVDR data may assist in identification of potentially dangerous driving events related to near crash and crash occurrences, where reducing the risk factors related to the events will contribute to a reduction in crashes.

### 1.2. The study focus

Previous studies of the IVDR data in Israel (Toledo et al., 2008; Prato et al., 2010; Lotan et al., 2010; Farah et al., 2013, 2014) demonstrated the driving events' contribution for identification and treatment of driver-related risk factors. Since the data collected enable to restore the driving event locations on the road map, further analysis of the events considering infrastructure-related factors can be undertaken. Such a consideration became possible as, in recent years, the National Road Infrastructure Company in Israel established a Safety Management System (SMS) that provides data and tools for black-spot identification and carrying out various analyses of the interurban road network (Gitelman et al., 2015). Building the SMS database, a detailed road survey was conducted, in which the infrastructure characteristics of interurban roads were recorded.

This led to possibility of a research based on matching of two existing databases: driving events obtained from the IVDR systems (as introduced in Lotan et al., 2010; Farah et al., 2013, 2014), and the infrastructure characteristics at the event (and other) locations produced from the SMS data. The main objective of this study was to examine the relationship between the occurrence of driving events and the infrastructure characteristics, on interurban roads in Israel. Another objective was to explore a relationship between the driving events and crashes. The main hypothesis underlying this study was that the relationship between drivers' behavior (as expressed through the driving events) and the infrastructure characteristics can be established, where the findings of the analysis can contribute to a deeper understanding of the interactions between the infrastructure characteristics and drivers' behavior. Furthermore, understanding such relationship may assist in early identification of high-risk locations on the road network, once a relation between driving events and crashes is ascertained. The idea of using driving events for identification of high-risk locations was also raised by other studies (Pande et al., 2014, Wolshon and Parr, 2015), where, e.g., Wolshon and Parr (2015) demonstrated statistical correlations between clusters of high magnitude jerk events (rate of change of acceleration) and long-term crash rates, on two urban arterials.

The methodological approach used in the study was to utilize advanced technology data and statistical tools to find the best fit between the infrastructure variables and driving events, and the events and crashes, using a fine resolution of road segments. This approach is similar, e.g., to Quddus (2013), who explored the relationship between speed and crash rates. Disaggregated segment-based traffic, road geometry, and crash data from road segments in motorways and trunk roads around London were used in the analysis. Average speeds were not found to be significantly associated with crash rates, when controlling for other factors affecting crashes such as traffic volume, road geometry (e.g., grade and curvature), and number of lanes; yet, speed variation was found to be positively associated with crash rates. A site-based approach was also applied by Shi and Abdel-Aty (2015), who investigated real-time traffic monitoring data produced by a microwave vehicle detection system that is deployed on an expressway network in Orlando, to unveil the effects of traffic dynamics on rear-end crash occurrence. Crashes and non-crash cases at the same road locations were selected for the analysis accompanied by traffic and speed characteristics in time and space, and the infrastructure characteristics (number of lanes, speed limits, curvature). The study showed significant associations between congestion indicators

**Table 1**
Events defined by the "green-box" system (Lotan et al., 2010).

| Event No. | Description | Event No. | Description | Event No. | Description |
|---|---|---|---|---|---|
| 1 | Braking | 8 | Accelerate while in turn | 15 | Trip start |
| 2 | Accelerating | 9 | Braking while in turn | 16 | Trip end |
| 3 | Braking into turn | 10 | Accelerate while exiting turn | 17 | Estimated trip start |
| 4 | Accelerate into turn | 11 | Braking while exiting turn | 18 | Estimated trip end |
| 5 | Accelerate while in turn | 12 | Lane change | 19 | Speed alert |
| 6 | Sudden brake in turn | 13 | Bypass | 20 | Collision suspect |
| 7 | Turning | 14 | Lane handling | | |

and crashes having controlled for other variables. In general, such segment- or site-based approach is common in road safety studies where a model for predicting crash occurrences is typically fitted using traffic exposure and road infrastructure characteristics (Elvik et al., 2009; Bonneson and Pratt, 2009).

The study database included combined information on driving events, infrastructure characteristics, travel exposure, and crash numbers collected for selected interurban roads, in similar periods: mostly, the year 2010, except for the crash numbers supplied for 2008–2010. This paper presents the results of the analyses carried out on the study database aiming to examine the relationship between the occurrence of driving events and road infrastructure characteristics, accounting for traffic exposure, as well as to explore the relationship between the road crashes and the driving events, given the exposure. It was expected that this study may contribute to general knowledge on driving events' occurrence in the context of road infrastructure characteristics, and on the relationship between driving events and crashes, under various road conditions. Once the relations are unveiled, driving events may serve as crash substitutes, enabling to identify potential dangerous locations on the road network and providing a basis for preventive road safety work.

## 2. Methodology

### 2.1. Data preparation

The preparation of the research database included two stages: (a) mapping the IVDR events' data on the road network and (b) preparing the integrated database for the research, including driving events, infrastructure characteristics, exposure indicators and crash numbers.

The IVDR data used in the current analysis corresponds to a "control group" of drivers in Lotan et al. (2010), Farah et al. (2013, 2014), i.e. those driving the vehicles with the IVDR devices but without any feedback concerning their driving style and the events recorded. This group included 64 vehicles out of over 200 vehicle fleet equipped with the IVDR. The IVDR data collected enable to restore the vehicle's location on the road map, at the beginning and the end of each trip and at regular time points during the trip, e.g. every minute and at the time moments adjacent to the driving event occurrences. A one-year database of the IVDR records comprised 3.4 million of GPS observations distributed throughout the country. The "green box" IVDR system identifies and classifies 20 different events, as presented in Table 1. The event classification is based on parameters of the detailed trajectory of the vehicle during the maneuver, such as its duration and extent of sudden changes in the vehicle movement, and on the speed at the maneuver point. Fig. 1 presents the distribution of the events by type, for the whole "control group" in Lotan et al. (2010). The results clearly show a

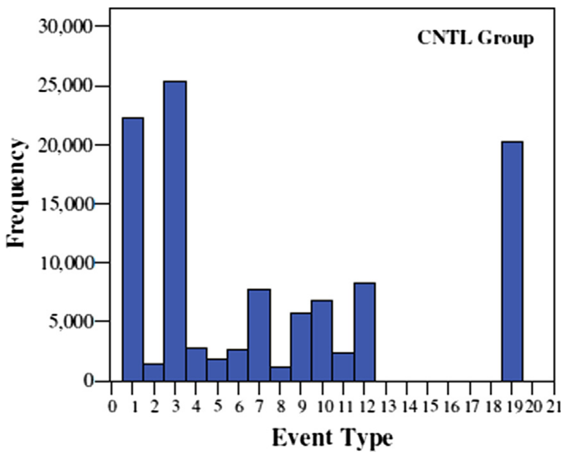



**Fig. 1.** Distribution of events by type, in the "control group" (Lotan et al., 2010).

**Table 2**
Road sections included in the analysis, with the number of segments and frequencies of the IVDR-vehicle travels.

| Road No | Total road length (km) | GPS observations (thousands) | Road length included in the analysis (km) | Number of 200 m segments | Average number of IVDR trackings per segment | Standard deviation of IVDR trackings per segment |
|---|---|---|---|---|---|---|
| 1 | 95.7 | 72.4 | 67.4 | 337 | 188.5 | 76.0 |
| 2 | 99.2 | 77.1 | 87.8 | 439 | 128.0 | 87.8 |
| 4 | 194.8 | 134.2 | 155.6 | 778 | 137.1 | 131.0 |
| 40 | 301.5 | 59.3 | 195.0 | 975 | 36.4 | 53.2 |
| 44 | 40.0 | 9.6 | 30.4 | 152 | 57.3 | 50.7 |
| 65 | 90.2 | 31.9 | 80.2 | 401 | 23.1 | 53.5 |
| 70 | 75.4 | 20.0 | 16.4 | 82 | 12.4 | 5.1 |
| 90 | 468.2 | 9.9 | 30.6 | 153 | 6.7 | 1.5 |
| 444 | 38.8 | 43.6 | 36.6 | 183 | 92.1 | 94.9 |
| Total | 1403.8 | 457.9 | 700.0 | 3500 | 85.3 | 102.1 |

concentration of “braking” events (1 and 3) and “speed alert” events (19) related to other event types.

Given the large number of the IVDR driving data and a significant investment of resources needed for the assignment of these data on the road network, not all roads were included in the current analysis, but a selected subset of the interurban roads in Israel. For the preparation of the study database, nine interurban roads were selected: No 1, 2, 4, 40, 44, 65, 70, 90, and 444, representing the main roads in the center region with an addition of typical roads used for recreation and tourist travels in the north and the south of the country. The roads selected comprise over a quarter of the total rural network of the country, where they present a combination of different road types. On the nine roads, 62 IVDR vehicles were observed (out of 64) indicating that the group of roads selected for the analysis was representative for the drivers' population covered by the original study.

Since both GPS observations from IVDR vehicles and road infrastructure data contain errors, a map-matching algorithm was developed to minimize such errors. The algorithm has two main characteristics: (a) each GPS sampling point is allowed to match more than one “candidate” road segment, and (b) drivers choose the shortest route between two consecutive GPS sampling points. The assumption above may fail when the sampling rate drops significantly. Since the sampling rate of GPS points is relatively frequent, this assumption is acceptable and logical. The path between an origin and destination is then formed by joining the shortest paths between two consecutive sampling points. Note that this path may be substantially different from the shortest path between the start and endpoints of a given trip.

The roads were divided into analysis segments of 200 m long, in a way similar to Paefgen et al. (2014). This resolution was selected accounting for the frequency of the IVDR data, changes in road conditions and accuracy in reporting road crashes. The bottom threshold was 100 m length as dictated by the units of road conditions. On the other hand, the longer the segments the more data would need averaging; thus, a close to the bottom value of 200 m was selected. Each segment is defined by road number, driving direction, start km and end km. For each segment, the data on the total number of travels – the IVDR trackings, the number of drivers and the number of vehicles traveled through the section, in the year 2010, were prepared. In addition, the number of driving events that occurred throughout the year was computed for each segment. To allow for meaningful analysis, the “effective” road length was reduced to include the segments with at least 5 IVDR-vehicle trips and with at least 3 different vehicles involved, as aggregated figures for 2010. These thresholds were selected arbitrarily, aiming to exclude road segments that were not used by the IVDR vehicles, but still keeping the amount of units enabling the analysis. The threshold of 3 different vehicles was applied as it corresponded to 5% of the total vehicles involved in the data collection, and enabled us to keep in the study dataset 10% of segments, mostly from remoted areas. In addition, the minimum number of vehicle trips per road unit was similar for segments with 3, 4, and 5 different vehicles. Table 2 shows the total road length and the number of segments that were included in the analysis, with the IVDR tracking statistics and the amount of processed GPS observations. The data demonstrate that applying the filtering thresholds the total length of the segments studied was reduced by half. Fig. 2 shows a country's map with the road segments covered by the study. It can be seen that road units included in the research provide a wide coverage of the roads examined. On most study roads, a substantial amount of the IVDR-vehicle travels was observed.

The infrastructure characteristics were produced based on the SMS data for each predefined segment unit. These characteristics are: road type; traffic volume; road layout characteristics – lane width, right shoulder width, left shoulder width (on a divided road only); a horizontal radius and a vertical grade as well as the characteristics reflecting the change of the infrastructure along the road section such as: the unit's proximity to an intersection/interchange when approaching a junction (“junction ahead”) or distancing it (“junction behind”); the length of road section to which the unit belongs (a distance between the main junctions); a change in the number of lanes with regard to the previous unit; a change in the width of the lanes with regard to the previous unit. All infrastructure characteristics are categorical variables, which were defined in accordance with existing road types and accounting for the demands of current design guidelines. Three rural road types were considered in the analysis, as follows: a *freeway* that corresponds to a divided multi-lane road without at-grade junctions; a *dual-carriageway* road that is a multi-lane road with a median (typically, with a barrier) and at-grade junctions; and a *single-carriageway* road that is a two-lane road without median. Table 3 provides the definitions of traffic and road infrastructure characteristics included in the analysis. To note, traffic volumes were assigned based on the Central Bureau of Statistics (CBS) on-road measurements that take place, on average, once in two years. Due to lower accuracy of these values related to the IVDR-travels, a categorical instead of a continuous variable was defined (as an indication of the level of general traffic volume on the road segment).

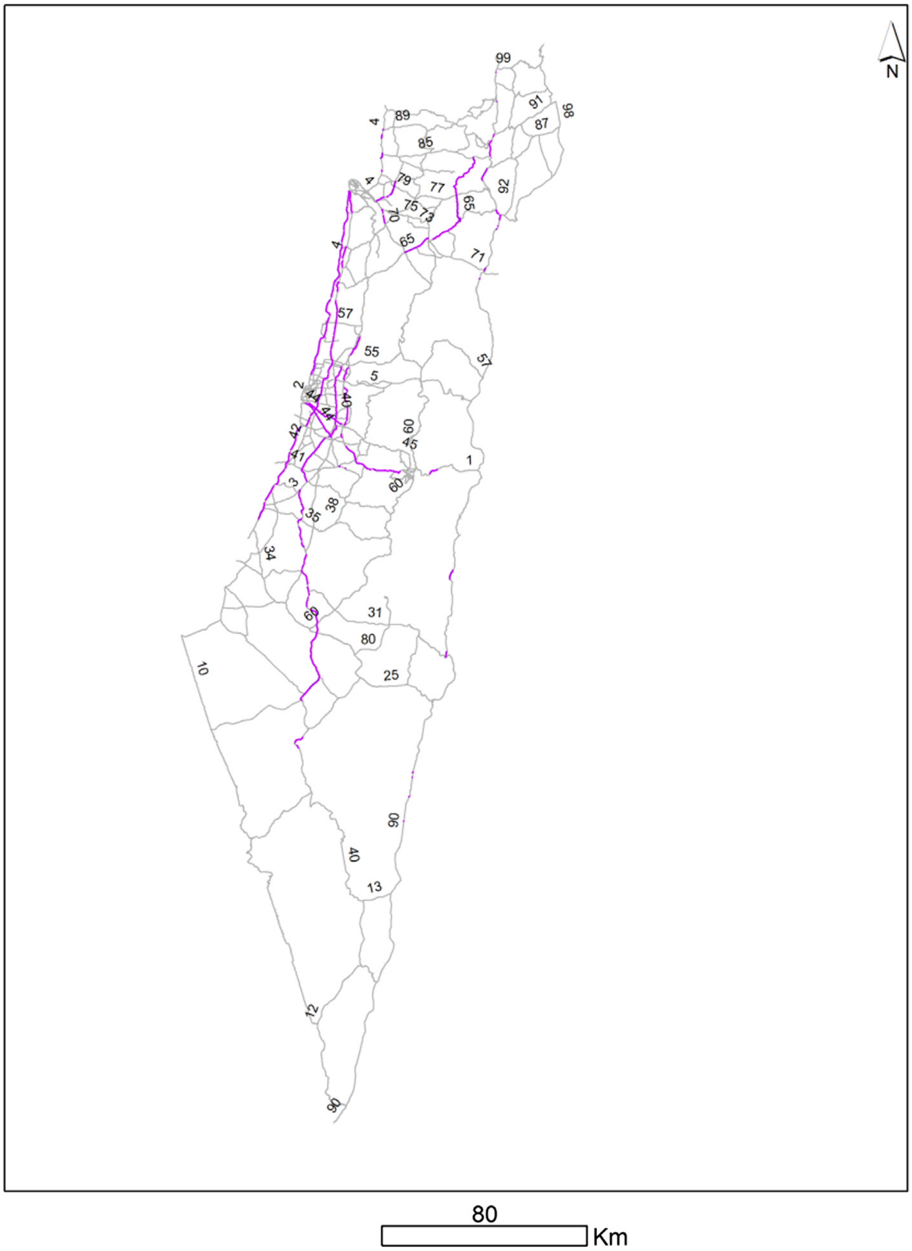



**Fig. 2.** The road segments included in the study.

The dataset generated for the study included 3500 analysis units (segments). Each segment contains the number of driving events, the number of trips by the IVDR-vehicles, and traffic and infrastructure characteristics of the segment. Both the driving events and the infrastructure characteristics belong to the year 2010. In addition, for each unit the total number of road crashes for the years 2008–2010 was produced from the CBS crash files. The crash numbers included in the analysis were: (1) all injury crashes; (2) injury crashes excluding pedestrian crashes; (3) "general with casualties" crashes; (4) total crashes, as a sum of 1 and 3 groups (all injury plus "general with casualties" crashes).

It should be noted here that in Israel two types of crash files are collected by the police: (a) injury crash file, with cases investigated by the police, which serves as a basis for the official crash and injury figures, and (b) "general with casualties" crash file, with cases reported to the police but not investigated. The first file includes all injury severity levels: fatal, serious and slight, where slight injury crashes should satisfy certain selection criteria concerning the time passed since the occurrence and crash participants. The second file includes cases with slight injuries only (not hospitalized), which did not satisfy the selection criteria of the "injury" file. As the amount of records in the second file versus the first one presents a 78% to 22% relation (CBS, 2012), it is customary to conduct crash analyses involving both crash files.

**Table 3**
Category definitions of traffic and road infrastructure variables considered in the study.

| Variable | Category definitions | | | |
|---|---|---|---|---|
| Road type | 1 – single-carriageway road, 2 – dual-carriageway road, 3 – freeway | | | |
| Lane width, m | 1 – standard (3.6 m and over); 2 – narrow (3.3–3.6 m); 3 – very narrow (below 3.3 m) | | | |
| Right shoulder width, m | 1 – wide/standard (2.5 m and over); 2 – medium (1.5–2.5 m); 3 – narrow (1.5 m and below) | | | |
| Left shoulder width[a], m | 1 – wide/standard (2.5 m and over); 2 – medium (1.5–2.5 m); 3 – narrow (1.5 m and below); 0 – not relevant (single-carriageway road) | | | |
| Traffic volume, vehicles per day | Categories by road type | Single-carriageway road | Dual-carriageway road | Freeway |
| | 1 – low | Below 6000 | Below 20,000 | Below 40,000 |
| | 2 – moderate | 6,000–15,000 | 20,000–40,000 | 40,000–80,000 |
| | 3 – heavy | Over 15,000 | Over 40,000 | Over 80,000 |
| Horizontal radius, m | 1 – small | Below 170 | Below 340 | Below 570 |
| | 2 – medium | 170–800 | 340–1400 | 570–2300 |
| | 3 – large | Over 800 | Over 1400 | Over 2300 |
| Vertical grade, % | 0 – none | Below 2% | Below 2% | Below 2% |
| | 1 – low | 2%-5% | 2%-4% | 2%-3% |
| | 2 – moderate | 5%-8% | 4%-6% | 3%-5% |
| Section length (distance between the main junctions)[b], km | 1 – short | Below 2 | Below 4 | Below 8 |
| | 2 – medium | 2–4 | 4–8 | 8–16 |
| | 3 – long | Over 4 | Over 8 | Over 16 |
| Junction ahead | 1 – yes (if the segment edge approaching a junction is within 100, 200 and 900 m from the beginning of junction's zone, for single-carriageway road, dual-carriageway road and freeway, respectively); 0 – no (if the segment edge approaching a junction is within longer distance from the beginning of junction's zone); 2 – the segment includes a junction | | | |
| Junction behind | 1 – yes (if the segment edge distancing from a junction is within 50, 100 and 700 m from the end of the junction's zone, for single-carriageway road, dual-carriageway road and freeway, respectively); 0 – no (if the segment edge after a junction is within longer distance from the end of junction's zone); 2 – the segment includes a junction | | | |
| Change in the number of lanes compared to a previous unit | 0 – no, 1 – yes | | | |
| Change in the lane width compared to a previous unit | 0 – no, 1 – yes (the difference is over 0.3 m) | | | |

[a] Relevant for dual-carriageway roads and freeways, only.
[b] Main junctions are those with left turns and essential traffic volumes on secondary roads – over 1000 vehicles per day.

A preliminary examination of the frequency of IVDR events in the study database revealed that driving events of any type were found on 30.2% of all analysis units. Most events were of two types only: "braking" (type 1 in Table 1) and "speed alert" (type 19 in Table 1), which were observed on 7.1% and 20.6% of the study units, respectively. Hence, the consequent analyses of the relationship between the driving events and the infrastructure characteristics, and the driving events and road crashes focused on three types of events: "braking", "speed alert" and total events. The meaning of the specific events are as follows: "braking" – relates to strong braking while driving on a straight section, without turning of the steering wheel (while for braking in other occasions, with turning of the steering wheel, other driving events were recorded); "speed alert" – refers to driving at speed over 120 km/h as a uniform threshold on all road types (this value was much higher than the speed limits, which, in 2010, were 100, 90 and 80 km/h for freeway, dual- and single-carriageway roads, respectively).

### *2.2. Modeling approach*

The modeling approach proposed in this paper considers the events as an intermediate variable. The idea is to find a relationship between the events and infrastructure characteristics, and then relate the number of events to the number of crashes.

First, the relationship between each infrastructure characteristic and the event counts was examined, given the exposure. The dependent variable (number of events) is a count variable. The event distributions across the study units were skewed, with low means and an over-dispersion. Negative binomial (NB) regression is commonly used for modeling over-dispersed count variables (Cameron and Trivedi, 1986). The NB model can be considered as a generalization of Poisson regression since it has the same mean structure as Poisson regression and an extra parameter to model the over-dispersion. Considering the "excess zero" problem which may necessitate using a zero-inflated model, it was shown (Allison, 2012) that an NB model may fit data with many zeros, as well. Hence, to provide mathematical form of the relations, we adopted NB regression models in this study. To note, other studies of the IVDR events – Farah et al. (2014), Musicant et al. (2014), applied NB distribution for modeling events' count.

NB regression models were fitted to the event counts, separately for the three pre-defined types of events, i.e. "braking", "speed alert" and total events. A multivariate model allowing correlations between various types of events was not tried due to minor numbers of other event types in the database and the lack of correlation between "braking" and "speed alert" events that was found on the stage of descriptive statistics (see Section 2.3). Having examined separate impacts of each infrastructure characteristic, multivariable models were developed for the relationship between the whole set of infrastructure characteristics and driving events, accounting for the exposure. NB regression models were applied, with the log of the expected number of the events as dependent variable, logarithm of exposure as an offset and infrastructure characteristics selected as best predictors. Model explanatory variables were selected by stepwise selection, using Akaike information criterion. (The issue of correlations of the event counts among road segments was explored but found not significant and, hence, was not included in the models). Both the sets of relationships fitted to separate infrastructure characteristic and the multivariable models were considered to conclude as to the impact of road infrastructure characteristics on driving events' occurrences.

Second, the relationship between the driving events and crashes, given the exposure, was explored using NB regression models. Previous analysis demonstrated various forms of relations between the exposure and the events, for different road types (freeways, dual-carriageway and single-carriageway roads). Hence, at this stage, separate models were fitted to particular road types, where accounting for four crash types and three event types considered, in total, 36 models were developed. Multivariate models with simultaneous Modeling of several crash severity levels were not applied as only injury crashes were analyzed in the study, i.e. all injury crashes, injury crashes excluding pedestrian crashes and injury crashes from another police file (see Section 2.1).

At the initial step of each analysis, the form of the relationship between the predicting and predicted variables (e.g. exposure and events) was explored using Generalized Additive Models (GAM) (Wood, 2006). If a non-linear relation was suspected, then the need for a piecewise-linear approximation (Hastie et al., 2009) was considered and two ranges of values of the predicting variable were defined subdivided by a “breaking point” (where the form of the relation may change). The significance of such “breaking points” was further examined in the NB modeling.

Concerning the exposure variable, it can be employed in the NB regression model in two forms: as an offset variable or as a direct explanatory variable. The first option means fixing the coefficient of the log exposure to one, and actually modeling rates. In most cases in this study, we found it safer to let the coefficient of the log exposure to be estimated optimally by the model instead of fixing it to one. Only in the case of multivariable models with wide range of potential predictors (for the events), the exposure variable was used as an offset.

The original models for the relationship between the driving events and crashes, given the exposure, were as follows. If $Y_i$ represents the number of crashes on section $i$, then the model is expressed as:

$$Y_i \sim NB(\lambda_i, k) \tag{1}$$

$$\lambda_i = E(Y_i) = \exp(\beta_0 + X_i\beta) \tag{2}$$

where: $X_i$ is the vector of explanatory variables; $\beta$ is the vector of parameters to be estimated; $\boldsymbol{k}$ is the scale parameter for the NB distribution, i.e. $V(Y_i) = \lambda_i + k\lambda_i^2$.

The model above follows a common methodology of developing regression models for crashes – see Hauer (2015). However, such a model may not be capable of taking into account a possible effect of spatial correlation, and, hence, may inflate type I errors due to over-estimated effective sample size. There is a growing number of road safety studies which examine possible spatial correlation among road segments or study areas, while fitting explanatory models for crashes (Quddus, 2013; Wang and Kockelman, 2013; Barua et al., 2016). In the literature, several methods were developed to correct for the effects of spatial autocorrelation. Dormann et al. (2007), for example, overview the following methods that account for spatial autocorrelation in analyses of spatial data: autocovariate regression, spatial eigenvector mapping, generalized least squares, conditional autoregressive models, simultaneous autoregressive models, generalized linear mixed models and generalized estimation equations (GEE).In this study, the focus of interest was regression coefficients only, thus, originally, we used the GEE approach for our analyses. Furthermore, in order to examine the impact of spatial correlation among the study segments on the results, we re-fitted the original explanatory models. For this, we used SAS 9.4 GLIMMIX procedure (SAS, 2013) to fit a marginal GEE-type model with the same mean function as before, but this time allowing for correlation among the segments. To be specific, suppose that $Y$ represents the (nx1) vector of observed data, $E(Y_i)$ is as in Eqs. (1) and (2) above, but the variance of the observations is$V(Y) = A^{1/2}RA^{1/2}$. The matrix **A** is a diagonal matrix, where A(i,i) $=$ $\lambda_i + k\lambda_i^2$ for an NB distribution and A(i,i)$=\lambda_i$ for a Poisson distribution. The matrix **R** is a covariance matrix.

We used two different covariance structures (i.e. two different **R** options), to analyze the same data. In the 1st structure, we allowed for exponential spatial covariance structure (using *SP(EXP)* option of the SAS GLIMMIX procedure), where the covariance between two observations depends on the distance between the centers of the two segments, and different roads are assumed to be independent. In the 2nd structure, the roads were split into several sub-roads, when consecutive segments were far more than 0.5 km, and we used *AR(1)* autocorrelation model between segments within the same sub-road. (Different sub-roads are assumed to be independent.) To be more specific:

– Using *SP(EXP)* option models, the covariance between two road segments (segments $i$ and $j$), depends on the distance, $d_{ij}$, between their centers, hence R(i,j) equals to $\sigma^2 e^{\frac{-d_{ij}}{\alpha}}$. The GLIMMIX constrains $\alpha$ to be positive. $\alpha = 0$ means zero correlation among all segments.
– Using *AR(1)* option specifies a first-order autoregressive structure, where the covariance between two road segments (segments $i$ and $j$), depends on the number of road sections, the difference, |i-j|, between them, and hence R(i,j) equals to $\sigma^2\rho^{|i-j|}$. $\rho = 0$ means zero correlation among all segments.

Both the original models for the event-crash relationships and the results of their re-fitting while accounting for spatial correlation among the study segments, are presented in Section 3.2.

**Table 4**
Descriptive statistics of the study dataset.[a]

a – Traffic and road infrastructure characteristics

| Characteristic | Distribution of units according to characteristic categories |
|---|---|
| Road type | 32% freeways, 46% dual-carriageway roads, 22% single-carriageway roads |
| Traffic volume | 54% heavy traffic, 31% moderate traffic, 14% low traffic volume |
| Lane width | 90% standard, 7% narrow, 3% very narrow |
| Right shoulder width | 20% narrow, 26% medium, 41% wide shoulder, 13% unknown |
| Left shoulder width | 30% narrow, 21% medium, 19% wide shoulder, 8% unknown (22% not relevant, for single-carriageway roads) |
| Horizontal radius | 84% large, 15% medium, 1% small |
| Vertical grade | 64% none, 28% low, 5% moderate, 3% high |
| Junction ahead | 15% yes, 72% – no, in 13% segment includes a junction |
| Junction behind | 10% yes, 77% – no, in 13% segment includes a junction |
| Section length | 54% short, 28% medium, 6% long sections (in 13% not relevant as segment includes a junction) |
| Change in number of lanes | 5% yes, 95% no |
| Change in lane width | 45% yes, 55% no |

b – Driving events, IVDR vehicle travels and crash statistics

| Event/IVDR travel/crash type | Mean | Standard deviation | Range |
|---|---|---|---|
| "Braking" events | 0.10 | 0.49 | 0–11 |
| "Speed alert "events | 0.31 | 0.75 | 0–12 |
| Total events | 0.51 | 1.27 | 0–30 |
| IVDR vehicle travels | 85 | 102 | 5–908 |
| IVDR vehicles involved | 14 | 11 | 3–79 |
| IVDR drivers involved | 16 | 14 | 1–125 |
| All injury crashes | 1.54 | 4.4 | 0–65 |
| Injury crashes excluding pedestrian crashes | 1.49 | 4.3 | 0–62 |
| "General with casualties" crashes | 6.41 | 35.2 | 0–1112 |
| Total crashes | 7.95 | 38.0 | 0–1130 |

[a] N = 3500 units (road segments), each of 200 m length.

### 2.3. Descriptive statistics

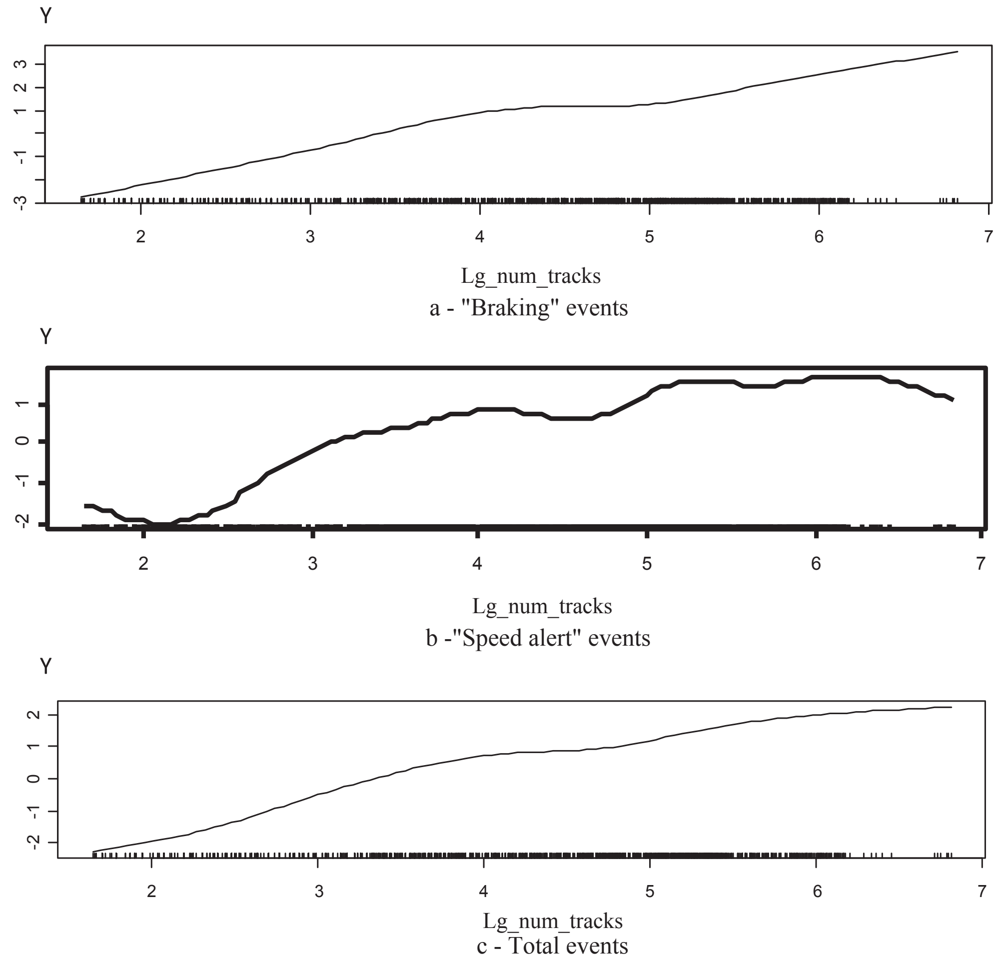
Table 4 provides descriptive statistics of the study dataset that indicates high traffic volumes on a substantial share of the segments, relatively high frequency of junctions and certain presence of geometric restrictions related to horizontal curves, vertical slopes and cross-section deficiencies, e.g. narrow shoulders. Such road and traffic characteristics are known as having effects on crash occurrences (e.g. Elvik et al., 2009; Bonneson and Pratt, 2009), where they may also have an impact on driving task performance and, therefore, on driving events.

The driving events per road segment were not frequent, with an average of 0.1 for "braking", 0.31 for "speed alert" and 0.51 for total events. A Pearson correlation analysis revealed significant positive correlation between "braking" and total events, "speed alert" and total events (with correlation coefficients of 0.55 and 0.57, respectively, $p < 0.001$), but no correlation between the two particular types of the events (Pearson correlation coefficient of $-0.004$, $p = 0.80$). Thus, in consequent analyses, separate models were fitted to each event type. (To note, individual differences between the drivers were not considered by this study but were examined by other studies, e.g. Musicant et al., 2014; Toledo et al., 2014). As a sum of the years 2008–2010, the average number of crashes per study segment was: 1.54 all injury crashes, 1.49 injury crashes excluding pedestrian crashes, 6.4 "general with casualties" crashes and 8.0 total crashes.

### 2.4. Accounting for exposure

The number of travels by IVDR-vehicles, per year, was selected as a natural indicator of exposure of the study road segments to the IVDR-events. It had an average of 85, with the range of 5–908. A log-transformation of the indicator's distribution demonstrated that the main range of the values was 7–446. The GAM component smooth functions fitted to the exposure-events relations, for various event types – Fig. 3, indicated that linear positive relation could be noticed between the log-values of the exposure and the logarithm of the expected number of event counts, of each type. The possibility of a piecewise-linear relation between the exposure and events was tested, but not found to be significant.

Being aware of the importance of a traditional exposure measure – the road traffic volume, for crash prediction (Elvik et al., 2009), a need in the use of this measure for predicting the events was examined. This was done by using two exposure measures: the number of IVDR-vehicle travels and the category of traffic volume (see Table 3), in the explanatory models for driving events. We found that adding the second exposure measure did not increase the explanatory power of the models, where, for each type of the events, the additional variance explained by this variable lied between 0.9% and 2.3%. Thus, the category of traffic volume was not used in the modeling of driving events. However, for modeling crashes, we applied both exposure measures, to account for both the exposure to the IVDR events and to general vehicle traffic.




Notes: *Lg_num_tracks* is logarithm of the number of IVDR-vehicle travels per segment (exposure); *Y* is a component smooth function *s(Lg_num_tracks)* reflecting the form of the exposure-event relation, with log-values of the expected events.

**Fig. 3.** GAM component smooth functions fitted to the exposure-event relations, by driving event type. Notes: *Lg_num_tracks* is logarithm of the number of IVDR-vehicle travels per segment (exposure); *Y* is a component smooth function *s(Lg_num_tracks)* reflecting the form of the exposure-event relation, with log-values of the expected events.

**Table 5**
Explanatory models for the number of driving events, with exposure and *junction ahead* variables.

| | a – "braking" events | | | b – "speed alert" events | | | c – all events | | |
|---|---|---|---|---|---|---|---|---|---|
| Variables | Estimate | z value | Pr(> \|z\|) | Estimate | z value | Pr(> \|z\|) | Estimate | z value | Pr(> \|z\|) |
| Intercept | −6.788 | −17.61 | < 0.001 | −4.408 | −24.80 | < 0.001 | −4.435 | −28.91 | < 0.001 |
| Lg_num_tracks | 0.951 | 11.98 | < 0.001 | 0.804 | 21.21 | < 0.001 | 0.875 | 26.96 | < 0.001 |
| junction ahead – "yes" | 0.463 | 2.59 | 0.010 | −0.564 | −5.42 | < 0.001 | −0.242 | −2.89 | 0.004 |
| junction ahead – "the segment includes a junction" | 1.063 | 5.69 | < 0.001 | −1.388 | −8.05 | < 0.001 | 0.078 | 0.84 | 0.401 |
| $R^2$_dev | 23.6% | | | 24.5% | | | 29.3% | | |

Notes: *Lg_num_tracks* is logarithm of the number of IVDR-vehicle travels per segment; the effects of both categories for *junction ahead* are given related to category "no" (the segment is not approaching a junction).

**Table 6**
Explanatory models for the number of driving events, with exposure and right shoulder width' variables.

| | a – "braking" events | | | b – "speed alert" events | | | c – all events | | |
|---|---|---|---|---|---|---|---|---|---|
| Variables | Estimate | z value | Pr(> \|z\|) | Estimate | z value | Pr(> \|z\|) | Estimate | z value | Pr(> \|z\|) |
| Intercept | −6.647 | −15.74 | < 0.001 | −4.327 | −22.20 | < 0.001 | −4.372 | −25.81 | < 0.001 |
| Lg_num_tracks | 0.935 | 11.01 | < 0.001 | 0.797 | 19.80 | < 0.001 | 0.873 | 25.11 | < 0.001 |
| shoulder width – "medium (1.5–2.5 m)" | 0.253 | 1.31 | 0.190 | −0.178 | −1.97 | 0.049 | −0.016 | −0.21 | 0.837 |
| shoulder width – "narrow (1.5 m and below)" | 0.493 | 2.69 | 0.007 | −0.971 | −8.73 | < 0.001 | −0.284 | −3.45 | 0.001 |
| $R^2$_dev | 20.2% | | | 24.1% | | | 28.9% | | |

Notes: *Lg_num_tracks* is logarithm of the number of IVDR-vehicle travels per segment; the effects of medium and narrow shoulders are given related to category "wide/standard (2.5 m and over)".

## 3. Results

### 3.1. Exploring relations between infrastructure characteristics and events

At this stage, initially, a relationship between each infrastructure characteristic and the event counts, given the exposure, was examined. Separate NB regression models were adjusted to each type of the events, i.e. "braking", "speed alert" and total events. The models were fitted to the main range of the exposure values. Table 5 provides examples of the models that were fitted for the relationship between segment's proximity to a junction – *junction ahead* variable, and driving events' counts. For all event types, a significant impact of junction proximity on the number of events was observed, accounting for the exposure effect. However, while for "braking" events, the situations of close junction or of a segment including a junction are associated with an increase in the events, the same situations bring to a decrease in the "speed alert" events. For total events, the impact of junction proximity was inconsistent, indicating a decrease in the events where the segment is close to junction and an increase, where the segment includes a junction (the latter is not significant).

Table 6 illustrates the models adjusted for a relationship between right shoulder width and driving events' counts. In all the models, the total impact of the shoulder width was significant, yet not always significant for the medium width category. In line with the previous case, an inverse effect of the characteristic was observed on "braking" compared to "speed" events, where medium or narrow shoulders (compared to wide) were associated with more "braking" but fewer "speed alert" events. In addition, narrow shoulders were related to fewer total events that obviously reflects the dominating amount of "speed" events in the total.

Similar explanatory models were fitted to examine the impact on the events' occurrence of other infrastructure characteristics, given the exposure. Table 7 summaries the findings of the models developed. It is evident from Table 7 that most of the infrastructure characteristics, except for the change in the number of lanes between sequential segments, were found to have a significant effect on

**Table 7**
Summary of the models fitted to relations between separate infrastructure characteristics and driving events.

| Infrastructure characteristic examined | | Effect on the number of "braking" events | Effect on the number of "speed alert" events | Effect on the number of all events |
|---|---|---|---|---|
| Junction ahead | Segment close to a junction | Increase | Decrease | Decrease |
| | Junction within a segment | Increase | Decrease | Increase |
| Junction behind | Segment close to a junction | Increase | Decrease | Decrease |
| | Junction within a segment | Increase | Decrease | Increase |
| Section length (distance between main junctions) | Medium-long section | Decrease | Increase | Increase |
| | Junction within a segment | Increase | Decrease | Increase |
| Change in the number of lanes between sequential segments | | – | – | – |
| Change in the lane width between sequential segments (yes) | | Increase | Decrease | – |
| Road type | Dual carriageway road[a] | Increase | Increase | Increase |
| | Freeway[a] | Decrease | Increase | Increase |
| Right shoulder width | Medium or narrow shoulder (compared to wide shoulder) | Increase | Decrease | Decrease |
| Left shoulder width | Wide or medium shoulder on a dual-carriageway road or freeway[a] | Decrease | Increase | Increase |
| Lane width | Narrow lane compared to a standard lane | – | Decrease | Decrease |
| Horizontal radius | Medium or large radius (compared to small radius) | Increase | Increase | Increase |
| Vertical grade | Most grades (compared to ungraded sections) | Decrease | Decrease | Decrease |

Notes: In all cases specified in the table a significant effect of the infrastructure characteristic was found ($p < 0.05$), given the exposure effect. In cases marked with "–" the effect of the infrastructure characteristic, when exposure is present, was not significant.

[a] Compared to a single-carriageway road.

driving events' occurrence. The models showed that the number of “braking” events increases near junctions, where the right shoulder width is below a standard value, where the vehicle travels on a dual-carriageway road (compared to a single-carriageway road), and on sections with larger horizontal radiuses compared to small ones. On the other hand, the number of such events decreases on freeway road sections (compared to a single-carriageway road), on longer road sections (compared to short ones), where a wider left shoulder is present on a divided road or freeway, and where a vertical grade is present (compared to an ungraded road). Some of the findings allegedly look as contradicting to engineering expectations, for instance, concerning the impact of horizontal radius or vertical grade on “braking” events, where larger radius brings to an increase in the events, presence of vertical grades – to a decrease. However, they become reasonable if we recall the meaning of “braking” events, which relate to braking while driving on a stretch section.

According to the results in Table 7, the number of “speed” events increases on a dual-carriageway road or freeway (compared to a single-carriageway road), on a medium or long road section (compared to a short one) and on sections with larger horizontal radiuses compared to small ones. At the same time, the number of “speed” events decreases near junctions, where there is a change in the roadway width, the right shoulder width is below a standard value, lane width is narrow or where a vertical grade is present.

It can be noted that a substantial part of the infrastructure characteristics showed a reverse effect on “braking” events compared to the “speed” events, while this difference seems reasonable due to an “opposite” character of these events. In general, junction proximity or geometry constraints bring, in most cases, to an increased number of “braking” events and to a decreased number of “speed” events. Conversely, improved road conditions and driving on longer sections, without at-grade disruptions, are associated with a decrease in the “braking” events and an increase in the “speed alert” events.

The shaded cells in Table 7 show an inconsistent effect for the impact of junction proximity (*junction ahead* and *junction behind* variables) on the total number of events. The inconsistency is related to a decrease in the events when a segment is close to junction, and an increase in the events when the segment includes a junction. Concerning the effects of other infrastructure characteristics, such as section length, right and left shoulder width, road type, lane width, radius and grade, on the total events, those were similar to the impacts observed for “speed alert” events, evidently reflecting the prevalence of “speed” events among the total driving events.

Following the examination of separate impacts of each infrastructure characteristic, multivariable models were developed for the relationship between the infrastructure characteristics and driving events. NB regression models were adjusted, with a logarithm of events as dependent variable and logarithm of exposure as an offset. Tables 8 and 9 show the multivariable models developed for “braking” and “speed alert” events, respectively.

Among the road characteristics that remained as predictors in the multivariable models, we find: road type, junction proximity, section length, left shoulder width (for both event types) and right shoulder width (for speed events only). On the other hand, lane width and its changes between sequential segments, horizontal radius and vertical grade did not appear among the variables predicting driving events. The impact of lane width, curvature and grade on various types of driving events was found in the univariate analyses of the data (see Table 7), i.e. their impact on the events do exist. However, in the multivariable models the impact of these characteristics is masked by other, stronger variables, which left in the models, where the most essential of them is, probably, road type.

The multivariable models show that when a segment is situated on a longer road section, more “speed” events and fewer “braking” events are expected. On the other hand, for a segment near junction or including a junction, the number of “speed” events decreases and that of “braking” events increases. Higher-standard road types are associated with an increase in “speed” events related to single-carriageway roads, where for “braking” events an increase was found for dual-carriageway roads and a decrease for freeways (as opposed to single-carriageway roads). More narrow right shoulder brings to a decrease in “speed” events. Wider left shoulder leads to an increase in “speed” events. With regard to the impact of the left shoulder width on “braking” events (see Table 8),

**Table 8**
Multivariable explanatory model for “braking” events.[a]

| Variables | Estimate | Std. error | z value | Pr(> \|z\|) |
|---|---|---|---|---|
| Intercept | −6.564 | 0.244 | −26.857 | < 0.001 |
| Junction ahead – “yes” | 0.412 | 0.181 | 2.274 | 0.023 |
| Junction ahead – “the segment includes a junction” | 0.427 | 0.190 | 2.245 | 0.025 |
| Section length – “medium” | −0.532 | 0.269 | −1.980 | 0.048 |
| Section length – “long” | −1.162 | 1.069 | −1.087 | 0.277 |
| Road type – “dual-carriageway” | 0.571 | 0.255 | 2.241 | 0.025 |
| Road type – “freeway” | −1.350 | 0.325 | −4.153 | < 0.001 |
| Left shoulder width – “wide ” | −0.094 | 0.353 | −0.267 | 0.790 |
| Left shoulder width – “medium” | 0.141 | 0.336 | 0.421 | 0.674 |
| Left shoulder width – “narrow” | n/a | | | |
| Road type “dual-carriageway” and left shoulder width “wide ” | −0.298 | 0.424 | −0.703 | 0.482 |
| Road type “dual-carriageway” and left shoulder width “medium” | −0.952 | 0.425 | −2.241 | 0.025 |

Model statistics: Dispersion parameter 0.526 (0.102); Null deviance: 957.3 on 2884 degrees of freedom; Residual deviance: 778.3 on 2874 degrees of freedom; 2 × log-likelihood: −1550.2; AIC: 1574.2; $R^2$_dev = 18.7%.

[a] With logarithm of events as dependent variable, logarithm of the number of IVDR-vehicle travels as an offset. For detailed definitions of categories see Table 3; the effects are given related to other category of each variable such as: “no” for *junction ahead*; “short” for *section length*; “single-carriageway” for *road type*; “not relevant” for *left shoulder width*.

**Table 9**
Multivariable explanatory model for "speed alert" events*

| Variables | Estimate | Std. error | z value | Pr(> \|z\|) |
|---|---|---|---|---|
| Intercept | −9.694 | 1.005 | −9.644 | < 0.001 |
| Junction ahead – "yes" | −0.487 | 0.108 | −4.501 | < 0.001 |
| Junction ahead – "the segment includes a junction" | −0.895 | 0.182 | −4.917 | < 0.001 |
| Junction behind – "yes" | −0.399 | 0.117 | −3.393 | 0.001 |
| Section length – "medium" | 0.422 | 0.088 | 4.811 | < 0.001 |
| Section length – "long" | 1.699 | 0.249 | 6.832 | < 0.001 |
| Right shoulder width – "medium" | 0.004 | 0.089 | 0.044 | 0.965 |
| Right shoulder width – "narrow" | −0.330 | 0.120 | −2.752 | 0.006 |
| Road type – "dual-carriageway" | 3.363 | 1.009 | 3.333 | 0.001 |
| Road type – "freeway" | 3.769 | 1.009 | 3.735 | < 0.001 |
| Left shoulder width – "wide " | 0.722 | 0.122 | 5.904 | < 0.001 |
| Left shoulder width – "medium" | 1.154 | 0.112 | 10.278 | < 0.001 |
| Left shoulder width – "narrow" | n/a | | | |

Model statistics: Dispersion parameter 2.691 (0.519); Null deviance: 2136.6 on 2745 degrees of freedom; Residual deviance: 1600.8 on 2734 degrees of freedom; 2 × log-likelihood: −3318.3; AIC: 3344.3; $R^2$_dev = 25.1%.

* With logarithm of events as dependent variable, logarithm of the number of IVDR-vehicle travels as an offset. For detailed definitions of categories see Table 3; the effects are given related to other category of each variable such as: "no" for *junction ahead* and for *junction behind*; "short" for *section length*; "wide" for *right shoulder width*; "single-carriageway" for *road type*; "not relevant" for *left shoulder width*.

it was not significant for freeways, whereas for dual-carriageways roads a significant decrease in "braking" events was observed for a medium value (1.5–2.5 m) and a non-significant decrease for a wide value (over 2.5 m). In line with previous findings, the infrastructure characteristics that remained in the multivariable models generally demonstrate a reverse effect on two event types.

The results show clearly that higher road types, better road conditions and driving on longer road sections, without at-grade junctions, are associated with an increase in the "speed alert" events and a decrease in the "braking" events. Conversely, junction proximity and road layout constraints increase the number of "braking" events and decrease the number of "speed" events.

### 3.2. Exploring relations between driving events and crashes

The second analysis was devoted to the relationship between the driving events and crashes. Since for both variables, travel exposure is a leading predictor of occurrences, their relationship was examined given an exposure. The latter was presented by two characteristics: the number of yearly IVDR-vehicle travels through the segment (as in the analysis of relations between road characteristics and events) and the level of daily traffic volume on the road section (as is common in road safety models). Initially, the form of the relation between the number of IVDR-vehicle travels and crashes was tested using the GAM-models – Fig. 4. They indicated a positive relation between the two variables, for various crash types, with a possible "breaking point" near 4.5 log-value of the exposure (90 travels). Thus, the possibility of a change in the exposure-crash relation at the level of 90 travels was allowed in the consequent fitting of mathematical forms of the models. Similarly, event-crash relations were considered for various types of both variables, using the GAM-models; they indicated a possibility of "breaking point" in some relationships but at a very low number of the events (mostly, around 1), thus, reducing the usability of such a "breaking point". Therefore, the models were fitted allowing a linear relation between the events and crashes.

Regression models were adjusted to the data, with an NB distribution for crash numbers and a log-link function between the predictor variables and the resulting one. Previous analyses pointed out the impact of road type on the event occurrence. Hence, separate models were fitted at this stage to three road types: freeways, dual-carriageway and single-carriageway roads, where road type reflects various road conditions. Accounting for four crash types, three event types and three road types, in total, 36 models were developed. (To avoid "error" results in processing logarithms, before a log-transformation of the event counts, 0.5 was added to each raw number; and a few observations with very high crash numbers were excluded from the dataset.)

Tables 10 and 11 present examples of the models fitted to injury crashes on various road types, using "braking" and "speed alert" events. In all the models, a positive impact of both indicators of exposure was found; yet, for freeways, the impact of the number of IVDR-vehicle travels was not significant but the impact of level of the traffic volume was. In most cases, the "breaking point" in the exposure-crash relation was not significant for the occurrence of injury crashes, except for dual-carriageway roads, where for higher level of the IVDR travels (over 90), a decrease in the crash rate related to exposure is expected (in the model for "braking" events).

For "braking" events, a significant positive impact on injury crashes, given the exposure, was found on single-carriageway and dual-carriageway roads and non-significant impact on freeways. The effect of the "speed" events on injury crashes was negative on all road types, significant for dual-carriageway roads and freeways and close to significant for single-carriageway roads. The results of the models are as follows: an increase in crashes with more "braking" events (for two road types) and a consistent decrease with more "speed" events (for all road types), look reasonable as "braking" events generally reflect the presence of constraints and changes in road infrastructure conditions while "speed" events point to the possibility of undisturbed traffic flow, under better road conditions. For freeways, the effect of "braking" events was not ascertained as such driving events, due to their nature, should be infrequent for this road type.To examine the impact of spatial correlation on the event-crash relationship the original explanatory models were re-

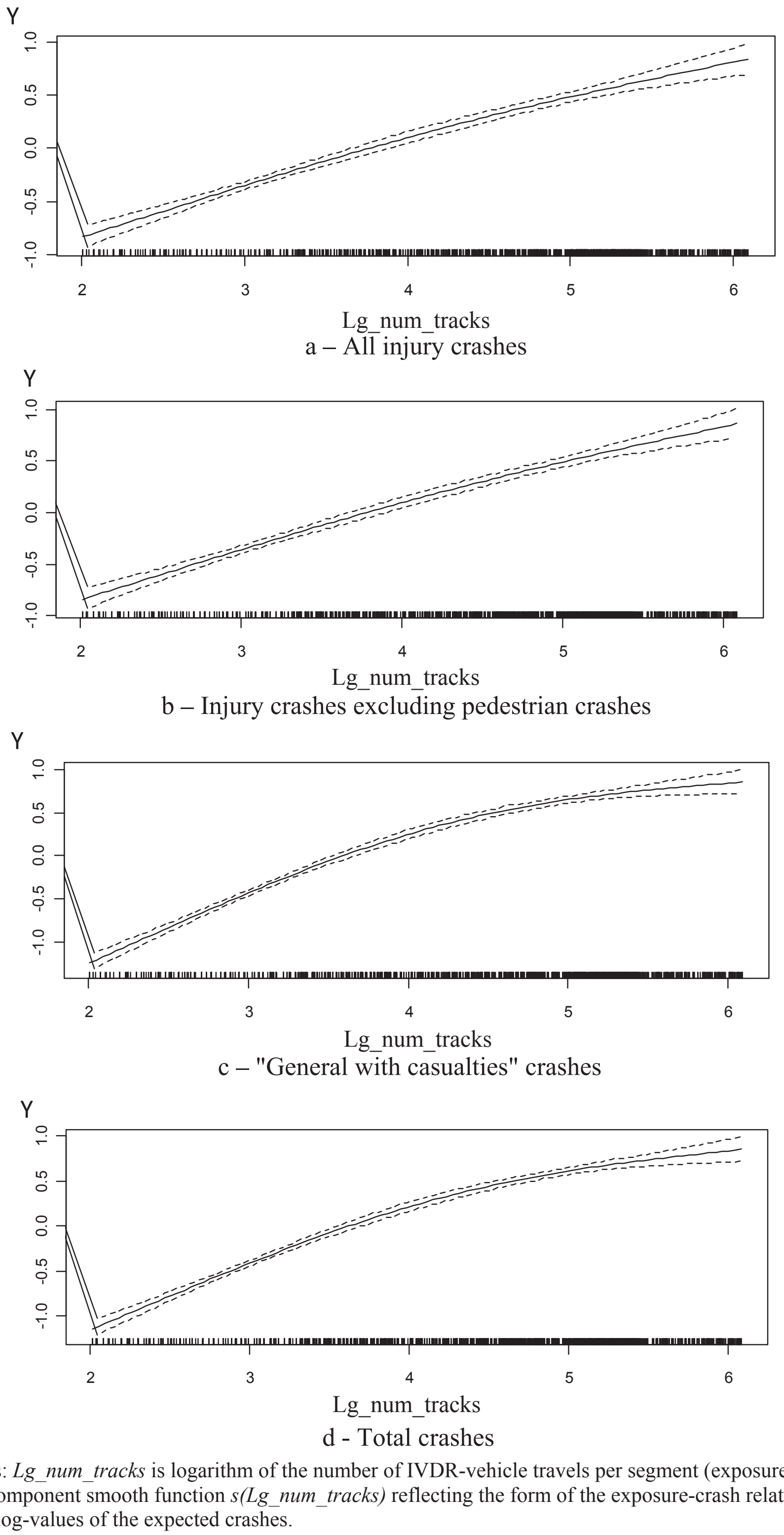



**Fig. 4.** GAM component smooth functions fitted to the exposure-crash relations, by crash type, with 95% confidence intervals. Notes: ***Lg_num_tracks*** is logarithm of the number of IVDR-vehicle travels per segment (exposure); ***Y*** is a component smooth function ***s(Lg_num_tracks)*** reflecting the form of the exposure-crash relation, with log-values of the expected crashes.

**Table 10**
Explanatory models for number of injury crashes, on various road types, with exposure and "braking" events.

| | Single-carriageway road | | | Dual-carriageway road | | | Freeway | | |
|---|---|---|---|---|---|---|---|---|---|
| Variables | Estimate | z value | Pr(> \|z\|) | Estimate | z value | Pr(> \|z\|) | Estimate | z value | Pr(> \|z\|) |
| Intercept | −2.72 | −3.18 | 0.002 | −0.99 | −4.55 | < 0.001 | −1.00 | −1.93 | 0.054 |
| Lgn | 0.45 | 4.67 | < 0.001 | 0.27 | 5.21 | < 0.001 | 0.05 | 0.34 | 0.731 |
| I((Lgn – brk) * (Lgn – brk > 0)) | −0.30 | −0.66 | 0.509 | −0.40 | −2.81 | 0.005 | 0.26 | 1.18 | 0.239 |
| Lj | 1.26 | 5.03 | < 0.001 | 0.49 | 5.89 | < 0.001 | −0.07 | −0.43 | 0.667 |
| Traffic volume –"moderate" | 1.15 | 1.37 | 0.170 | 0.64 | 4.76 | < 0.001 | 0.64 | 2.81 | 0.005 |
| Traffic volume –"heavy" | 2.28 | 2.73 | 0.006 | 0.81 | 6.25 | < 0.001 | 1.51 | 6.20 | < 0.001 |
| $R^2$_dev | 26.6% | | | 6.7% | | | 13.3% | | |

Notes: *Lgn* is a logarithm of the number of IVDR-vehicle travels per segment; *brk* is a "breaking point" in exposure-crash relation, equal to 4.5 (90 travels); *Lj* is a logarithm of the event count. The effects of *traffic volume* are given related to "low" category (see Table 3 for definitions). The models were fitted using 590, 1463 and 1077 observations (segments) for single-carriageway roads, dual-carriageway roads and freeways, respectively.

**Table 11**
Explanatory models for number of injury crashes, on various road types, with exposure and "speed alert" events.

| | Single-carriageway road | | | Dual carriageway road | | | Freeway | | |
|---|---|---|---|---|---|---|---|---|---|
| Variables | Estimate | z value | Pr(> \|z\|) | Estimate | z value | Pr(> \|z\|) | Estimate | z value | Pr(> \|z\|) |
| Intercept | −3.52 | −4.27 | < 0.001 | −1.85 | −8.85 | < 0.001 | −1.61 | −3.09 | 0.002 |
| Lgn | 0.46 | 4.73 | < 0.001 | 0.34 | 6.55 | < 0.001 | 0.23 | 1.43 | 0.154 |
| I((Lgn − brk) * (Lgn − brk > 0)) | 0.74 | 1.73 | 0.085 | −0.20 | −1.42 | 0.157 | 0.10 | 0.45 | 0.651 |
| Lj | −0.56 | −1.62 | 0.106 | −0.72 | −7.21 | < 0.001 | −0.38 | −6.29 | < 0.001 |
| Traffic volume –"moderate" | 0.87 | 1.14 | 0.256 | 0.55 | 4.05 | < 0.001 | 0.50 | 2.13 | 0.034 |
| Traffic volume –"heavy" | 1.76 | 2.30 | 0.022 | 0.72 | 5.45 | < 0.001 | 1.31 | 5.26 | < 0.001 |
| $R^2$_dev | 23.9% | | | 7.4% | | | 15.3% | | |

Notes: *Lgn* is a logarithm of the number of IVDR-vehicle travels per segment; *brk* is a "breaking point" in exposure-crash relation, equal to 4.5 (90 travels); *Lj* is a logarithm of the event count. The effects of *traffic volume* are given related to "low" category (see Table 3 for definitions). The models were fitted using 591, 1462 and 1080 observations (segments) for single-carriageway roads, dual-carriageway roads and freeways, respectively.

**Table 12**
Results of re-fitting models for injury crashes (see Tables 10 and 11) accounting for spatial correlation among the study segments.

| Type of model: type of events, type of road | Re-fitting of original models in SAS without spatial parameters[a]: significance of the events' impact ($L_j$) | *AR(1)* estimate (p-value) | *SP(EXP)* estimate |
|---|---|---|---|
| "braking" events, single-carriageway roads | Direction 1 (N = 233) sig.Direction 2 | 0.030 (0.82) | 0 |
| | (N = 357) ns | Did not converge | 0 |
| "braking" events, dual-carriageway roads | Direction 1 (N = 738) sig.Direction 2 | Did not converge | 0 |
| | (N = 725) sig. | −0.017 (0.68) | 0 |
| "braking" events, freeways[b] | Direction 1 (N = 579) nsDirection 2 | −0.027 (0.55) | 0 |
| | (N = 498) ns | −0.011 (0.82) | 0 |
| "speed alert" events, single-carriageway roads[c] | Direction 1 (N = 233) ns | 0.497 (0.0017) | Did not converge |
| | Direction 2 (N = 358) ns | Did not converge | 0 |
| "speed alert" events, dual-carriageway roads | Direction 1 (N = 739) sig. | Did not converge | 0 |
| | Direction 2 (N = 723) sig. | −0.022 (0.60) | 0 |
| "speed alert" events, freeways | Direction 1 (N = 580) sig. | −0.027 (0.56) | 0 |
| | Direction 2 (N = 500) sig. | −0.042 (0.43) | 0 |

[a] The original models were developed in *R software*. To examine the impact of spatial dependency, the original models were re-fitted in SAS (2013), for both directions of travel on the road.
[b] $L_j$ (events' impact) was not significant in the original model.
[c] In re-fitting the model in SAS, the p-value of $L_j$ (events' impact) changed from 0.106 to 0.336 (not significant).

fitted as explained in Section 2.2. Table 12 summarizes the results of re-fitting models for injury crashes (the original models presented in Tables 10 and 11) while accounting for spatial correlation. We found that for the 1st data covariance structure – *SP(EXP)* estimates, $\alpha$ was equal to 0, for all models except for one, for which the model run did not converge. When $\alpha$ equals to 0, there is no change in the original model estimated parameters.In the analyses with the 2nd data covariance structure – *AR(1)* estimates, either the model did not converge, or $\rho$ was found to be highly non-significant (p-values of 0.43–0.82), while the re-fitted model parameters related to the events' impact ($L_j$) were very close to the original models, both in their values and significance. In one case (a model with "speed alert" events on single-carriageway roads) $\rho$ was found to be significant. However, the events' impact ($L_j$) was not significance both in the re-fitted model with spatial correlation and in the original model without spatial correlation, meaning that

**Table 13**
Summary of the models fitted to the event-crash relations, on different road types, given the exposure: the coefficients relating events to crashes and their significance level.

| Driving event type | Road type | Crash type | | | |
|---|---|---|---|---|---|
| | | Injury crashes | Injury crashes excluding pedestrian crashes | "General with casualties" crashes | Total crashes: injury + "general with casualties" crashes |
| "Braking" | Single-carriageway | 1.26[##] | 1.14[##] | 1.28[##] | 1.29[##] |
| | Dual-carriageway | 0.49[##] | 0.45[##] | 0.015 | 0.14[*] |
| | Freeway | −0.07 | −0.14 | 0.003 | −0.01 |
| "Speed alert" | Single-carriageway | −0.56[**] | −0.74[*] | −0.16 | −0.27 |
| | Dual-carriageway | −0.72[##] | −0.68[##] | −0.65[##] | −0.66[##] |
| | Freeway | −0.38[##] | −0.39[##] | −0.33[##] | −0.33[##] |
| Total events | Single-carriageway | 1.24[##] | 1.05[##] | 1.71[##] | 1.49[##] |
| | Dual-carriageway | 0.20[##] | 0.18[#] | −0.02 | 0.02 |
| | Freeway | −0.34[##] | −0.34[##] | −0.29[##] | −0.29[##] |

[##] Significant with $p < 0.001$.
[#] Significant with $p < 0.01$.
[*] Significant with $p < 0.05$.
[**] Close to significant ($p = 0.106$). The rest of the coefficients are insignificant.

the addition of spatial dependency did not change the model findings. To summarize, we found that after controlling for the explanatory variables, introducing spatial correlations did not change any conclusion about the events' impact ($L_j$) on crashes, hence, we did not continue to pursue spatial correlation in Modeling event-crash relationships in this study.

Table 13 summarizes the findings of all the models fitted for the event-crash relationship, while accounting for the exposure (measured by two values as introduced above) and considering the road type (the original models, without spatial correlations, are presented). For each case, the coefficients relating events to crashes in the model and their significance level, are detailed. One can note that in a substantial part of the cases a significant relation between the events and crashes was found. However, the form of this relation differs for various road types. The findings across various crash types reveal that:

- For single-carriageway roads, a consistent positive and significant relation was identified between the "braking" events and crashes, as well as between the total events and crashes, of all crash types. As to the "speed alert" events, a negative relation with crashes was observed that was significant (or close to significant) for two categories of injury crashes. For "braking" and total events, the strength of the relation between the driving events and crashes was essential (high coefficient values), indicating a potential of the IVDR-events for predicting crashes on single-carriageway roads.
- For dual-carriageway roads, a consistent negative relation was identified between the "speed alert" events and all crash types. In addition, a significant positive relation was found between the "braking" events and crashes, for both injury crash types and total crashes. In most cases, the strength of the relation was relatively high, indicating the explanatory potential of both types of the events for predicting crashes on dual-carriageway roads. As to the relation between the total events and crashes, a significant positive impact was found for injury crash types but not very strong, and no impact for other crashes.

For both road types, a positive relation between the "braking" events and crashes and a negative relation between the "speed" events and crashes, apparently reflect the impact of road conditions on driver behavior and crash occurrence, where an uninterrupted travel under better road conditions may invite higher speeds but not be associated with higher crash counts, and conversely, worse road conditions and more interruptions create more "braking" events and lead to more crashes. These results can be seen in line with the design consistency concept (Lamm et al., 1999), which assumes that when the road design is consistent and uniform, fewer driver errors are expected ("braking" events, in our case) and safer driving can be attained (i.e. fewer crashes). On the other hand, many studies showed that better road design conditions are associated with higher travel speeds and actually with better safety records (e.g. Edquist et al., 2009; Ivan et al., 2009).

- For freeway road segments, across all the crash types, no significant relation was found between the "braking" events and crashes, showing that this type of events has a negligible potential for predicting crashes on this road type. On the other hand, a consistent pattern of relations, across all the crash types, was found between the "speed alert" events and crashes, demonstrating a decrease in the number of crashes with the events' increase (yet the strength of the relation is lower than for other road types). As to the total events, their crash-relation patterns actually repeat those observed for the "speed alert" events demonstrating a negative event-crash link, for all the crash types considered.

The result on a negative relation between the number of "speed alert" events and the number of crashes on freeways is similar to that found for dual-carriageway roads, where it probably reflects the impact of better road conditions. The latter may enable higher speeds while it is not associated with higher crash counts.

## 4. Discussion and conclusions

In line with the growing interest in naturalistic driving studies (Uchida et al., 2010; Antin et al., 2011; Regan et al., 2012; Hallmark et al., 2013), this paper examined driver behavior data collected by an IVDR-technology. A basic assumption behind this approach, is that behavior data collected automatically may assist in identification of potentially dangerous driving events, which are related to crash occurrences, where reducing the risk factors related to the events will contribute to a reduction in crashes.

Previous studies of the IVDR-based data (Toledo et al., 2008; Prato et al., 2010; Farah et al., 2013, 2014) demonstrated usability of the driving events, recognized by this technology, for identification and treatment of driver-related risk factors. Unlike previous research, this study focused on examination of the IVDR-based driving events aiming to recognize infrastructure-related risk factors. Based on the geographically matched data from the IVDR records, road infrastructure characteristics' database of the interurban roads and crashes, and applying statistical models, this study explored two topics: the relationship between the occurrence of driving events and road infrastructure characteristics, accounting for traffic exposure, and the relationship between the recorded driving events and road crashes. As common analysis units, short road segments were applied on which the IVDR data (the number of travels and driving events), road infrastructure characteristics and crashes were matched, for a similar period.

This study aspired to contribute to general knowledge on driving events' occurrence in the context of road infrastructure characteristics, and on the relationship between driving events and crashes, under various road conditions. Once such relations are unveiled, driving events may serve for identification of potential dangerous locations on the road network and providing a basis for preventive road safety work. The idea of using driving events for identification of high-risk locations on the road network was also raised by other studies (Pande et al., 2014, Wolshon and Parr, 2015).

Naturalistic driving studies are led by an underlying assumption that surrogate safety events are intended to serve as substitutes to crash risk measures, and that components of this relationship should be supported by empirical evidence (Shankar et al., 2008; Guo et al., 2010; Hallmark et al., 2011). Statistical relations between the driving events and crashes were examined in this study actually aiming to explore the applicability of the IVDR-events as crash-substitutes. One can admit that fitting a model for a relationship between driving events and crashes does not imply that driving events explain or cause crashes. However, statistical correlations need to be demonstrated, otherwise, the findings concerning the event occurrences are not transferable on crashes.

In general, this study demonstrated a significant relationship between the driving event counts and road infrastructure characteristics at the event locations, on the interurban roads in Israel. Based on the models fitted, certain relations between the driving events and crashes were shown, given the exposure and road conditions. In all the models developed, an exposure measure in terms of the number of yearly IVDR-vehicle trips per road segment and/or the level of daily vehicle traffic, played an essential role in predicting the driving events and crash occurrences, thus, supporting the need in exposure-based measures for performing NDS data analyses (Shankar et al., 2008; Jovanis et al., 2011). Both driving event counts and crashes were found to be directly related to the exposure, where the impact of infrastructure characteristics on the events and of the events on crashes was shown given the exposure effect.

The relationship between the driving events and infrastructure characteristics and, more generally, a triple relationship between crashes, infrastructure characteristics and events was not explicitly examined in the past (Regan et al., 2012). The current research showed that components of this relationship can be ascertained but the whole picture is complex. According to the findings, various event types have different relation to the infrastructure characteristics and a different effect on crashes.

Most infrastructure characteristics considered by the study such as indicators of junction proximity, length of road sections, change in the road width, road type, right and left shoulder width, lane width, horizontal radius, vertical grade, were found to have a significant effect, at least on some event types. According to the multivariable models fitted by the study, for the driving events examined, among the most influential road characteristics were road type, junction proximity, section length, and the widths of road shoulders. However, a substantial part of the infrastructure characteristics showed an inverse effect on “braking” events compared to the “speed” events. Typically, junction proximity or geometry constraints bring to an increased number of “braking” events and to a decreased number of “speed” events. Conversely, improved road conditions and driving on longer sections, without at-grade junctions, are associated with a decrease in the “braking” events and an increase in the “speed alert” events. Such differences in the impact seem reasonable due to an “opposite” character of these events.

The effects of many infrastructure characteristics on the total events was similar to those observed for the “speed alert” events, where an increase in the events was observed on longer road sections, without geometric restraints and at-grade junctions, and on higher-level roads (freeway or dual-carriageway road compared to single-carriageway road). Hence, a surprising finding of the first analysis was that, contrary to the expectations, an increase in the “speed alert” and in the total number of events indicates the presence of improved road conditions and, therefore, will not necessarily be associated with higher crash rates.

Considering an event-crash relationship we found that, under certain road conditions, driving event counts can contribute to the prediction of crash occurrences. For single-carriageway roads, better explanatory potential for predicting both injury and total crashes was found for “braking” events and the total events, which were positively related to crashes. For dual-carriageway roads, for predicting various crash types, the “speed alert” events are more suited, where those are negatively related to crashes. In addition, “braking” events can be applied for predicting injury crashes on this road type, for which a positive relation to crashes is expected. For freeway road segments, the “speed alert” and total events are applicable for predicting crashes, where a negative event-crash relation is expected. These results are in line with the engineering judgement as more “braking” events are associated with worse road conditions and more interrupted travel that may lead to more crashes, whereas more “speed” events actually reflect better road conditions, with no junction proximity and non-interrupted travel, that is generally associated with lower crash rates. They can also be seen in line with the design consistency concept (Lamm et al., 1999), and with findings of studies that examined a relationship

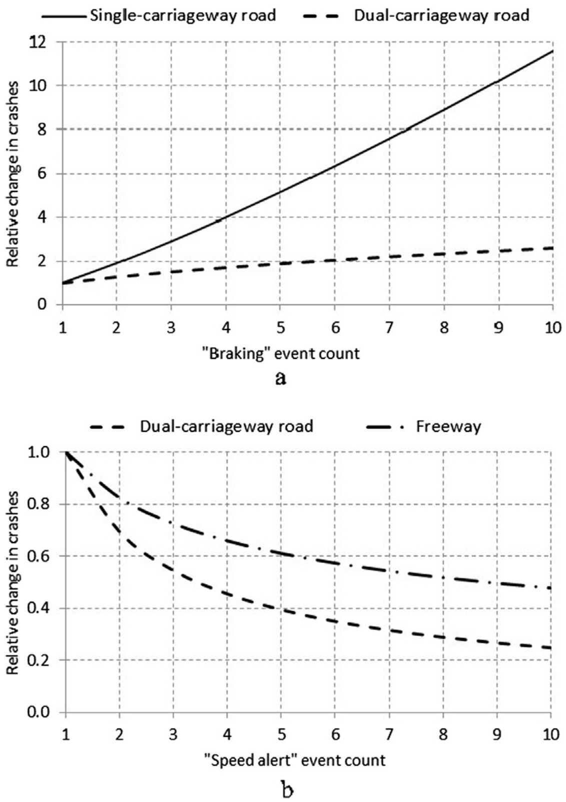



**Fig. 5.** Relative change in injury crashes associated with increase in the event counts, based on the study models: (a) for "braking" events, (b) for "speed alert" events.

between road design conditions, travel speed and crashes (Edquist et al., 2009; Ivan et al., 2009).

Fig. 5 provides a visual presentation of relations between the event counts and a relative change in injury crashes, for road types for which a significant association was found between the two variables (see Tables 10 and 11). As evident, an increase in "braking" events is associated with a moderate increase of crashes on dual-carriageway roads and with a substantial increase – on single-carriageway roads. Conversely, an increase in "speed" events is associated with a moderate reduction of crashes on freeways and with a stronger reduction – on dual-carriageway roads. An exponential link between the events and crashes can be noted, while controlling for both exposure measures (the IVDR-travels and the level of general traffic volume on the road). Based on Table 13, similar relations can be shown for other crash types.

The study showed significant associations between the IVDR driving events and crashes, depending on road type. However, negative relations between the "speed alert" events and crashes were found for all road types, together with an earlier observation that more "speed" events actually indicate the presence of better road conditions. Such results may be viewed as contradicting to the expectations that surrogate crash measures produced by the NDS should characterize near-crash situations or reflect crash prevention boundaries (Songchitruksa and Tarko, 2006; Shankar et al., 2008; Hallmark et al., 2011), i.e. assuming a direct relation between the surrogate measures and the crash likelihood. In this sense, the driving events defined based on the IVDR records may not satisfy the definitions of ideal surrogate crash measures. In particular, the driving events produced by the IVDR technology considered in the current study (the "green box") do not always indicate an exceptional event that leads to crashes. The definition of "speed alert" events seems to be too rough in order to enable detailed analyses. At the same time, this study results demonstrated that the IVDR events may assist in recognizing road conditions associated with higher crash risk and that a relation between the amount of events and crashes is significant. In addition, one cannot ignore empirical evidence from the previous studies that driving events were positively correlated with drivers' crash records (Toledo and Lotan, 2007; Toledo et al., 2008).

Summing up, we believe that the IVDR events are suitable for some applications but may not fit others and, thus, their usability for specific purposes should be examined by empirical studies. The results reported by this study should be interpreted in terms of the driving events applied; the results can be different if, for example, a different definition of braking or speed events is used. While recording driving events by means of advanced technologies, more attention should be given to the events' definitions, especially, if those are intended to be used for analyses based on the data matched with external sources. In general, the IVDR data can be treated as less accurate than the NDS data because NDS studies also collect video-data, which can be used for a verification at any time frame and are systematically applied for the event classification (Dingus et al., 2016). At the same time, the IVDR-data are easily accessible and less resource consuming and, thus, can be collected on wider vehicle fleets than the NDS data (Toledo et al., 2014; Paefgen et al., 2014).

The study brought new insights into the relationship between the IVDR-based driving events, road infrastructure characteristics and crashes, demonstrating significant effects and consistent forms of the relations, under certain conditions. Some components of the relationship were ascertained, where selected event types were proven as suitable for indicating better or worse road infrastructure conditions and for predicting crashes, on various road types. However, we cannot conclude yet that the events explored are good predictors for road crashes and that they are applicable as crash substitutes for an identification of high-risk locations on the road network. The heterogeneity of findings of this study indicates that further research of the above relationship is needed for a better understanding of the interactions between the "triangle" components. In the future research, a particular focus should be on the definitions of driving events produced by the IVDR or other technologies, that will contribute both to understanding of the relations found and to a transferability of results.

## Acknowledgments

Appreciation is extended to the National Road Infrastructure Company of Israel – Netivey Israel which commissioned this research and to Or Yarok Association which provided the IVDR dataset for the analysis.

## References

Allison, P., 2012. Do We Really Need Zero-inflated Models? < http://statisticalhorizons.com/zero-inflated-models > .
Antin, J., Lee, S., Hankey, J., Dingus, T., 2011. Design of the in-vehicle driving behavior and crash risk study. Transportation Research Board: Washington DC, USA.
Barua, S., El-Basyouny, K., Islam, M., 2016. Multivariate random parameters collision count data models with spatial heterogeneity. Anal. Method. Accid. Res. 9, 1–15.
Blanchard, R.A., Myers, A.M., Porter, M.M., 2010. Correspondence between self-reported and objective measures of driving exposure and patterns in older drivers. Accid. Anal. Prevent. 42, 523–529.
Bonneson, J.A., Pratt, M.P., 2009. Roadway safety design workbook. Report no. FHWA/TX-09/0-4703-P2. Texas Transportation Institute and Federal Highway Administration.
Cameron, A.C., Trivedi, P.K., 1986. Econometric models based on count data: comparisons and applications of some estimators and tests. J. Appl. Econ. 1, 29–53.
Cannon, B.R., Sudweeks, J., 2011. Geospatial analysis of high-crash intersections and rural roads using naturalistic driving data. Final Report 11-UT-013, the National Surface Transportation Safety Center for Excellence. Virginia Tech Transportation Institute.
CBS, 2012. Road accidents with casualties. General summaries: Part A. Publication No 1492. Central Bureau of Statistics (CBS), Jerusalem.
Dingus, T.A., Guo, F., Lee, S., Antin, J.F., Perez, M., Buchanan-King, M., Hankey, J., 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data. Proceedings of the National Academy of Sciences of the United States of America 113 (10), 2636–2641.
Dormann, C.F., McPherson, J.M., Araujo, M.B., et al., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609–628.
Edquist, J., Rudin-Brown, C.M., Lenne, M., 2009. Road design factors and their interaction with speed and speed limits. Monash University, Accident Research Centre, Victoria, Australia.
Elvik, R., Hoya, A., Vaa, T., and Sorensen, M. (2009). The Handbook of Road Safety Measures. 2nd ed. Emerald.
Farah, H., Musicant, O., Shimshoni, Y., Toledo, T., Grimberg, E., Omer, H., Lotan, T., 2013. The first year of driving - Can an In-Vehicle Data recorder and parental involvement make it safer? Transport. Res. Record: J. Transport. Res. Board 2327, 26–33.
Farah, H., Musicant, O., Shimshoni, Y., Toledo, T., Grimberg, E., Omer, H., Lotan, T., 2014. Can providing feedback on driving behavior and training on parental vigilant care affect male teen drivers and their parents? Accid. Anal. Prevent. 69 (2014), 62–70.
Farmer, C.M., Kirley, B.B., McCartt, A.T., 2010. Effects of in-vehicle monitoring on the driving behavior of teenagers. J. Saf. Res. 41 (1), 39–45.
Fleishman, H., Mulchadsky, I. Raz, O., 2003. System and method for vehicle driver behavior analysis and evaluation. United States of America Patent US 7,389,178 B2.
Gitelman, V., Hakkert, S., Doveh, E., 2015. Establishing a safety management system for the non-urban road network – models developed and future needs. In: Proceedings of IRF Europe & Central Asia Regional Congress, Istanbul, Turkey. TS 5.5, Paper 132.
Guo, F., Klauer, S.G., Hankey, J.M., Dingus, T.A., 2010. Near crashes as crash surrogate for naturalistic driving studies. Transport. Res. Rec.: J. Transport. Res. Board 2147, 66–74.
Hallmark, S., Hsu, Y.-Y., Boyle, L., Carriquiry, A., Tian, Y., and Mudgal, A., 2011. Evaluation of Data Needs, Crash Surrogates, and Analysis Methods to Address Lane Departure Research Questions Using Naturalistic Driving Study Data. SHRP 2 Report S2-S01E-RW-1, National Academy of Sciences, USA.
Hallmark, S., McGehee, D., Bauer, K.M., Hutton, J.M., Davis, G.A. et al., 2013. Initial Analyses from the SHRP 2 Naturalistic Driving Study. Addressing Driver Performance and Behavior in Traffic Safety. SHRP 2 Safety Project S08, National Academy of Sciences, USA.
Hanowski, R.J., Medina, A.L., Wierwille, W.W., Lee, S.E., 2004. Incident clustering. Diagnostic approach for assessing usability of intersections and other road sites. Transport. Res. Rec.: J. Transport. Res. Board 1897, 173–179.
Hanowski, R.J., Hickman, J.S., Wierwille, W.W., Keisler, A., 2007. A descriptive analysis of light vehicle-heavy vehicle interactions using in situ driving data. Accid. Anal. Prevent. 39, 169–179.
Hanowski, R.J., Hickman, J.S., Olson, R.L., Bocanegra, J., 2009. Evaluating the 2003 revised hours-of-service regulations for truck drivers: the impact of time-on-task on critical incident risk. Accid. Anal. Prevent. 41, 268–275.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, 2nd ed. Springer-Verlag.
Hauer, E., 2015. The art of regression modelling in road safety. Springer, Heidelberg.
Heinrich, H.W., 1959. Industrial Accident Prevention. McGraw-Hill, New York.
Hyden, C., 1987. The Development of a Method for Traffic Safety Evaluation: the Swedish Traffic Conflicts Technique. Department of Traffic Planning and Engineering, Lund University, Sweden.
Ivan, J.N., Garrick, N.W., Hanson, G., 2009. Designing roads that guide drivers to choose safer speeds. Joint Highway Research Advisory Council of the University of Connecticut and the Connecticut Department of Transportation. Report No. JHR 09-321.
Jovanis, P.P., Aguero-Valverde, J., Wu, K.F., Shankar, V., 2011. Analysis of Naturalistic Driving Event Data. Transport. Res. Record: J. the Transport. Res. Board 2236 (1), 49–57.
Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., Ramsey, D.J., 2006. The impact of driver inattention on near-crash/crash risk: an analysis using the 100-car naturalistic driving study data. Report No. DOT HS 810 594, NHTSA: Washington, D.C.
Koppel, S., Charlton, J., Kopinathan, C., Taranto, D., 2011. Are child occupants a significant source of driving distraction? Accid. Anal. Prevent. 43 (3), 1236–1244.
Lamm, R., Psarianos, B., Mailaender, T., 1999. Highway Design and Traffic Safety Engineering Handbook. Mc-Graw Hill.
Lee, S.E., Olsen, E.C.B., Wierwille, W.W., 2004. A comprehensive examination of naturalistic lane-changes (Report No. DOT HS 809 702). Washington, DC: National Highway Traffic Safety Administration (NHTSA).
Lee, S.E., Simons-Morton, B.G., Klauer, S.E., Ouimet, M.C., Dingus, T.A., 2011. Naturalistic assessment of novice teenage crash experience. Acc. Anal. Prevent. 43 (4), 1472–1479.
Lotan, T., Albert, G., Ben-Bassat, T., Ganor, D., Grimberg, E., Musicant, O., Hakkert, S. and Toledo T., 2010. Potential benefits of in-vehicle systems for understanding driver behavior; a series of small-scale ND studies in Israel. PROLOGUE Deliverable D3.2. Or Yarok, Hod Hasharon, Israel.

McGehee, D.V., Raby, M., Carney, C., Lee, G.D., Reyes, M.L., 2007. Extending parental mentoring using an event-triggered video intervention in rural teen drivers. J. Saf. Res. 38 (2), 215–227.

Musicant, O., Bar-Gera, H., Schechtman, E., 2014. Temporal perspective on individual driver behavior using electronic records of undesirable events. Accid. Anal. Prevent. 70 (2014), 55–64.

Olson, R.L., Hanowski, R.J., Hickman, J.S., Bocanegra, J., 2009. Driver distraction in commercial vehicle operations. Report No: FMCSA-RRR-09-042. U.S. Department of Transportation, Federal Motor Carrier Safety Administration. Washington DC: FMCSA.

Paefgen, J., Staake, T., Fleisch, E., 2014. Multivariate exposure modeling of accident risk: insights from Pay-as-you-drive insurance data. Transport. Res. A 61, 27–40.

Pande, A., Loy, J., Dixit, V., Spansel, K., Wolshon, B., and Kent, J.D., 2014. Exploration of naturalistic driving data for identifying high crash risk highway locations. Transportation Research Board 93rd Annual Meeting Compendium of Papers, #14-5291, Washington DC.

Perkins, S.R. and Harris, J.I., 1968. Traffic Conflict Characteristics – Accident Potential at Intersections. Highway Research Record, 225, 35-43. Highway Research Board, National Research Council, Washington DC.

Prato, C.G., Toledo, T., Lotan, T., Taubman-Ben-Ari, O., 2010. Modelling the behavior of novice young drivers during the first year after licensure. Accid. Anal. Prevent. 42 (2), 480–486.

Quddus, M., 2013. Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS. J. Transport. Saf. Secur. 5 (1), 27–45.

Regan, M.A., Williamson, A., Grzebieta, R., Tao, L., 2012. Naturalistic Driving Studies: Literature Review and Planning for the Australian NDS. Proc. Australasian College of Road Safety Conference – A Safe System: Expanding the reach, Sydney.

Retting, R.A., Greene, M.A., 1997. Influence of traffic signal timing on red-light running and potential vehicle conflicts at urban intersections. Transport. Res. Record: J. Transport. Res. Board 1595, 1–7.

Sayed, T., Zein, S., 1999. Traffic conflict standards for intersections. Transport. Plan. Technol. 22, 309–323.

SAS, 2013. SAS/STAT® 13.1 User's Guide. Cary, NC: SAS Institute Inc.

Sayer, J., Devonshire, J., Flannagan, C., 2007. Naturalistic driving performance during secondary tasks. In: Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, 9–12 July, Stevenson, Washington.

Shankar, V., Jovanis, P.P., Aguero-Valverde, J., Gross, F., 2008. Analysis of Naturalistic Driving Data: Prospective View on Methodological Paradigms. Transportation Research Record: Journal of the Transportation Research Board, No. 2061, Transportation Research Board of the National Academies, Washington, D.C., 1–8.

Shi, Q., Abdel-Aty, M., 2015. Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. Transport. Res. C: Emerging Technol. 58, 380–394.

Shiftan, Y., Toledo, G., 2015. Can Feedback from In-vehicle Data Recorders Improve Driver Behavior and Reduce Fuel Consumption? Transportation Research Board 94th Annual Meeting Compendium of Papers, Paper #15-2181.

Songchitruksa, P., Tarko, A., 2006. The extreme value theory approach to safety estimation. Accid. Anal. Prevent. 38, 811–822.

Toledo, T., Lotan, T., 2007. In-vehicle data recorder for evaluation of driving behavior and safety. Transport. Res. Rec.: J. Transport. Res. Board 1953, 112–119.

Toledo, T., Musicant, O., Lotan, T., 2008. In-vehicle data recorders for monitoring and feedback on drivers' behavior. Transport. Res. Part C: Emerg. Technol. 16, 320–331.

Toledo, T., Farah, H., Morik, S., Lotan, T., 2014. Driving exposure of Israeli young male drivers within a graduated driver licensing system. Transport. Res. Part F 26 (2014), 180–189.

Transportation Research Board [TRB], 2014. < http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Public/Pages/The_SHRP_2_Naturalistic_Driving%20Study_472.aspx > (accessed 20.12.14)

Treat, J.R., Tumbas, N.S., McDonald, S.T., Shinar, D., Hume, R.D., Mayer, R.E., Stansifer, R.L., Castellan, N.J., 1979. Tri-level study of the causes of traffic accidents final report. Report DOT HS-034-3-535, National Highway Traffic Safety Administration, U.S. Department of Transportation.

Twisk, D.A.M., van Nes, N., Haupt, J., 2012. Understanding safety critical interactions between bicycles and motor vehicles in Europe by means of Naturalistic Driving techniques. Proceedings, International Cycling Safety Conference 2012, 7–8 November 2012, Helmond, The Netherlands.

Uchida, N., Kawakoshi, M., Tagawa, T., Mochida, T., 2010. An investigation of factors contributing to major crash types in Japan based on naturalistic driving data. IATSS Res. 34 (1), 22–30.

Wang, Y., Kockelman, K., 2013. A poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accid. Anal. Prevent. 60, 71–81.

Wolshon, B., Parr, S.A., 2015. Identifying high-risk roadways for infrastructure investment using naturalistic driving data. Final research report. Maritime Transportation Research & Education Center, University of Arkansas.

Wood, S.N., 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.

Wu, K., Jovanis, P.P., 2012. The relationship between crash and crash-surrogate events in naturalistic driving data. Accid. Anal. Prevent. 45, 507–516.

Wu, K., Jovanis, P.P., 2013. Defining and screening crash-surrogate events using naturalistic driving data. Accid. Anal. Prevent. 61, 10–22.

Wu, K.-F., Aguero-Valverde, J., Jovanis, P.P., 2014. Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. Accid. Anal. Prevent. 72, 210–218.