Received: 1 April 2022 | Revised: 24 August 2022 | Accepted: 12 September 2022

DOI: 10.1049/itr2.12281

IET Intelligent Transport Systems

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Developing speed-related safety performance indicators from floating car data

Jiří Ambros[1] | Davide Shingo Usami[2] | Veronika Valentová[1]

[1]CDV – Transport Research Centre, Líšeňská 33a, Brno, Czech Republic

[2]Centre for Transport and Logistics, Sapienza University of Rome, Rome, Italy

Correspondence
Jiří Ambros, CDV – Transport Research Centre, Líšeňská 33a, 636 00 Brno, Czech Republic.Email: jiri.ambros@cdv.cz

Funding information
Ministerstvo Dopravy

Abstract
In the road traffic safety domain there is a need for using proactive (non-crash-based) indicators, known as safety performance indicators (SPIs). Traffic speed based on big data (floating car data [FCD]) could help develop network-wide SPIs, but related knowledge and experience are insufficient so far. The authors attempted to fill this gap by using nationwide Italian FCD to develop speed-related SPIs and validating their relationship to crashes to see their potential explanatory value. The authors calculated the coefficient of variance (CV), congestion index (CI), and the number of incidents as candidate SPIs. For validation, the authors used linear correlation, crash frequency model, and ranking consistency. Incidents turned out to be the best SPI, especially for motorways.

1 | INTRODUCTION

From a global perspective, the level of road safety is unacceptable. According to the latest overview [1], road crashes bring 1.35 million deaths each year and present the first leading cause of death for children and young adults (5–29 years of age). The long-term targets, which aimed at a 50% reduction in the number of deaths, have not yet been achieved [1]. Specifically in the European Union, the trend is positively downward; however, in recent years, it has stagnated and currently every year 25,000 people die on EU roads and more than 135,000 are seriously injured [2, 3, 4].

It is generally accepted that safety plans and targets need to be monitored periodically to verify the progress made and to adopt necessary changes based on recent trends observed [5, 6]. The number and severity of traffic crashes have been used as traditional indicators–however, these present only 'the tip of the iceberg' since they occur as the worst case of unsafe operational conditions of the road traffic system [7]. This has motivated the development of so-called 'safety performance indicators' (SPIs), originally defined as any measurement that is causally related to crashes or injuries and used to monitor safety performance or understand the process that leads to crashes [8]. Later in the frame of the SafetyNet project, the following purposes of SPIs were summarized [9]:

- to reflect the current safety conditions of a road traffic system,
- to measure the influence of various safety interventions,
- to compare different countries, regions etc.

The European Transport Safety Council (ETSC) [8] report introduced a set of seven SPIs focused on: alcohol and drugs, speed, protective systems, daytime running lights, vehicles, roads, and trauma management. In the current EU Road Safety Policy Framework 2021 to 2030, the set increased to eight SPIs (labelled as 'key performance indicators' [KPIs]): speed, safety belts, child restraint systems, protective equipment, driving under the influence of alcohol, driver distraction by handheld devices, vehicle safety, infrastructure safety, post-crash care [4].

In parallel, there have been numerous studies using 'surrogate measures of safety' (SMoS), which may serve as an alternative or complement to traditional (crash-based) safety analyses. Most often SMoS are used to assess the safety of road infrastructure (for an overview see, e.g., [10, 11, 12]). In addition, SMoS data collection and analysis methods have recently employed various advanced sensing technologies [13], such as automated video analysis, Light Detection and Ranging (LiDAR), or floating car data (FCD). Nevertheless, these techniques are usually limited to individual sites; the exception is FCD, which is big data collected from dedicated sensors embedded in-vehicle probes travelling in the road network and is thus relatively unlimited in time and space [14].

SMoS have also been used for studying driving behaviour, assessing driving styles or driver profiling, and using naturalistic driving in instrumented vehicles or with on-board devices. The obtained information may then be utilized in various usage-based insurance (pay-as-you-drive/pay-how-you-drive) schemes. More information is available, for example, in

reviews of Horrey et al. [15], Tselentis et al. [16], or Singh and Kathuria [17].

Especially, speed is interesting from the perspective of SMoS since there is a well-documented relationship between speed and safety. For example, the Power Model [18] related the effects of mean speed changes to the number of crashes of different severity and was validated in various road environments and on various levels [19–21]. In addition, speed may be used to further calculate derivatives such as acceleration or jerk. There are a number of approaches to speed estimation [22]; speed can also be conveniently obtained from FCD, either from global positioning system (GPS) or accelerometer data. Using FCD to obtain speed data has several advantages over traditional measurement techniques, such as roadside traffic counters or inductive loops. FCD enables network-wide data collection, as well as the availability of historical data, which is an ideal source for before-after studies.

We decided to focus on speed-related indicators, which can be conveniently obtained from FCD and used to assess the safety of countries or regions similarly to SPIs. More details on FCD-based and speed-related SMoS were recently summarized by Ambros et al. [23].

Our objective was to develop speed-related SPIs from FCD and explore their validity. The subsequent text introduces a Literature review, followed by Data and methods, Results, and Discussion and conclusions.

2 | LITERATURE REVIEW

In principle, both SMoS and SPIs/KPIs share similar motivations and definitions. The difference between them is that they have been applied on different spatial levels: SMoS on the micro-level (road sections or intersections), and SPIs/KPIs on the macro-level (countries and regions).

When developing an alternative indicator, its reliability and validity should be ensured. In the context of FCD-based speed, reliability means the relationship to the 'ground truth', that is, speed obtained from traditional sensors. Validity may be interpreted as the relationship to crashes as a traditional safety indicator; as noted by Tarko et al. [12], to be useful for transportation safety applications, there must be a possibility for non-crash events (i.e. SMoS/SPIs) to be converted into a corresponding crash frequency and/or severity.

The following paragraphs present examples of previous studies, which used FCD to obtain speed-related indicators, and included some form of reliability and/or validity test.

Firstly, several studies focused on speed and safety on selected roads, for example:

- Pei et al. [24] used GPS data from 480 taxis and assigned it to 112 road segments in Hong Kong. Their analysis confirmed that average speed plays a significant role in crash risk.
- Wang et al. [25, 26] used FCD from 15,000+ taxis travelling on urban arterials in Shanghai to derive speed and speed variation. Using statistical models, both variables were shown to be positively associated with crashes: an increase in speed was associated with an increase in total crashes, and larger speed variation was also associated with increased crash frequency.
- Using a database of FCD from private cars, buses, and trucks, Gitelman et al. [27] developed models for 179 road sections in Israel, which linked speed, injury crash data, traffic volume, and road infrastructure characteristics. The models demonstrated a positive relationship between mean speeds and crashes while controlling for traffic and road characteristics.
- In a series of Canadian studies [28–31], several SMoS were derived from smartphone GPS data collected in Quebec City, Montreal, and Ottawa. The measures, including braking, congestion, average speed, and speed variation, were incorporated into statistical models and found related to crash frequency and severity.
- Based on FCD from HERE, National Performance Management Research Dataset (NPMRDS) was made available in the USA and used in several studies focusing on speed–safety relationship: for example, Banihashemi et al. [32] in the Washington State, Ederer et al. [33] on Georgia State Route 6, Das et al. [34] in Ohio and Washington states.
- In Italy, the research on speeds from FCD has not been focused on safety. For example, De Fabritiis et al. [35] developed and tested two algorithms to perform travel speed predictions for a specific road link by directly using current and near-past average FCD travel speeds. Fusco et al. [36] also proposed a method for short-term speed predictions on large networks based on raw FCD. Only one paper was found where drivers' speeds from FCD are compared to theoretical safety speeds determined from road geometry [37].

Secondly, several studies collected FCD on a wider scale (across a country). While they tested reliability (comparability of FCD-based speed to fixed loops), they did not consider validity (relationship to crashes)

- Bekhor et al. [38] estimated speed from FCD on approx. 1600 road sections in Israel and compared it with speeds obtained from loops. The authors reported a 'relatively good fit between speeds measured by the two methods, with a downward bias for the GPS speeds'.
- In a network of 500 rural road locations, Diependaele et al. [39] obtained speed from FCD. When compared with loop data, FCD speed was on average almost 10 km/h higher.
- Jurewicz et al. [40] determined speed from probe data on 235 locations in the Australian state of Victoria. The obtained FCD speed was lower than the speed measured by loops.

More examples of comparative studies are provided in a review by Ambros et al. [23]: interestingly, some found that FCD-based speeds were higher than loop-based speeds, and some found an opposite tendency. This may stem from differences between time mean speed and space mean speed [38], as well as from the complexity of estimation of free-flow speed (FFS) from FCD and uncertain representativeness of the FCD fleet to a driving population [41]. This emphasizes the necessity of testing the validity of FCD-based speed indicators;

as indicated by Gitelman et al. [27] and Jurewicz et al. [40], using FCD data may lead to the development of a new generation of Power Models.

In sum, the first group of speed-safety studies adopted a detailed approach on a sample of sites, usually cities or US states; their goal was not to represent the whole country, as is the idea behind SPI. On the other hand, the studies in the second group were closer to SPI principles; however, they focused only on speed reliability and did not test the relationship between speed and safety.

Based on this overview of the state-of-the-art, we attempted to fill the identified gap by using nationwide FCD to develop speed-related indicators and validating them against crashes to see their potential value as SPIs.

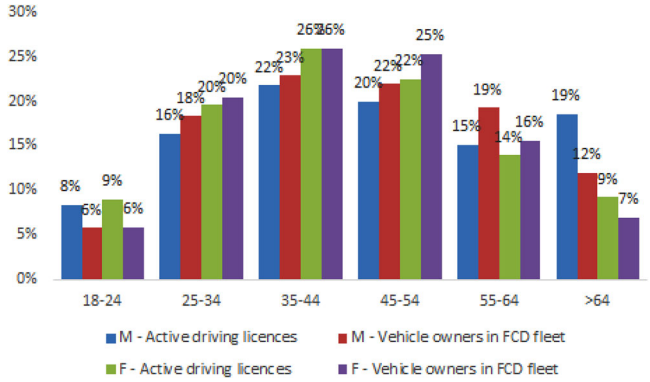


FIGURE 1 Active driving licenses as of 10 January 2011 by age group and gender in Italy and the FCD dataset.

3 | DATA AND METHODS

3.1 | Data description

The primary data source was FCD from OCTO Telematics. This big data contains both personal and freight vehicles, and private and commercial vehicles. Each monitored vehicle is equipped with a black box that records GPS measurements (position, heading, speed, quality) which are periodically transmitted to a data processing centre every 3 min. The data processing centre matches the received data to the road network to estimate link travel speeds. According to De Fabritiis et al. [35], the accuracy of the provided estimates is over 90% on the entire road network, compared with measures coming from Automatic Vehicle Identification (AVI).

We selected all FCD from March 2010. The total number of vehicles with at least one trip recorded was about 700,000 vehicles, representing about 2% of the car fleet. The penetration rate in 2008 was about 1.7% of the Italian privately owned car fleet and is projected to rise to 3% by the end of 2009 according to De Fabritiis et al. [35].

To assess the representativeness of the sample, we compared the distribution of monitored vehicles by the 110 Italian provinces in the FCD dataset with one of the national vehicle fleet obtained from the Italian National Institute of Statistics (Istat), excluding motorcycles, buses, and trailers. The correlation coefficient was 0.94. In terms of percentage points, the maximum difference found was 2.5%, while in 80% of provinces the difference was below 0.55%. Based on these findings, we assume that the sample represents the geographical distribution of vehicles quite well.

We also compared the percentage distribution of Italian driving licenses as of 9 January 2011 (the data was not available for 2010) by gender and age (data source: Italian Ministry of Infrastructure and Transport) with the corresponding distribution in the FCD dataset (see Figure 1). It has to be noted that in this dataset the information available refers to the owner of the vehicle and that for about 19% of the cases gender and age group are not known because it is a company vehicle.

The correlation coefficient between the distributions by age group is 0.93. Considering the distributions concerning the female gender, the correlation is even higher (0.98), while for the male gender it is lower (0.81). The latter value is mainly due to the difference for the over 64 age group, which accounts for 19% of licenses in the national driving population and 12% in the FCD dataset. When assessing representativeness, however, it should be borne in mind that the use of a car by drivers in this age group (> 64) is lower, so the FCD sample is likely to be even more representative in terms of mileage.

From available data, we selected the following variables:

- average speed (km/h),
- speed variance,
- number of incidents (defined as events with serious average impact where acceleration/deceleration values remain, for a sufficient period, equal to or greater than 2 g),
- vehicle-kilometers (VKM),

Average speed $\bar{V}$ and speed variance $VAR(V)$ were calculated as follows: let $d_{i,p,r,t}$ be the distance travelled by vehicle i in province p on road type r during time interval t, and let $t_{i,p,r,t}$ be the time spent by vehicle i to cover the distance $d_{i,p,r,t}$.

$$\bar{V} = \frac{\sum_{i=1}^{N} d_{i,p,r,t}}{\sum_{i=1}^{N} t_{i,p,r,t}}$$

$$VAR\,(V) = \sqrt{\frac{\sum_{i=1}^{N} {d_{i,p,r,t}}^2 / t_{i,p,r,t}}{\sum_{i=1}^{N} t_{i,p,r,t}} - \bar{V}^2}$$

The secondary source was crash data from the CARE database, which hosts detailed data on individual road crashes resulting in death or injury in the European Member States. Data are standardized using a Common Accident Data Set (CaDaS; [42]). Severity levels were distinguished as fatally injured (FAT) or injured (INJ). Since fatalities had relatively low frequency (on average 6 % FAT vs 94 % INJ), we combined both categories (FAT+INJ).

TABLE 1 Descriptive characteristics of data

		Min.	Max.	Mean	Std. Dev.
Speed (km/h)	Urban	20.270	48.987	30.259	4.408
	Motorway	64.293	122.442	105.393	9.568
	Rural	32.205	73.572	48.294	6.691
Speed variance (CV)	Urban	0.148	0.235	0.187	0.019
	Motorway	0.000	0.083	0.046	0.024
	Rural	0.077	0.154	0.110	0.013
Congestion index (CI)	Urban	0.001	0.039	0.016	0.006
	Motorway	0.000	0.038	0.002	0.007
	Rural	0.001	0.041	0.015	0.007
Incidents	Urban	1	294	22.654	41.861
	Motorway	1	16	2.786	3.403
	Rural	0	86	7.818	11.614
Vehicle-kilometres (VKM)	Urban	$1.858 \times 10^{+08}$	$3.331 \times 10^{+10}$	$3.081 \times 10^{+09}$	$4.616 \times 10^{+09}$
	Motorway	$3.465 \times 10^{+07}$	$1.814 \times 10^{+10}$	$1.874 \times 10^{+09}$	$2.503 \times 10^{+09}$
	Rural	$2.974 \times 10^{+08}$	$1.966 \times 10^{+10}$	$3.291 \times 10^{+09}$	$3.243 \times 10^{+09}$
Crashes	Urban	2	1652	118.700	218.978
	Motorway	1	144	11.808	19.821
	Rural	4	175	27.741	23.125

We did not obtain raw data, but several aggregations. From possible alternatives, we selected a combination aggregated by

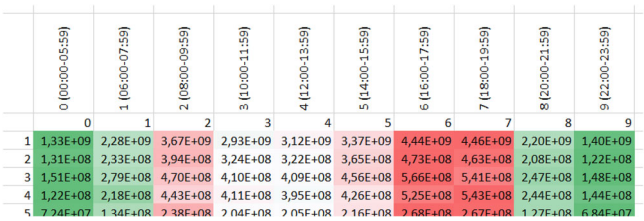
- provinces
- road types
  - urban roads
  - motorways
  - rural roads
- 10 time intervals (night defined between midnight to 6, day split into nine 2-h intervals)
  - 0 (00:00–05:59)
  - 1 (06:00–07:59)
  - 8 (20:00–21:59)
  - 9 (22:00–23:59)

Descriptive characteristics of data, aggregated by provinces, are listed in Table 1. In addition to the mentioned variables (speed, incidents, vehicle kilometres, crashes) it contains also additional variables (speed variance and congestion index) which are described in the following subsection.

3.2 | Methods

Based on the state-of-the-art and available data we anticipated a link between the three elements

- potential SPIs
  - speed
  - number of incidents
- number of crashes



	0 (00:00-05:59)	1 (06:00-07:59)	2 (08:00-09:59)	3 (10:00-11:59)	4 (12:00-13:59)	5 (14:00-15:59)	6 (16:00-17:59)	7 (18:00-19:59)	8 (20:00-21:59)	9 (22:00-23:59)
	0	1	2	3	4	5	6	7	8	9
1	1,33E+09	2,28E+09	3,67E+09	2,93E+09	3,12E+09	3,37E+09	4,44E+09	4,46E+09	2,20E+09	1,40E+09
2	1,31E+08	2,33E+08	3,94E+08	3,24E+08	3,22E+08	3,65E+08	4,73E+08	4,63E+08	2,08E+08	1,22E+08
3	1,51E+08	2,79E+08	4,70E+08	4,10E+08	4,09E+08	4,56E+08	5,66E+08	5,41E+08	2,47E+08	1,48E+08
4	1,22E+08	2,18E+08	4,43E+08	4,11E+08	3,95E+08	4,26E+08	5,25E+08	5,43E+08	2,44E+08	1,44E+08
5	7,24E+07	1,34E+08	2,38E+08	2,04E+08	2,05E+08	2,16E+08	2,68E+08	2,67E+08	1,27E+08	6,84E+07

FIGURE 2 Illustration of visual identification of peak hours from VKM sums as time intervals 2, 6, 7.

Instead of using average speed and/or speed variance, we decided to use the coefficient of variance (CV). This standardized speed indicator is computed as standard deviation (i.e. square root of speed variance) divided by the mean (i.e. average speed).

$$CV = \frac{standard\ deviation}{mean} = \frac{\sqrt{speed\ variance}}{average\ speed}$$

Following Stipancic et al. [28–31], we added the congestion index (CI), which expresses the relationship between FFS and average speed. Using MS Excel Conditional Formatting of VKM sums (see Figure 2) we visually defined peak hours as time intervals 2, 6, 7 (i.e. 8–10 and 16–20 o'clock). Then FFS was defined as speed from off-peak hours. Average speed relates to all-time intervals together, that is, both peak and off-peak hours.

$$CI = \begin{cases} \frac{FFS - average\ speed}{FFS} & if\ FFS > average\ speed \\ 0 & otherwise \end{cases}$$

FIGURE 3 Expected link between speed, incidents, and crashes

FIGURE 4 Scheme of combinations used in calculation of correlation

CV and CI are indicators already tested in other studies that showed a good relationship with the frequency and severity of accidents in the off-peak period [28]. These indicators (CV and CI) were calculated at the level of the provincial road network. They were also calculated for different time intervals and three road categories. Since they are normalized measures of speed, they allowed comparisons both between different road categories and time intervals and with results obtained in other contexts.

The expected link between the elements is visualized in a causal chain (Figure 3).

Should there be a statistical relationship between the elements, the causal chain will hold, that is, there will be a valid link. In this case, speed and incidents indicators may be used as SPIs.

According to an overview of SMoS validation studies [11], there are several different approaches to validation, from which we chose the following two: linear correlation and the crash frequency model. Next, we added ranking consistency, which is similar to the 'method consistency test' as used with hotspot identification methods [43, 44-46].

Validation 1: Linear correlation

We proved by the Shapiro–Wilk test that the samples did not follow a normal distribution ($p = 0.000 < 0.05$). Therefore, we used the Spearman correlation coefficient, which is recommended for non-normal data. We calculated the correlation coefficients in three pairs, as indicated in Figure 4. Originally, we planned to use aggregations by province, road type, and time intervals; however, splitting by time intervals turned out to result in relatively small samples. Therefore, we used only aggregation by province and road type.

Validation 2: Crash frequency model

The model was built in IBM SPSS with

- number of crashes as a response variable,
- speed, incidents, and VKM as potential explanatory variables.

As mentioned above, the input samples were not normally distributed. Thus following safety modelling state-of-the-art [47–50], we developed generalized linear models (GLMs), assuming negative binomial error distribution and logarithmic link function (i.e. VKM was added as a natural logarithm).

Our goal was to add explanatory variables as long as they:

1. have a statistically significant influence on at least 90% level ($p \leq 0.10$),
2. have a logical sign (VKM, incidents, speed should be associated with crash increase).

Validation 3: Ranking consistency

Additionally, we deemed it practical to enable ranking the provinces based on their safety level. This means we needed to ensure that ranking based on potential SPIs is consistent with ranking based on crash numbers. The idea of the test is that a list of provinces, which were identified as 'unsafe' (i.e. on the top of the list ranked in the descending order) by the tested method, should be similar to the list of segments identified by the standard method. Then the tested method, which identifies segments with the largest overlap with the list of segments by the standard method, is more consistent.

The results of the validation are reported in the following section.

4 | RESULTS

Validation 1: Linear correlation

Spearman correlation coefficient was calculated in IBM SPSS for the correlation pairs indicated in Figure 3. Table 2 reports the results for the road types, together with achieved levels of statistical significance p (in italics). We adopted the threshold level $p \leq 0.10$; the values under this threshold (with two exceptions) are listed in bold.

In sum, the highest correlation coefficients (approx. 0.6 and more) were found between incidents and crashes. The remaining pairs (speed + incidents, speed + crashes) had coefficients in the range 0.2 to 0.6, with the lowest values on rural roads and the highest values on motorways. It is also interesting to note that in urban areas, congestion seems to have a higher correlation with incidents than speed variability, whereas the opposite is true on rural roads.

The relatively low values may not be too surprising. For example, Stipancic et al. [31] reported Spearman correlation coefficients between speed and crashes, of which none exceeded 0.6.

Validation 2: Crash frequency model

Parameters of built crash frequency models are listed in Table 3:

TABLE 2 Spearman correlation coefficients and achieved levels of statistical significance (in italics)

Urban	Incidents	Crashes
Speed CV	**0.22**	−0.10
	0.022	*0.310*
Speed CI	**0.29**	**0.30**
	0.002	*0.002*
Incidents		**0.65**
		0.000
Motorway	**Incidents**	**Crashes**
Speed CV	**0.65**	**0.34**
	0.000	*0.003*
Speed CI	**0.64**	**0.35**
	0.000	*0.002*
Incidents		**0.81**
		0.000
Rural	**Incidents**	**Crashes**
Speed CV	**0.27**	−0.16
	0.004	*0.107*
Speed CI	**0.19**	**0.17**
	0.046	*0.083*
Incidents		**0.59**
		0.000

TABLE 3 Parameters of crash frequency models

		B	SE	*p*	R^2
Urban	(Intercept)	−10.615	1.219	0.000	**0.56**
	ln VKM	0.695	0.059	0.000	
	Speed CI	18.972	9.849	0.054	
	(ODP)	0.296	0.045		
Motorway	(Intercept)	−14.926	2.881	0.000	**0.73**
	ln VKM	0.815	0.133	0.000	
	Speed CI	19.482	12.700	0.125	
	(ODP)	0.238	0.082		
	(Intercept)	−13.833	2.349	0.000	**0.76**
	ln VKM	0.654	0.116	0.000	
	Speed CV	39.446	8.647	0.000	
	(ODP)	0.129	0.049		
	(Intercept)	−8.767	3.879	0.024	**0.83**
	ln VKM	0.520	0.182	0.004	
	Incidents	0.112	0.044	0.011	
	(ODP)	0.199	0.071		
Rural	(Intercept)	−10.251	1.050	0.000	**0.57**
	ln VKM	0.616	0.049	0.000	
	Speed CI	9.732	6.119	0.112	
	(ODP)	0.146	0.025		

FIGURE 5 Graphs of ranking consistency

- B (SE) … standardized regression coefficient (plus its standard error)
- p … achieved a level of statistical significance
- R^2 … model goodness-of-fit
- ODP … overdispersion parameter

In two cases the *p*-value (marked red in Table 3) exceeded the original threshold of 0.1. Since the exceedance was not much high (approx. 0.1), we kept the models in the list.

All explanatory variables (VKM, speed CV, speed CI, incidents) had a logical (positive) sign. Also, model goodness-of-fit was satisfactory: the R^2 values were between approx. 0.6 and 0.8, which is comparable to [31] who reported values up to 0.85.

Models for motorways had the highest R^2 values. Regarding the indicators, the model with incidents performed the best.

Validation 3: Ranking consistency

For illustrative purposes, ranking consistency was assessed for four variants of top selection from the ranked list of 110 provinces:

- 10% = 11
- 25% = 28
- 50% = 55
- 75% = 83

The results are reported in Table 4 and visualized in Figure 5. Based on the results we may comment that

- rural roads performed the worst, motorways performed the best (similarly to previous validation steps)
- incidents had the highest values of all indicators (approx. 60% and more)

For illustration, it may be noted that previous studies with the same ranking test [44–46] mostly reported values around 50%.

TABLE 4 Results of ranking consistency test

Top selection		Speed CV			Speed CI			Incidents		
		Urban	Motor-way	Rural	Urban	Motor-way	Rural	Urban	Motor-way	Rural
10%	11	36%	55%	9%	45%	27%	18%	73%	82%	55%
25%	28	36%	58%	25%	40%	58%	33%	58%	65%	58%
50%	55	49%	60%	53%	60%	51%	58%	73%	75%	76%
75%	83	80%	91%	74%	81%	76%	78%	87%	80%	85%

5 | DISCUSSION AND CONCLUSIONS

The objective of the study was to develop a speed-related indicator from big data (FCD) and explore its validity. We used both crash and FCD data aggregated at the Italian province level. Validity was tested by three different approaches, and we obtained the following results:

- Linear correlation:
  - highest between incidents and crashes
  - lowest on rural roads, highest on motorways
- Crash frequency model:
  - models for motorways had the highest R^2 values
  - the model with incidents performed the best
- Ranking consistency:
  - rural roads performed the worst, motorways performed the best
  - incidents had the highest values of all indicators

In sum, incidents turned out to be the best (i.e. the most valid) SPI, especially on motorways. This may be due to the fact that on motorways the geometric characteristics and traffic components are more homogeneous, in sufficient technical conditions, and with access restrictions. This homogeneity may explain the low values of speed variance and CI, as shown in Table 1, and its better ability to reflect the safety performance (incidents and crashes). In contrast, rural and urban roads have many heterogeneous characteristics and other risk factors may play a role, which masks the real impact of speed. This is especially true in urban areas with a high variety of users, the presence of intersections, pedestrian crossings etc.

However, some limitations of the study should be highlighted.

Aggregated data. It was not possible to access the raw data; the available data had already been aggregated at the level of Italian provinces and municipalities. The analysis was carried out at the level of provinces to achieve better representativeness, with more homogeneous and significant penetration rates. It would be beneficial to undertake further analyses at a lower level, potentially including more road-related attributes, such as number of lanes, traffic lights, or roadside conditions. Access to raw data would also allow for better map-matching and overcoming some of the problems highlighted in the next point.

Definition of road categories. The definition of urban and rural road categories is also complex. In metropolitan areas, it is difficult to perfectly distinguish between urban and rural roads; in rural areas, the road cross-section may vary between a typical two-lane road and a national road similar to a motorway. Such a mix of characteristics may create a bias and affect the results.

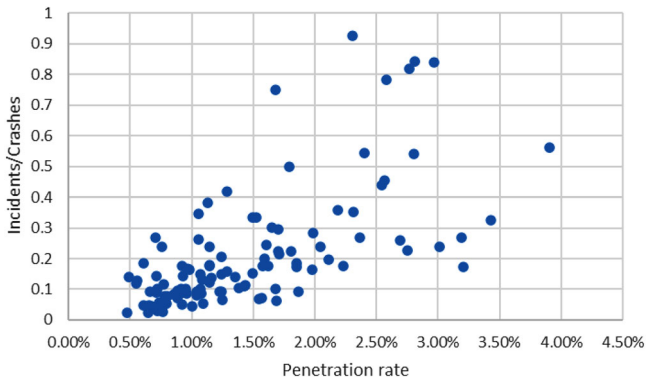


FIGURE 6 Provinces by penetration rate and incidents per crash ratio

Definition of traffic flow. Traffic may be either free-flow or congestion, and each of these traffic flow conditions has a different influence on speed, as well as road safety [51]. It would be thus beneficial to distinguish these conditions in related analyses. For this purpose, knowledge of gaps between the vehicles would be necessary—but it is in principle unavailable when using data from floating vehicles, which represent only a sample of driving vehicles. Without this information, as was the case in our study, the results represent a mix of both free-flow and congested conditions.

Representativeness of the fleet. The monitored vehicle fleet (determined as the total number of vehicles that made at least one trip in the month observed) represents about 1.7% of the total Italian fleet. The penetration rate is variable from province to province and varies from a minimum of 0.48% to a maximum of 3.91%.

A relationship is observed between the fleet penetration rate and the number of incidents per crash in a province. As the penetration rate increases, the ratio of incidents per crash increases (correlation coefficient of 0.65; see Figure 6). In general, a penetration rate higher than 2% corresponds to an incident per crash ratio greater than 0.17.

Seeing the highlighted limitations, the presented paper may be considered a rather preliminary (exploratory) study, which provides points for further research. Future studies should analyze the relationship between speed, incidents, and crashes while

taking into account the geometric and functional characteristics of roads that influence vehicle speeds (such as the number of lanes, the presence of roadside verges, or traffic islands). The possibility of accessing raw FCD data and the availability of up-to-date geographical road databases such as OpenStreetMap would make it possible to map-match the road network accurately and to better define road categories that take these aspects into account. Using other approaches to statistical validation may be considered as well.

The main conclusions of this study can be summarized as follows:

- The number of incidents derived from FCD in a certain area can be used as an SPI thanks to the relevant correlation with road crashes, allowing its use for even daily monitoring, for example, with the elaboration of rankings of provinces by safety level where data are available.
- The speed-based indicators (CV and CI) aggregated at the provincial level have a good level of correlation with the number of crashes in the case of motorways, but the correlation values for urban and rural roads are rather low. This link should be further explored with dedicated analysis.

AUTHOR CONTRIBUTIONS

Jiří Ambros: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Writing – original draft; Writing – review & editing. Davide Shingo Usami: Conceptualization; Data curation; Formal analysis; Methodology; Writing – review & editing. Veronika Valentová: Funding acquisition; Methodology; Supervision; Validation; Writing – review & editing.

ACKNOWLEDGEMENTS

The authors thank OCTO Telematics and Istat for providing data. The authors appreciate the help of Robert Zůvala with statistical analyses. The paper was produced with the financial support of the Czech Ministry of Transport within the program of the long-term conceptual development of research institutes.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

FUNDING INFORMATION

The paper was produced with the financial support of the Czech Ministry of Transport within the program of the long-term conceptual development of research institutes.

DATA AVAILABILITY STATEMENT

Data are subject to third party restrictions.

ORCID

Jiří Ambros https://orcid.org/0000-0003-2707-6243

REFERENCES

1. WHO (World Health Organization. Global status report on road safety 2018. World Health Organization, Geneva (2018). https://www.who.int/publications/i/item/9789241565684
2. Adminaitė-Fodor, D., Carson, J., Jost, G.: Ranking EU progress on road safety. 15th Road Safety Performance Index (PIN) report. Brussels: European Transport Safety Council (2021). https://etsc.eu/15th-annual-road-safety-performance-index-pin-report/
3. EC (European Commission). Road safety targets – Monitoring report November 2020. European Commission, Brussels, (2020). https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/monitoring_report_november_2020.pdf
4. EC (European Commission). EU road safety policy framework 2021–2030 – Next steps towards 'Vision Zero'. European Commission, Brussels, (2020). https://op.europa.eu/cs/publication-detail/-/publication/d7ee4b58-4bc5-11ea-8aa5-01aa75ed71a1
5. ETSC (European Transport Safety Council). A methodological approach to national road safety policies. European Transport Safety Council, Brussels, (2006). https://etsc.eu/wp-content/uploads/A-methodological-approach-to-national-road-safety-policies1.pdf
6. OECD (Organisation for Economic Co-operation and Development). Towards Zero: Ambitious Road Safety Targets and the Safe System Approach. Organisation for Economic Co-operation and Development, Paris (2008). https://www.nrspp.org.au/resources/towards-zero-ambitious-road-safety-targets-and-the-safe-system-approach/
7. Gitelman, V., Vis, M., Weijermars, W., Hakkert, S.: Development of road safety performance indicators for the European countries. Adv. Soc. Sci. Res. J. 1, 138–158 (2014)
8. ETSC (European Transport Safety Council). Transport safety performance indicators. European Transport Safety Council, Brussels, (2001). https://etsc.eu/wp-content/uploads/Transport-safety-performance-indicators.pdf
9. Hakkert, A.S., Gitelman, V., Vis, M.A. (eds.). Road Safety Performance Indicators: Theory. SafetyNet project deliverable D3.6. (2007). http://www.dacota-project.eu/Links/erso/safetynet/fixed/WP3/sn_wp3_d3p6_spi_theory.pdf
10. Chang, A., Saunier, N., Laureshyn, A.: Proactive Methods for Road Safety Analysis. SAE International, Warrendale (2017). https://ictct.net/wp-content/uploads/SMoS_Library/LIB_Chang_2017.pdf
11. Johnsson, C., Laureshyn, A., De Ceunynck, T.: In search of surrogate safety indicators for vulnerable road users: A review of surrogate safety indicators. Transp. Rev. 38, 765–785 (2018)
12. Tarko, A., Davis, G., Saunier, N., Sayed, T., Washington, S.: White Paper Surrogate Measures of Safety. Transportation Research Board, Washington (2009). https://sites.google.com/site/surrogatesafety/documents
13. Polders, E., Brijs, T.: How to analyse accident causation? A handbook with focus on vulnerable road users. InDeV project deliverable 6.3. Hasselt: Hasselt University, (2018). https://www.bast.de/InDeV/EN/Handbook/Handbook-InDeV.html
14. Neilson, A., Indratmo Daniel, B., Tjandra, S.: Systematic review of the literature on big data in the transportation domain: Concepts and applications. Big Data Res. 17, 35–44 (2019)
15. Horrey, W.J., Lesch, M. F., Dainoff, M. J., Robertson, M. M., Ian Noy, Y.: On-board safety monitoring systems for driving: Review, knowledge gaps, and framework. J. Saf. Res. 43, 49–58 (2012)
16. Tselentis, D.I., Yannis, G., Vlahogianni, E.I.: Innovative motor insurance schemes: A review of current practices and emerging challenges. Accid. Anal. Prev. 98, 139–148 (2017)
17. Singh, H., Kathuria, A.: Analyzing driver behavior under naturalistic driving conditions: A review. Accid. Anal. Prev. 150, 105908 (2021)
18. Nilsson, G.: Traffic safety dimensions and the Power Model to describe the effect of speed on safety. Bulletin 221. Lund University, Lund (2004). https://portal.research.lu.se/en/publications/traffic-safety-dimensions-and-the-power-model-to-describe-the-eff
19. Elvik, R.: The Power Model of the relationship between speed and road safety: Update and new analyses. Report 1034/2009. Institute of Transport Economics, Oslo (2009). https://www.toi.no/getfile.php?mmfileid=13206
20. Elvik, R.: A re-parameterisation of the power model of the relationship between the speed of traffic and the number of accidents and accident victims. Accid. Anal. Prev. 50, 854–860 (2013)

21. Elvik, R., Vadeby, A., Hels, T., Van Schagen, I.: Updated estimates of the relationship between speed and road safety at the aggregate and individual levels. Accid. Anal. Prev. 123, 114–122 (2019)
22. Fernández Llorca, D., Hernández Martínez, A., García Daza, I.: Vision-based vehicle speed estimation: A survey. IET Intell. Transp. Syst. 15, 987–1005 (2021)
23. Ambros, J., Jurewicz, C., Chevalier, A., Valentová, V.: Speed-related surrogate measures of road safety based on floating car data. In: Macioszek, E., Sierpiński, G. (eds.) Research Methods in Modern Urban Transportation Systems and Networks, pp. 129–144. Springer, Nature Switzerland AG, Cham (2021)
24. Pei, X., Wong, S.C., Sze, N.N.: The roles of exposure and speed in road safety analysis. Accid. Anal. Prev. 48, 464–471 (2012)
25. Wang, X., Fan, T., Chen, M., Deng, B., Wu, B., Tremont, P.: Safety modeling of urban arterials in Shanghai, China. Accid. Anal. Prev. 83, 57–66 (2015)
26. Wang, X., Zhou, Q., Quddus, M., Fan, T., Fang, S.: Speed, speed variation and crash relationships for urban arterials. Accid. Anal. Prev. 113, 236–243 (2018)
27. Gitelman, V., Doveh, E., Bekhor, S.: The relationship between travel speeds, infrastructure characteristics, and crashes on two-lane highways. J. Transp. Saf. Secur. 10, 545–571 (2018)
28. Stipancic, J., Miranda-Moreno, L., Saunier, N.: Impact of congestion and traffic flow on crash frequency and severity: Application of smartphone-collected GPS travel data. Transp. Res. Rec. 2659, 43–54 (2017)
29. Stipancic, J., Miranda-Moreno, L., Saunier, N., Labbe, A.: Surrogate safety and network screening: Modelling crash frequency using GPS travel data and latent Gaussian Spatial Models. Accid. Anal. Prev. 120, 174–187 (2018)
30. Stipancic, J., Miranda-Moreno, L., Saunier, N., Labbe, A.: Network screening for large urban road networks: Using GPS data and surrogate measures to model crash frequency and severity. Accid. Anal. Prev. 125, 290–301 (2019)
31. Stipancic, J., Racine, E. B., Labbe, A., Saunier, N., Miranda-Moreno, L.: Massive GNSS data for road safety analysis: Comparing crash models for several Canadian cities and data sources. Accid. Anal. Prev. 159 (106232) (2021)
32. Banihashemi, M., Dimaiuta, M., Zineddin, A., Spear, B., Smadi, O., Hans, Z.: Using linked SHRP2 RID and NPMRDS data to study speed-safety relationships on urban interstates and major arterials. 98th TRB Annual Meeting, Washington (2019)
33. Ederer, D.J., Rodgers, M. O., Hunter, M. P., Watkins, K. E.: Case study using probe vehicle speeds to assess roadway safety in Georgia. Transp. Res. Rec. 2674(11), 554–562 (2020)
34. Das, S., Geedipally, S.R., Fitzpatrick, K.: Inclusion of speed and weather measures in safety performance functions for rural roadways. IATSS Res. 45, 60–69 (2021)
35. De Fabritiis, C., Ragona, R., Valenti, G.: Traffic estimation and prediction based on real time floating car data. In: 11th International IEEE Conference on Intelligent Transportation Systems (ITSC), Beijing, pp. 197–203 (2008)
36. Fusco, G., Colombaroni, C., Isaenko, N.: Short-term speed predictions exploiting big data on large urban road networks. Transp. Res. Part C 73, 183–201 (2016)
37. Colombaroni, C., Fusco, G., Isaenko, N.: Analysis of road safety speed from floating car data. Transp. Res. Procedia 45, 898–905 (2020)
38. Bekhor, S., Lotan, T., Gitelman, V., Morik, S.: Free-flow travel speed analysis and monitoring at the national level using global positioning system measurements. J. Transp. Eng. ASCE 139, 1235–1243 (2013)
39. Diependaele, K., Riguelle, F., Temmerman, P.: Speed behavior indicators based on floating car data: Results of a pilot study in Belgium. Transp. Res. Procedia 14, 2074–2082 (2016)
40. Jurewicz, C., Espada, I., Makwasha, T., Han, C., Alawi, H., Ambros, J.: Use of connected vehicle data for speed management in road safety. In: 28th ARRB Conference, Brisbane (2018)
41. Teuchies, M., Vadeby, A., Van Schagen, I., Riguelle, F., Tutka, P.: Methodological guidelines – KPI speeding. Baseline project deliverable. Vias institute, Brussels (2021). https://baseline.vias.be/en/publications/
42. DG MOVE (Directorate-General for Mobility and Transport). CARE Database – CaDaS Common Accident Data Set (2021). https://ec.europa.eu/transport/road_safety/sites/default/files/cadas_glossary_v_3_8.pdf
43. Ambros, J., Havránek, P., Valentová, V., Křivánková, Z., Striegler, R.: Identification of hazardous locations in regional road network – comparison of reactive and proactive approaches. Transp. Res. Procedia 14, 4209–4217 (2016)
44. Cheng, W., Washington, S.: New criteria for evaluating methods of identifying hot spots. Transp. Res. Rec. 2083 76–85 (2008)
45. Montella, A.: A comparative analysis of hotspot identification methods. Accid. Anal. Prev. 42, 571–581 (2010)
46. Yu, H., Liu, P., Chen, J., Wang, H.: Comparative analysis of the spatial analysis methods for hotspot identification. Accid. Anal. Prev. 66, 80–88 (2014)
47. Ambros, J., Jurewicz, C., Turner, S., Kieć, M.: An international review of challenges and opportunities in development and use of crash prediction models. Eur. Transp. Res. Rev. 10(35) (2018)
48. Lord, D., Mannering, F.: The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transp. Res. Part A 44, 291–305 (2010)
49. Reurings, M., Janssen, T., Eenink, R., Elvik, R., Cardoso, J., Stefan, C.: Accident prediction models and road safety impact assessment: A state-of-the-art. RIPCORD-ISEREST project deliverable 2.1. European Commission, Brussels (2005). http://ripcord.bast.de/pdf/RI-SWOV-WP2-R1-State_of_the_Art.pdf
50. Yannis, G., Dragomanovits, A., Laiou, A., La Torre, F., Domenichini, L., Richter, T., Ruhl, S., Graham, D., Karathodorou, N.: Road traffic accident prediction modelling: A literature review. Proc. Institut. Civil Engineers Transp. 170, 245–254 (2017)
51. Ambros, J., Kyselý, M.: Free-flow vs car-following speeds: Does the difference matter? Adv. Transp. Stud. 40, 17–26 (2016)

How to cite this article: Ambros, J., Usami, D.S., Valentová, V.: Developing speed-related safety performance indicators from floating car data. IET Intell. Transp. Syst. 1–9 (2022). https://doi.org/10.1049/itr2.12281