doi.org/10.26398/asaproc.0013

Road accidents in Italy: New indicators, at province level, based on geographic information system open data

Marco Broccoli[a], Silvia Bruzzone[b]
[a] Directorate for Methodology and Statistical Process Design, Istat, Rome, Italy
[b] Directorate for Social Statistics and Welfare, Istat, Rome, Italy

1. Introduction

Road safety performance indicators (RSPIs) give a multidimensional approach for accidents investigation concerning roads, vehicles and persons involved. Combining the use of statistical surveys, administrative geographical information systems (GIS) and big data (BD) sources, the result gives new elements on planning infrastructure solutions and applying policies to reduce deaths and serious injuries.

Nowadays there is a clear information bias as regards the appropriate reference denominators placed as basis in construction of statistical indicators linked to road accidents. Resident population, used as a common proxy for exposed at risk in a specific geographical area, usually is not an appropriate solution due to the seasonal nature of road accidents. Vehicle fleet might provide a comprehensive overview of the vehicles present within a certain region and should give a more accuracy on the statistical indicators. The "deductible distortion" is primarily attributed to the mobility of road users. The length of the road network gives a consistent first set of information concerning the different territories. This information is not available from official statistics at national level. Although there are archives and detailed road graphs for each municipality, province and region, a harmonised and systematic national road registry has not been established yet.

The present research focuses on the use of existing administrative sources, the scouting of new sources and the analysis of integrated and auxiliary data to calculate road accidents rates, mortality and harmfulness indexes, comparing these measures to the correspondent length in meters of carriageway by road direction from the Open Street Map (OSM). This approach has been used for the first time, by the authors, to build road accidents indicators and adopted by the "experimental statistics" at Istat (Broccoli and Bruzzone 2019; 2021). Although OSM is mentioned as a data source in many scientific articles, the main purpose was for GIS-based traffic simulation (Zilske et al., 2011). The final purpose of the project is to expand statistical information with the supply of traffic flows (vehicles / km) on the national road network. This would make it possible to calculate the probability of being involved in the accident, considering the different exposure to risk of accidents.

2. Methodology

GIS is a geographic system designed to capture, store, manipulate, analyse, manage and present spatial or geographic data. To relate information from different sources, GIS uses a spatial location as the key index variable (key reference by position). Just as a relational database containing text or numbers can relate many different tables using common key index variables, GIS can relate otherwise unrelated information by using location as the key index variable. This key characteristic of GIS is nowadays a valid alternative on statistical dissemination. Any variable that can be located spatially using an x, y, and z coordinates, representing longitude, latitude, and elevation, respectively. These GIS coordinates may represent other quantified systems of territories (polygons), road networks (lines) and point of traffic (points).

"Join attributes by location"[1] is the algorithm that takes an input vector layer and creates a new vector layer, adding on the extended version the additional characteristics of the second layer. Specifically, shapes of the Italian localities, wide areas on which one or more houses are scattered, were used (2011 census latest available for the whole country in Italy) and consequently the Italian municipalities[2] were aggregated according to the locality layer. The aim of the project is to harmonise the location statistical variable of the Istat survey with the class of road graph. Therefore, the street layer is harmonized with the classification of the accidents location and permits to create new safety indicators by road type. The localities are classified as: urban areas, small-inhabited areas, productive areas and widespread houses[3] [4].

OSM[5] is a collaborative project aimed on creating free content maps of the world. The project aims at a collection world of geographical data, with the main purpose of creating maps and cartography. The key feature of the geographic data present in OSM is having a free license, the Open Database License. It is therefore possible use them freely for any purpose with the only constraint of mentioning the source. Everyone can contribute by populating or correcting data. The maps are created using the data recorded by portable GPS devices, aerial photographs, and other free sources.

Although OSM is an **o**pen-**s**ource tool based on information from a community, the product provides data to be considered reliable and consistent, so much that a major part of GPS Android and iOS navigation software on portable devices are powered by OSM, for example WisePilot, Maps.me, NavFree, Scout, ShareNav, MoNav, Navitel, etc.

The OSM vector layers, used in the present research, daily updated and free downloadable data consist in: road graph; point of traffic (POT); buildings; use of the land; natural places; POWF (point of worship); POIS (Point of interest); railways; transport; water areas; and water ways.

Table 1 contains the list of the main different type of road arch by OSM used to provide the length in meters of carriageways by road type and Italian provinces. The arch road types selected to calculate the indicators analysed in the present research are referred to the motorized vehicles flow: motorway, trunk, primary, secondary, tertiary, unclassified, residential, living street, motorway link, trunk link, primary link, secondary link, tertiary link, service, unknown. Pedestrian, track, track grade, bridleway, cycle way, footway, path, steps, also defined in OSM road types, are not considered part of the observed domain.

Besides, to calculate road accidents indicators, with denominator represented by the arches length in OSM, we built a "bridge matrix" between road categories, classified by functional road type and the categories linked to the roads holder, used by Istat road accidents survey. The systematic classification of road arches, classified in the categories used by Istat, has been modified, in respect of a first release (Broccoli and Bruzzone 2019; 2021).

A new analytical classification has been adopted, using a more refined technique of attributing single road arches, about three and a half million in total, to the Istat classification groups (Table 2).

---

[1] Join attributes by location is a functionality of QGIS software. The authors adapted the implementation of the algorithm for the specific data used in the present study - 27.1.17. Vector general — QGIS Documentation documentation

[2] Italian municipalities codes list at 1/1/2023 https://www.istat.it/it/archivio/6789

[3] 1.Urban area is an aggregate of contiguous or near houses with roads, squares and similar, or however areas characterized by services or public activities. 2.Small-inhabited area is an area without a place of collection, and it is based on a group of at least fifteen contiguous and near buildings, if the distance between the buildings does not exceed thirty meters and it is lower than the distance between the centre and the nearest of the houses clearly scattered. 3.Productive area is an extra-urban area not included in the centres or residential areas with more than 10 local units. 4.Widespread houses represent an area with scattered house in the municipal territory at a distance not enough to constitute a built-up area.

[4] Istat - Basi territoriali e variabili censuarie https://www.istat.it/it/archivio/104317 ; Istat - Descrizione dei dati geografici e delle variabili censuarie delle Basi territoriali per i censimenti: anni 1991, 2001, 2011 Descrizione Dati Geografici (istat.it)

[5] OpenStreetMap provides geographic data on thousands of websites, mobile and hardware devices. OpenStreetMap is built by a community of mappers, who contribute, update and monitor data on roads, cafes, railway stations and much more, all over the world - OSM: OpenStreetMap

Table 1: ***Open Street Map main road arches classification***

OSM Road type	Road type Description
Motorway	Free or by toll payment highways, including motorway connections.
Motorway Link	The link roads (slip roads/ramps) leading to/from a motorway from/to a motorway or lower-class highway. Normally with the same motorway restrictions
Trunk	Roads type between Motorway, Motorway connections and Primary Road. The junction section of a motorway-ring road that leads to the city centre can also be classified as trunk.
Trunk Link	The link roads (slip roads/ramps) leading to/from a trunk road from/to a trunk road or lower-class.
Primary	Roads of national and regional importance not classified as motorways, trunks, or their link. They connect the main cities to each other. Mostly, they are classified as SS (Main Roads) or SR (Regional).
Primary link	The link roads (slip roads/ramps) leading to/from a primary road from to a primary road or lower-class highway.
Secondary	Secondary Roads of regional and provincial importance. They link together the main municipalities of a region. They are, usually classified as SP (provincial roads) but there are some exceptions. In urban areas, they are normally classified as important streets with two lanes in each direction
Secondary link	The link roads (slip roads/ramps) leading to/from or from/to a secondary road or lower-class highway.
Tertiary	Roads of local rank. They connect smaller municipalities together. In urban areas, they are side roads to primary and secondary roads with a medium flow of traffic.
Tertiary link	The link roads (slip roads/ramps) leading to/from or from/to a tertiary road or lower-class highway.
Residential	Roads in a residential area, which serve as an access to housing, without function of connecting settlements.
Living Street	Residential road where pedestrians have legal priority over cars, speeds are kept very low.
Service	Access roads or internal service areas, beaches, camping, industrial areas, shopping centres, residences, parking places, landfills, installations, etc.
Unknown/Unclassifie	Not classified

Table 2*: Bridge coding table between roads arches classification by OSM, localities and road type (a)*

Road Arches classification by Open Street Map	Localities at Census 2011			
	Urban areas + Small		Productive areas + Wide	
	Road Localisation by Istat Road accidents survey			
	Motorways	Urban Roads	Motorway	Rural Roads
Motorway and Trunk	x		x	
Primary, Secondary and Tertiary		x		x
Residential and Living Street		x		x
Motorway Link and Trunk link	x		x	
Primary, Secondary and Tertiary Link, Service		x		x
Unknown		x		x

3. Data and results

Using data from the OSM source, has been done new experimental statistics on road mortality and injuries. It has been compared the most common rates at province level out of the resident population or vehicles fleet amount. The rates consist in a ratio between a numerator based on data from Istat road accidents survey and a denominator given by the length of the road network, by type or the average resident population in Italy or the vehicle fleet counted by ACI.

All road accidents are resulting in deaths or injuries during the years 2018, 2020 and 2022, involving at least a vehicle circulating on the national road network and documented by police.[6] Moreover, to enrich data, an additional information has been obtained from OSM, in particular, data on the traffic points detected on the road arches (PoT). New road accident indicators, "weighted" with the information on traffic intensity, were built too, considering the kilometres of carriageway with the presence of a traffic point on the arch.

Finally, a group of composed and synthetic road accidents indicators completes the set of indicators. The method used is the arithmetic mean (z-scores) MZ method (Software Ranker) [7],

[6] Istat - Road accidents in Italy Istat.it Road accidents

[7] Istat Ranker tool desktop software: http://www.istat.it/en/tools/methods-and-it-tools/analysis-tools/ ranker; I.Ranker

(Mazziotta and Pareto, 2013). The values of indices have been calculated according to the MZ (arithmetic average Z scores)[8] method for road arches, vehicle fleet and population. This method allows to build indicators excluding the traffic volumes on the road and to consider the cause of the accident in a different light, linked to the driving behaviour and the characteristics of the infrastructures and vehicles, with or without the traffic flow influence. The results show that the ranking of the Italian provinces, in respect of the new synthetic indicator, is modified if compared with the ranking of the single indexes and underlines a new profile of accidents risk at territorial level.

3.1. Comparison of indicators and risk of road accident ranking in Italian provinces

The proposed indicators refer to the ratio between road accidents, deaths, injuries, and vehicles to the road length, considering the additional information on points of traffic too, the resident population and the vehicle fleet. In addition to what already disseminated by *Istat Experimental Statistics* (Broccoli and Bruzzone 2019 and 2021), updated information on road accidents for the years 2018, 2020 and for the last available 2022 were used in the present study. A comparison between the three key-years representative for the previous, during and post pandemic period, was useful to explain the change in mobility habits and the difference through the country counties. Some evidence and results are presented in the following Section 3.2.

Concerning the different indicators built, is interesting to highlight the difference between the measures of the same events, road accidents ratio or road mortality ratio, for example, but considering different denominators. In support of this thesis, an interesting result is given by the covariance matrix results, obtained by means the synthesis of all indicators by year, province and road type and processed with the tool *Ranker* (Ref. Footnote n.5). The application of different weighting criteria leads, in fact, to very divergent results. Table 3 shows the covariance values, referred to the variation of each variable contained in the matrix in respect of all others. The values show, in fact, that the risk to be involved in a road accident, within the province of residence (0.2929-0.3712) or within the vehicle registration province (0.4138-0.4861) is significantly lower if compared to the indicator out of the length road arch (including PoT), where the accident occurred. The road accidents indicators, referred to the road length by province, therefore, allows to obtain a result closer to the road accident risk measures, using traffic flows data. The last data would represent, in fact, the best and correct denominator for road accidents rates.

Focusing on the year 2022 on the Road accidents indicators by road length "weighted" by the PoT information (number of accidents, vehicles involved, deaths and injuries per 100 kilometres of carriageway in the province) "adjusted" by the effect on accidents of the presence of traffic points on the roads. Some evidence shows that a maximum exposure to the risk of accidents and number of vehicles involved, for motorways and urban roads in main cities. Milan records a maximum for accidents rates per 100 km of carriageway, on motorways and urban roads (respectively 929.37 and 629.48 per 100 km), followed by Rome (851.44 and 447.97) and Genoa (703.51 and 492.74). On rural roads, in contrast, the medium-sized provinces result most affected, Monza and Brianza has the maximum value (518.55 road accidents per 100 km). Similar results are detected for vehicles and injuries, slightly different the results for road mortality rates that highlight Rome disadvantage on urban roads. Meanwhile Naples on rural roads and Rimini on the motorways. Considering the rates out of the resident population, the ranking is completely different, for road accidents rates (per 1 ml of inhabitants), in fact, on motorways the highest value is for the province of Savona, on rural roads, for Latina and Genoa on urban roads.

Milan and Rome, in the first positions for road accidents risk on motorways, in respect of road length, went down in 8th and 10th positions out of the resident population (11th and 12th

web application https://i.ranker.istat.it

[8] The three methods evaluated are: MZ - arithmetic mean (z-scores); MR - relative indices method (IR); and Mazziotta-Pareto index (Mazziotta and Pareto, 2013). The criterion applied after the robustness analysis is MZ.

position for urban roads) (Figures 1 and 2).

Table 3*: Covariance matrix between computing indicators (a)*

Years	2018			2020			2022		
Denominators Type	Population	Vehicle fleet	Road graph	Population	Vehicle fleet	Road graph	Population	Vehicle fleet	Road graph
Population	1.0000	0.9066	0.3712	1.0000	0.9148	0.2929	1.0000	0.8914	0.3688
Vehicle fleet		1.0000	0.4861		1.0000	0.4138		1.0000	0.4802
Road graph			1.0000			1.0000			1.0000

a) Results obtained by Istat Ranker application.

Figure 1: ***Road accidents rate by province. Year 2022 (per 100 km road length - Motorways)***

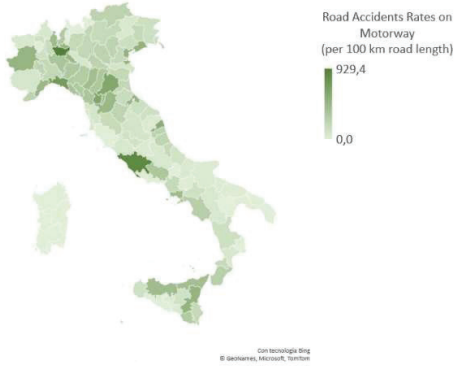
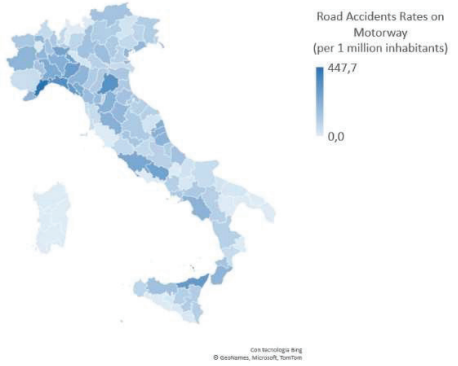


Figure 2: ***Road accidents rate by province. Year 2022 (per 1 million inhabitants - Motorways)***



The biggest provinces as Rome, Milan, Naples, Palermo and Turin with a high level of population lead to moderate values of indicators in comparison to reality. The high density of roads on these territories is correctly highlighted by road indicators ranking them as the worst provinces on exposure risk to be involved on an accident. At the same time, small provinces as Monza and Brianza and Varese, as the biggest one, get influenced on the difference between population and road extension on the two methods ranking. Traffic flows on this small area explode risk indicators to high-level values.

Some of medium size province as Grosseto, Savona, Rieti, Piacenza and Macerata using population reference give them a dangerous exposure risk harmonizing with their road infrastructure, the ranking gives a more accurate value with level of security better than the national average.

3.2. The effect of pandemic season on road accidents

During the year 2022, after the years in which the pandemic has seen the most acute phase, a clear recovery in mobility and in road accidents occurred. Road accidents, deaths and injuries compared on 2020, showed an overall growth above the months in which the traffic and mobility limitation measures enforced on containing of the virus. With respect to 2018, the road accidents and casualties in 2022, anyway, showed a decrease.

Concerning the comparison between the three years 2018, 2020 and 2022, the results in Table 4 show the ranking of the five best and worst performance by province in Italy of the synthesis indicators (method MZ processed by the Istat software Ranker, see footnote 5). Looking at the classification, Milan and Monza and Brianza appear constantly the provinces with the highest risk for the road safety, followed by Genoa, in 2018, the year of the collapse of Morandi Bridge, and by Rome, Naples, and Trieste in 2020 and 2022. The best performance, considering the compendium of the indicators, instead, is in towns of the southern regions: Basilicata, Sicily and Sardinia.

Table 4: ***Best and worst performance in the road accidents risk. Ranking first and last five provinces by year (ranking by arithmetic average Z scores). Years 2018, 2020 and 2022***

Best Z Road graph performance (a)							
2018		**2020**		**2022**		**2022 No Traffic**	
Isernia	0.9593	**Potenza**	0.9613	**Isernia**	0.9953	**Isernia**	0.9356
Agrigento	0.9376	**Isernia**	0.9447	**Benevento**	0.9488	**Benevento**	0.8894
Benevento	0.9224	**Benevento**	0.9320	**Sud Sardegna**	0.9057	**Campobasso**	0.8519
Oristano	0.9133	**Agrigento**	0.9094	**Campobasso**	0.8877	**Sud Sardegna**	0.8514
Nuoro	0.8514	**Nuoro**	0.8566	**Potenza**	0.8765	**Potenza**	0.8460
Worst Z Road graph performance (a)							
2018		**2020**		**2022**		**2022 No Traffic**	
Naples	-1.6180	**Genoa**	-1.3306	**Trieste**	-1.8683	**Naples**	-1.8181
Rome	-1.8673	**Naples**	-2.1097	**Naples**	-1.9743	**Trieste**	-1.9996
Genoa	-2.1538	**Rome**	-2.2092	**Rome**	-2.2733	**Rome**	-2.2437
Monza- Brianza	-2.9951	**Monza -Brianza**	-2.5266	**Monza-Brianza**	-2.2891	**Monza - Brianza**	-2.8428
Milan	-3.3958	**Milan**	-2.9771	**Milan**	-3.3239	**Milan**	-3.8396

(a) The values of indices have been calculated according to the MZ (arithmetic average Z scores)

4. Conclusion

Big data are not generated directly for statistical purposes, for a statistical use, a suitable methodology must be able to link events to refer mainly to units of the population of interest for official statistics. Other elements required are to process data with the aim to collect them coherently with the statistical framework, to weight data to guarantee representativeness with respect to the target population and to estimate aggregates of interest including quality measures by spatial and time dimensions. The project of the use of OSM for the investigation of road accidents pattern is adherent to the features cited and the activities in progress may provide new developments in the future.

References

Broccoli, M., Bruzzone, S. (2019). Use of the open street map to calculate indicators for road accidents on the Italian roads. Updating with 2017 data. *Istat Experimental Statistics*, Rome https://www.istat.it/en/archivio/257384.

Broccoli, M., Bruzzone S. (2021). Use of open street map for accident investigation on the road and motorways networks. *Rivista Italiana di Economia Demografia e Statistica*, **LXXV**(4), pp. 157-168.

Farchi, S., Molino, N., Giorgi Rossi, P. et al. (2006). Defining a common set of indicators to monitor road accidents in the European Union. *BMC Public Health,* **6**, pp. 183.

Hakkert, A.S., Braimaister, L. (2002). The uses of exposure and risk in road safety studies, *Report documentation SWOV Institute for Road Safety Research,* Leidschendam (NL).

Harbison, F.H., Maruhnic, J., Resnick, J. R. (1970). Quantitative analyses of modernization and development, *Princeton University Press,* New Jersey, pp. 224.

Mazziotta, M., Pareto, A. (2013). Methods for constructing composite indices: One for all or all for one?. *Rivista Italiana di Economia Demografia e Statistica,* **LXVII**(2), pp. 67-80.

Retallack, A.E., Ostendorf, B. (2020). Relationship between traffic volume and accident frequency at intersections. *International Journal of Environmental Research and Public Health,* **17**(4), pp. 139.

Zilske, M., Neumann, A., Nage, K. (2011). OpenStreetMap for traffic simulation. In: *Proceedings of the 1st European State of the Map: OpenStreetMap Conference*, Wien: OpenStreetMap Austria, pp. 126–134.