rivista di statistica ufficiale

REVIEW OF OFFICIAL STATISTICS

n. 3
2021

In this issue:

Istat
Istituto Nazionale di Statistica

rivista di statistica ufficiale

REVIEW OF OFFICIAL STATISTICS

n. 3
2021

In this issue:

Istat

rivista di statistica ufficiale
n. 3/2021

Four-monthly Journal: registered at the Court of Rome, Italy (N. 339/2007 of 19th July 2007).

e-ISSN 1972-4829
p-ISSN 1828-1982

The Scientific Committee, the Editorial Board and the authors would like to thank the anonymous reviewers (at least two for each article, on a voluntary basis and free of charge, with a double-anonymised approach) for their comments and suggestions, which enhanced the quality of this issue of the Rivista di statistica ufficiale.

Editorial Preface

This third Issue of Istat *Rivista di statistica ufficiale* closes the year 2021 with a monothematic content concerning work-related road safety.

On the one hand, indeed, road accidents continue to represent one of the main causes of death among the population and characterise an important public health problem due to their high costs, especially in human and social terms. On the other hand, at a more specific level of detail road accidents are a significant cause of death at work as well, confirming the particular danger of the road also as a "workplace". In this framework, the multidimensionality of the question emerges clearly, with the related importance of suitable data source availability, in order to both calculate targeted indicators and conduct in-depth analyses.

Precisely for this purpose, the four scientific articles that compose the present Issue illustrate the results of a joint research project, carried out through the constitution of a working group, with the involvement of the two Institutions that are essential to meet these needs. More specifically, the Italian National Institute for Insurance against Accidents at Work – Inail, with the *Work-related road accident data*, and the Italian National Institute of Statistics – Istat, with the survey on *Road accidents resulting in deaths or injuries*. This successful collaboration aims at connecting and integrating Inail and Istat archives, in particular with data on work-related road accidents by means of transport and other variables, in order to study their characteristics and determinants, so as to orient focussed prevention policies.

The first article, by Luca Taiano, Stefania Massari, Tiziana Tuoto, Luca Valentino, Silvia Bruzzone and Liana Veronico, is methodological. It deals with record linkage, comparing its two deterministic and probabilistic techniques, and showing how the latter provides more significant improvements than the sole deterministic approach, when integrating Inail and Istat data archives. The authors make use of *RELAIS* (*REcord Linkage At IStat*), the *ad hoc* open-source software developed, produced and disseminated by Istat, and prove that the probabilistic algorithm produce a significant increase in linked pairs, while deterministic filters add further accuracy.

The next article is authored by Silvia Bruzzone, Antonella Altimari, Giordana Baldassarre, Roberto Boscioni and Liana Veronico. It covers a descriptive and in-depth analysis using the integrated archive of road accidents and work-related road accidents for the year 2018, derived applying the methods described in the previous paper. By linking two different sources that have in common the accidents (with deaths and injuries), occurring on public roads and with means of transport, it is indeed possible to complete the information also on the persons involved in these accidents, both injured and deceased. This integrated database makes available enriched data that would otherwise not be accessible, and proves to be very useful for addressing policy-making and effective interventions in the field of both general road safety and safety at work.

Claudio Gariazzo, Alessandro Marinaccio, Silvia Bruzzone, Luca Taiano and Liana Veronico in the third article continue deepening the results obtained. Making use of time series, cluster and multiple correspondence analysis, they describe features and temporal developments, also focussing on the most representative road accidents among fatal and non-fatal events, both at work and during commuting. In more detail, urban and suburban areas result the most dangerous places where these events occur. Different types of collision become relevant accidents' origin, using private cars and motorcycles as preferred vehicle. Time series emphasise a strong weekly and seasonal variation. The cluster analysis allows identifying typical road accident profiles, including both accident parameters and economic sectors. Finally, both cluster and multiple correspondence analysis highlight that road accidents occurring on duty are characterised by certain specific economic sectors, while those during commuting can be related to a more heterogeneous range of economic sectors.

This Issue of Istat *Rivista di statistica ufficiale* concludes with the article by Antonella Pireddu, Antonella Altimari, Giordana Baldassarre, Alessandro Marinaccio and Luca Taiano. Still referring to the integrated archive of Inail and Istat data, the authors study approximately 129,000 work-related road accidents that occurred in Italy during the years 2014-2018, investigating on injury rates by road type and economic sector, whose processing has been made possible by the joint research work and record linkage results already described in the other articles. In the current paper, these indicators are examined using spatial and descriptive evaluations together with the analysis of variance - *ANOVA*. The main findings detect maximum injury rates in the macro sectors of sales, metals and machinery, transport and warehouses, health and social services, construction and plant. In addition, *ANOVA* shows that injury variances on rural and urban roads, as well as considering aggregated economic sectors, differ when going down to the detail of the Italian territorial divisions. Consequently, this study proves suitable and appreciable also for the local public health authorities responsible for risk assessment and management in specific contexts.

Patrizia Cacioli
Editor

Nadia Mignolli
Coordinator of the Editorial board

Work-related road accidents: a data linkage procedure applied to assess traffic accidents at work and commuting

Luca Taiano [1], Stefania Massari [1], Tiziana Tuoto [2], Luca Valentino [2], Silvia Bruzzone [2], Liana Veronico [1]

Abstract

Record linkage is a data integration technique whose goal is to identify the same unit represented in different data sources in different ways. Deterministic linkage and probabilistic linkage are two linkage techniques that have been already widely used. The goal of this paper is to show how a record linkage procedure based on a probabilistic approach provides an increase in linked pairs compared to the sole use of a deterministic approach and to provide a step procedure to sequentially apply multiple linkage techniques. Datasets are the Inail (Italian National Institute for Insurance against Accidents at Work) archive of work-related accidents occurring with the use of a vehicle and the Istat (Italian National Institute of Statistics) archive of road accidents resulting in death or injury. It was applied a deterministic linkage followed by probabilistic linkage on the unlinked records. Deterministic linkage undoubtedly considers two records to be a link if they agree on a selection of variables whereas probabilistic linkage assigns a probability of being a link to records. Results show that the probabilistic linkage produced an increase of 18% of the linked pairs compared to the sole use of the deterministic approach.

Keywords: Data integration, record linkage, deterministic linkage, probabilistic linkage, road accidents.

---

1 Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro (*Italian National Institute for Insurance against Accidents at Work*) - Inail.

2 Istituto Nazionale di Statistica (*Italian National Institute of Statistics*) - Istat.

In the writing of this article, the authors were supported by all the members of the Inail-Istat collaborative group: Antonella Altimari, Michela Bonafede, Roberto Boscioni, Claudio Gariazzo, Alessandro Marinaccio, Stefania Massari, Antonella Pireddu, Luca Taiano and Liana Veronico (for Inail); Giordana Baldassarre and Silvia Bruzzone (for Istat).

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

Among methodologies of data integration, the techniques of records linkage are a set of methods whose goal is to identify the same unit represented in different data sources in different ways. Methodologies of data integration such as record linkage create new information assets from the already available ones, adding value to the existing data archives, allowing better insights, new conclusions and reducing the necessities to carry out new surveys.

Two approaches can be used for record linkage: deterministic and probabilistic. The deterministic method establishes whether a pair of records is a link based on a set of given conditions with a determined outcome (match or non-match). Its efficiency, measured as the number of linked records, is limited by incorrectness and incompleteness of the information to be linked. The probabilistic method assigns to each pair a probability of being a link. It uses the approach described by Fellegi and Sunter (Fellegi and Sunter 1969; Scanu 2003) and can be implemented by the software *RELAIS* (REcord Linkage At IStat). The probabilistic approach has been widely used for integrating different data sources to enlarge the analysis and the comprehension of a given phenomenon (Tuoto *et al.*, 2015; Tuoto *et al.*, 2014; Tuoto *et al.*, 2012).

The novelty of this contribution is in the process that drives the linkage activities and the roles played by the data owners. The linked data archives are the Inail (Italian National Institute for Insurance against Accidents at Work) archive of commuting to (*in itinere*) or at work road accidents and the Istat (Italian National Institute of Statistics) archive of road accidents.

Record linkage creates a new dataset where each record ideally should refer to the same unit contained in both the input datasets. The new record has a complete set of information that is, most importantly, related.

The phenomena of road accidents is an example about how complementary information collected by different Authorities can be integrated to get insights. The Italian National Institute of Statistics - Istat registered road accidents that caused persons to die or to be injured. These road accidents have also an occupational origin. Workers use vehicles both for commuting (home-work travelling routes) and for their work (*e.g.* in the transport sector). Although the Istat archive contains pieces of information about the occupational origin, it is often incomplete or unfilled by police officers, and consequently the rate of road occupational accidents cannot be assessed from this dataset. The work-related component of road accidents is instead recorded by Inail (Italian National Institute for Insurance against Accidents at Work) being it an occupational accident. However, not all work-related road accidents are associated with a request for insurance compensation, particularly for those occurring during commuting. Reasons might be due to lack of knowledge about the possibility to claim for compensation, lack of time available to manage with administrative procedures or unregistered accidents. This can underestimate the overall work-related phenomena.

The interconnection of the two road accidents archives can provide advantages, not only in adding occupational information at each accident linked, but also in assessing the efficiency of Istat dataset in registering the occupational component. It can also provide advantages in identifying the unlinked records, potential work-related accidents registered by Istat but not included in the Inail, work-related, archive, with the aim to estimate the unclaimed work-related road accident phenomena. Brusco *et al.* (2019) earlier linked the two archives using a deterministic approach for road accidents occurred in Italy in the year 2015. They found a record linkage efficiency of about 23% of the number of records contained in the Inail archive, addressing the possible unmatched accidents with inaccuracies in the registration systems. To increase the efficiency a coupled deterministic and probabilistic approach could be used.

To test and verify this hypothesis, this work describes the integration of the above road accidents archives by means of a combined deterministic and probabilistic record linkage approach applied to data collected from the year 2014 to the year 2018.

2. Materials and methods

2.1 The datasets

The Inail occupational accidents data archive covers about 80% of the Italian workforce. Inail receives compensation claims applications for occupational injuries from all over the national territory, regarding all workers except for some categories (armed forces, firefighters and police workers, air transport personnel, autonomous tradespeople and professionals with VAT registration). The archive contains information about time and location of the accident, the economic branch of the victim, the occurrence on duty or during commuting and health consequences of the accident, such as body part injured, type of injury and health effects. The Inail dataset is made of the insurance claims of workers for road accidents during work or commuting between 2014 and 2018.

Data about road accidents are routinely collected by Istat to produce statistical reports on this phenomenon on the basis of data recorded from Local Authorities ("*Carabinieri*", Motorway Police, and Local Police). Data refer to road accidents in which an injury or a fatality occurred, involving at least one vehicle, on the public roads of the national territory, occurred between 2014 and 2018. The archive contains road and vehicle-related information of the accident, road and weather conditions, road signs and crossings presence, time, location and geographical coordinates of the accident.

2.2 Pre-processing of the datasets

Unavoidably, data coming from different sources need some pre-processing before they are usable in a linkage model. Data must be recorded and stored in the same way in order for the units to be compared. The procedure required a well-structured technique and different pre-processing steps.

2.2.1 From the accident to the person in the Istat dataset

The records contained in the Inail archive refer to a single individual, while the unit recorded in the Istat dataset is an accident, involving more individuals. The road accidents Istat database is structured according to four different dimensions: Accident, Road, Traffic Unit and Person. The structure used by Istat is the same recommended and implemented by the European Commission, in the CARE database (Community database on road accidents resulting in death or injury) (European Commission 2021). Following the 93/704/EC: Council Decision of 30 November 1993 on the creation of a Community database on road accidents, Italy updates, every year, road accidents data at national level.

Since the Istat dataset represents the accident, each record of the Istat dataset was required to be transformed into *n* records, each one representing a person involved in the accident. This was possible because personal data, for each person involved in the accident, was present in the accident record. Each record contained person identification data (name and surname) and attributes (*e.g.* gender, age) for each person involved in the accident.

To provide a collection of records referred to each person involved in the accident and harmonised with the Inail data set, the authors built a mirror database containing all injured and dead, excluding the unharmed drivers, identified by a seven-digit code. The first six digits identifies the road accident in the main database and the seventh digit corresponds to the ID of the person involved in the accident. In our case, name and surname are guaranteed to refer to the same person, whereas age is not, since the form filled by authorities at the time of the accident records but does not associate ages with persons.

This introduces an error in the form of a lower recall; persons with age wrongly associated can be excluded from the linked pairs, since a difference on a single field can be enough to prevent a link. A reduction of the accuracy is less likely since a single field wrongly filled is less likely to make the difference, alone, in forming a link. There is no bias since there is no kind of records affected more than others. It is useful to point this out since it can be a source of improvement in data collection. Time and location are guaranteed to be correct, being unique for the accident.

2.2.1 Data cleaning and formatting

The second pre-processing step was to ensure that the information contained in corresponding columns in the two archives was represented in the same way. Treatment and formatting were applied where needed. The following operations were carried out:

1. name and surname were converted into uppercase.
2. blank spaces were removed.
3. non-alphabetic characters were removed.
4. letters with signs on or above them were converted to their simple A-Z counterpart.
5. age was made into an integer.
6. a key column (a unique identifier for records in the dataset) was created if not already present.
7. date was put into a unique format (*e.g.* dd/mm/yyyy) and constructed where missing.
8. hour was made into an integer.
9. it was ensured that municipality codes belonged to the same yearly classification, and they were converted to the same yearly classification where that was not the case.
10. a new column was created containing the concatenation of surname and name (in some cases surname and name were both stored in the same column).
11. equal columns were given equal names.

Tables 2.1 and 2.2 show the variables contained in each dataset in their final form.

Table 2.1 - Inail dataset

Personal data			
Case code	Name	Surname	Age
Gender	Country of birth		
Site			
Macro-region	Region	Province	Municipality
Time			
Year	Month	Day (of month)	Date
Hour	Day (of week)		
Injury			
Outcome (injured/dead)		Type of injury	Body part injured
Work			
Economic activity		Commuting (yes/no)	
Insurance			
Insurance management group	Large tariff group (a)	Compensation type	Compensated days
Assumed grade of impairment		Actual grade of impairment	

Source: Authors' processing on Inail dataset
(a) The large tariff grouping that groups the tariff items, which associate the work with the premium rate.

Table 2.2 - Istat dataset

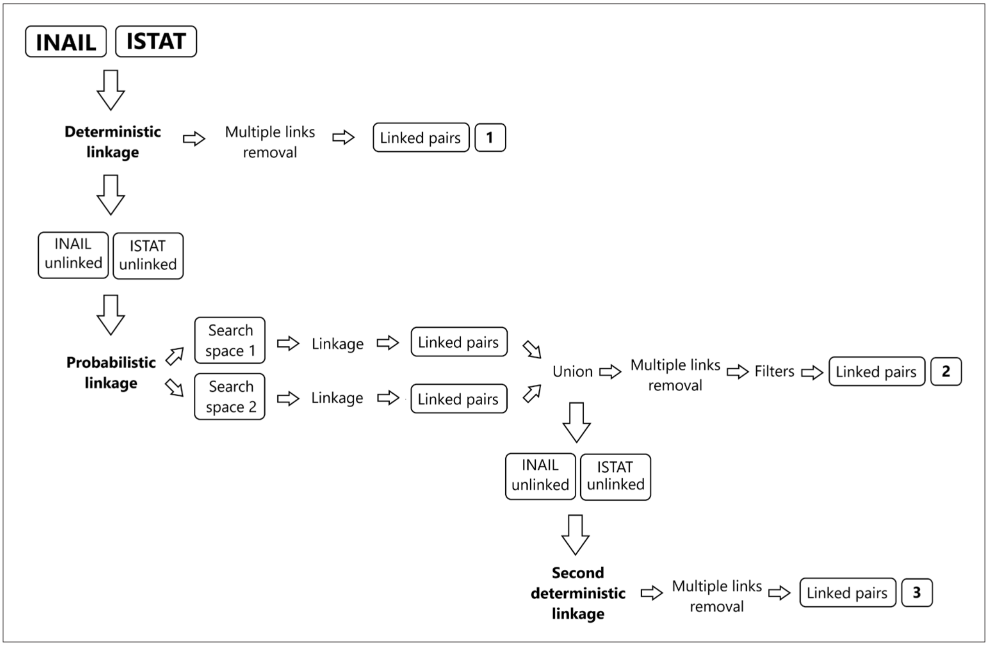
Personal data			
Name	Surname	Age	Gender
Driving license type	Role (driver/passenger/pedestrian)		
Site			
Province	Municipality	Locality	Coordinate type
Projection system	Latitude	Longitude	
Time			
Year	Month	Day (of month)	Hour
Minutes	Day (of week)	Period of day	
Injury			
Outcome (injured/dead)			
Work			
Professional condition (at work/commuting)			
Road			
Road identification code	National road or motorway section	Progressive mileage (Km)	Hectometric
Type of road	Pavement	Road-bed	Weather
Traffic signs	Junction / Non-junction	Localisation of the accident	
Hospital			
Name of the Hospital			
Vehicle			
Type of vehicle involved	Vehicle cylinder capacity	Vehicle license plate	
Accident			
Accident identification number		Road accident type	
Number of people dead in the accident		Number of people injured in the accident	

Source: Authors' processing on Istat dataset

2.3 The linkage procedure

The record linkage was carried out following a step procedure described in the following sub-sections. Figure 2.1 summarises the whole process.

Figure 2.1 - Linkage process



Source: Authors' processing on Inail and Istat datasets

2.3.1 Variables selection

In order to link records, the first step is to identify which variables can be used to perform the match. These variables must necessarily be looked for among those in common between the two datasets. The chosen variables are shown in Table 2.3.

Table 2.3 - Common variables

Personal data			
Name	Surname	Age	Gender
Site			
Province	Municipality		
Time			
Year	Month	Day (of month)	Date
Hour	Day (of week)		
Injury			
Outcome (injured/dead)			

Source: Authors' processing on Inail and Istat datasets

As privacy is concerned with such information, we applied procedures to prevent access to information by unauthorised users. After data linkage, data about individuals were removed for privacy reasons in compliance with the law.

2.3.2 Deterministic linkage

The first linkage operation consisted of a deterministic linkage. Records were matched by surname, name, date and municipality. Records reporting the same surname, name, date and municipality were undoubtedly considered representing the same accident. Clerical or data entry errors can produce a lower recall, rarely a lower accuracy, since it is more likely that two equal surnames are made different by an error than two different surnames are made equal by an error. For surname and name, their concatenation was used, so the equality condition was on their concatenation. A second check was performed in order to include accidents that were mistakenly recorded with name and surname inverted.

2.3.3 Multiple links removal

After this step, a check for multiple links was performed. Multiple links are present when a unit in one dataset is linked with more than one unit in the other. For example, it can happen if in one dataset the same accident is recorded twice, or in case of homonyms. Duplicates must then be removed and, among them, just one linked line must be kept since we want a 1:1 linkage. The key column we introduced in the previous step finds here one of its uses: if duplicates are present in a key column in the linked table, then that record has been linked to more than one record of the other dataset. Table 2.4 exemplifies this event:

Table 2.4 - Example of multiple links

Dataset A key	[other A columns …]	Dataset B key	[other B columns …]
001	[other A data …]	054	[other B data …]
001	[other A data …]	055	[other B data …]

Source: Authors' processing on Inail and Istat datasets

In this case, the element with key ‘001’ of dataset A matches two elements of dataset B. To select one link among the multiple links, equality of other common variables, like age and hour of the accident, were used, to discern which individual of dataset B represents the same accident of the record of dataset A.

2.3.4 Unlinked records archives

This step creates two sets of unlinked records. They are composed of the records remained unmatched, for each input dataset. The sets of unlinked records are the set difference between the input datasets and the linked dataset. The created key column finds here another use: the unlinked records of an input dataset are all of its elements whose key is not present in the linked dataset. A simple count check ensures that the operation has been correctly performed: the number of elements of the input dataset must be equal to the number of linked elements plus the number of the unlinked records. The datasets of unlinked records created in such a way were the input datasets for the next step.

2.3.5 Probabilistic linkage

The third step is to perform the probabilistic linkage. The input datasets are the datasets of unlinked records created in the previous step. The probabilistic linkage was performed according to the theory proposed by Fellegi and Sunter (Fellegi and Sunter 1969; Scanu 2003) and was run by the Istat software *RELAIS* (REcord Linkage At IStat). The Fellegi and Sunter theory is internationally recognised as the reference theory in record linkage (Christen, 2012*a*; Herzog, 2007), in particular its implementation due to Jaro (Jaro 1989). The size of the data processed in this paper required a reduction of the computational space for computational treatability; hence, a comparison space was created, using the Sorted Neighbourhood algorithm (Christen, 2012*b*) with a window size of 50. The Sorted Neighbourhood algorithm lists the elements of the two datasets in a single list, then sorts them according to a sorting variable. Then a fixed size window runs on the sorted list and all the pairs falling into the window are considered candidate pairs (Hernandez, 1995). This procedure was run twice, creating two sets of candidate links, one sorting on surname and name and the other sorting on the concatenation of surname and name. Then the same linkage model was applied to both of them; linked pairs were then unified in one dataset, taking the intersection and the set differences. The linkage model uses as matching variables surname, name, date and age;

for the latter a window of size 1 was considered for comparison, admitting a difference of one year in the ages of the compared records. The linkage model declared as matches the candidate pairs with a posterior linkage probability higher or equal than 0.8, and possible matches, to be reviewed by manual checks, those pairs with a posterior linkage probability in the range [0.5-0.8). At the end of this step, multiple links were removed and datasets of unlinked records calculated, as described in paragraphs 2.3.3 and 2.3.4.

2.3.6 Filtering

The dataset resulting from the probabilistic linkage is a dataset of linked pairs that we can inspect to elaborate filters (rules for clerical review and selection). The aim is to further increase the chances that the identified paired records refer to the same accident. Such a task reduces error tolerance in the matching procedure. For example, if date is different, what is the extent of the difference? If the extent is small and other fields match, can it be taken as a recording error, especially if data were collected manually?

The following filters were thus elaborated and run on the linked pairs. Filters were applied in succession; so filter 2 was applied on the pairs that did not pass filter 1, filter 3 was applied on the pairs that did not pass filter 1 and 2 and so on. Filters are expressed in the form of logical conditions that must all hold true to accept the pair.

Equal date, equal concatenation of surname and name, hour with difference not greater than one, age with a difference not greater than one.

1. On the possible matches only: equal date, equal municipality.
2. Equal date, concatenation of surname and name with a Levenshtein distance (edit difference) not greater than one, hour with a difference not greater than one, age with a difference not greater than one.
3. Equal date, equal surname, one name contained in the other, hour with a difference not greater than one, age with a difference not greater than one.
4. Equal date, equal name, one surname contained in the other, hour with a difference not greater than one, age with a difference not greater than one.

The constraint on date was then loosened: dates with a difference of one in day or month, in presence of an equality on other fields, is interpretable as a filling error (*e.g.* 07/03/2015, 07/04/2015). Having loosened the constraint on the date, the constraint on municipality was tightened, requiring the equality. It would be reasonable to assume that same age homonyms with an accident at the same time of the day are the same accident, but pairs with two inaccuracies (date and municipality) were chosen not to be included, which in any case resulted a scant minority.

5. Date with an inaccuracy of one, equal municipality, equal concatenation of surname and name, hour with a difference not greater than one, age with a difference not greater than one.
6. Date with an inaccuracy of one, equal municipality, concatenation of surname and name with a Levenshtein distance not greater than one, hour with a difference not greater than one, age with a difference not greater than one.
7. Date with an inaccuracy of one, equal municipality, equal surname, one name contained in the other, hour with a difference not greater than one, age with a difference not greater than one.
8. Date with an inaccuracy of one, equal municipality, equal name, one surname contained in the other, hour with a difference not greater than one, age with a difference not greater than one.
9. Date with an inaccuracy of one, equal municipality, equal anagram of the concatenation of surname and name (explained as typo errors, *e.g.* letter inversion), hour with a difference not greater than one, age with a difference not greater than one.
10. Date with an inaccuracy of one, equal municipality, equal concatenation of surname and name, age with a difference not greater than one.
11. Equal date, equal municipality, equal surname, equal first three letters of name.
12. Equal concatenation of surname and name, equal municipality, date with a difference of at most three days.
13. Equal date, equal municipality, concatenation of surname and name with a Levenshtein distance not greater than one, age with a difference not greater than one.

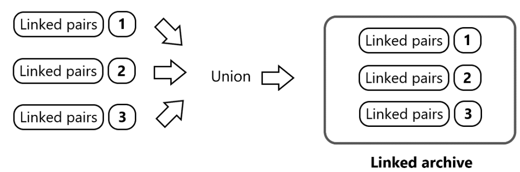
2.3.7 Second deterministic linkage

Given the filters that the probabilistic linkage allowed us to elaborate it seemed a natural extension to try to apply those filters to the unlinked records of the probabilistic linkage to try to get some other linked pairs. This step executes a deterministic linkage on the unlinked records of the probabilistic linkage applying the filters described in paragraph 2.3.6 as linking conditions. Multiple links removal was then performed.

2.3.8 Integration of the partial linked datasets

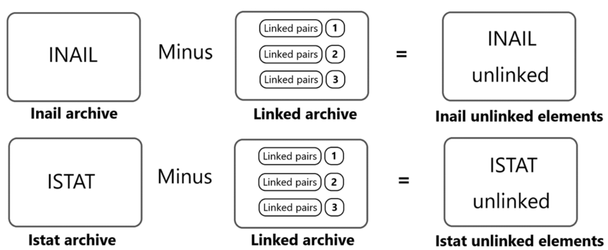
The outputs of the linkage steps described above (partial linked datasets) were then unified to form the final linked road accidents archive. Figure 2.2 sketches the partial linked sets integration. The information to identify which step linked the pair was added to the results, as a metadata, for documenting and improving future data collection and processing. Datasets of unlinked records were calculated at the end of this step by performing a set difference between the starting datasets and the linked dataset. Figure 2.3 sketches the unlinked records calculation.

Figure 2.2 - Union of linked pairs



Source: Authors' processing on Inail and Istat datasets

Figure 2.3 - Calculation of unlinked records



Source: Authors' processing on Inail and Istat datasets

Table 2.5 shows the structure of the linked dataset. The Inail occupational information integrate the Istat accident information.

Table 2.5 - Linked dataset

Personal data			
Age	Gender	Country of birth	Driving license type
Role			
Case			
Inail case code	Istat accident identification number		
Time			
Year	Month	Day (of month)	Date
Day (of week)	Hour	Minutes	
Site			
Macro-region	Region	Province	Municipality
Locality	Coordinate type	Projection system	Latitude
Longitude			
Road			
Road identification code	National road or motorway section	Progressive mileage (Km)	Hectometric
Type of road	Pavement	Road-bed	Weather
Traffic signs	Junction / Non-junction	Localisation of the accident	
Accident consequences			
Outcome	Number of people dead	Number of people injured	Road accident type
Type of injury	Body part injured	Compensation type	Compensated days
Assumed grade of impairment	Actual grade of impairment		
Work			
Commuting	Professional condition		
Vehicle			
Type of vehicle involved			
Economic area			
Economic activity	Large tariff group	Insurance management group	
Linkage info			
Linked by (algorithm step)			

Source: Authors' processing on Inail and Istat datasets

3. Results

Table 3.1 lists the number of accidents registered in each archive.

Table 3.1 - Number of records by archive and year

Archive	2014	2015	2016	2017	2018
Inail	93,056	91,418	93,243	93,673	94,553
Istat	254,528	250,348	252,458	250,128	246,253

Source: Authors' processing on Inail and Istat datasets

Table 3.2 shows the number of linked records by linkage step.

Table 3.2 - Number of linked records by linkage step

Step	Linked pairs	Overall percentage (%)
Deterministic	107,130	83.15
Probabilistic	20,169	15.65
Second deterministic	1,538	1.20
Total	**128,837**	**100.00**

Source: Authors' processing on Inail and Istat datasets

An increase of 18.83% is observed when the probabilistic linkage is applied after the deterministic approach (ratio: 20,169/107,130 = 0.1883).

Table 3.3 shows the number of linked records compared to the size of each dataset.

Table 3.3 - Number of linked records over datasets records

Year	Inail	Istat	Linked	Linked/Inail (%)	Linked/Istat (%)
2014	93,056	254,528	25,383	27.0	10.0
2015	91,418	250,348	24,824	27.0	10.0
2016	93,243	252,458	26,047	28.0	10.0
2017	93,673	250,128	25,872	28.0	10.0
2018	94,553	246,253	26,711	28.0	11.0
Total	**465,943**	**1,253,715**	**128,837**	**26.7**	**10.3**

Source: Authors' processing on Inail and Istat datasets

The number of linked pairs produced by the probabilistic algorithm before filters were applied was 31,147. Table 3.4 shows the proportions of accepted links by filter.

Table 3.4 - Proportions of accepted links by filter

Filter	Accepted links	Accepted links/Total unfiltered links (%)	Accepted links/Total accepted links (%)
Filter 1	14,651	47.04	72.64
Filter 1 on possible matches	95	0.31	0.47
Filter 2	1,660	5.33	8.23
Filter 3	1,857	5.96	9.21
Filter 4	137	0.44	0.68
Filter 5	485	1.56	2.40
Filter 6	0	0.00	0.00
Filter 7	0	0.00	0.00
Filter 8	0	0.00	0.00
Filter 9	98	0.31	0.49
Filter 10	309	0.99	1.53
Filter 11	546	1.75	2.71
Filter 12	137	0.44	0.68
Filter 13	194	0.62	0.96
Total	**20,169**	**64.75**	**100.00**

Source: Authors' processing on Inail and Istat datasets

4. Discussion

As an assessment of linkage quality, precision and recall are often used as measures of the performance of a record linkage process. Precision is the proportion of matches found by the linkage process that are correct, recall is the proportion of correct matches found by the process over all correct matches. The knowledge of correct matches is obtained either by a manual inspection of the data or from a database with known correct matches. We remind here that the goal of this article is not a performance measure, neither a comparison between different algorithms nor to reduce the computational load on probabilistic linking by first running a deterministic linkage, but to observe the improvement in linked pairs produced by a probabilistic linkage run on the unlinked pairs of a deterministic pass. Nonetheless, we are interested in the process accuracy, as we are interested that the linked pairs are correct, so that the improvement that we observe is an improvement of correct matches.

For the deterministic pass, accuracy is enforced by the nature of the variables selected. Municipalities in Italy are very small territorial entities (more than 7,900 in effect on February 2021). The surface that each one encompasses is thereby very small compared to the whole national territory. Thereby it is extremely unlikely the happening of two different accidents in the same municipality, on the same day, with persons having the same name and surname. Thereby is extremely high the chance that such records are referring to the same accident. Moreover, 98% of the linked records with a recorded hour fall in the same 2-hour class, and hour *was not used as linking variable*.

For the probabilistic pass including filters applied afterwards, we observe that filters 1-4 account for most of the linked pairs (more than 90%). More than 95% of those pairs are localised in the same province or municipalties geographically close, yet *no geographical information was used as linking variable*. This gives us enough confidence on the accuracy of the algorithm. In addition, the equality of the identifying fields makes extremely narrow the chances that they can refer to distinct accidents.

Some questions may arise. Would it have been possible to run multiple probabilistic passes using filter variables? Would it have been possible to obtain the same, or better, result with a differently designed probabilistic or deterministic process? Would it have been possible to obtain the same, or better, result with a more sophisticated deterministic strategy applying only deterministic conditions? Probably, but the goal of this article is not to examine *all possible linkage strategies* and pick the best one, the goal of this article is to examine *this strategy* and how this strategy works when applied to road accidents data. The comparison with other designs can then be an open point for further research.

Concluding remarks

This study is part of the Italian National Statistical Programme 2017-2019 (IST-02463 - Analysis of social and health aspects tied to the road accidents phenomenon through Record Linkage with other information sources). Results show that the probabilistic algorithm produced a significant increase in linked pairs compared to the sole use of a deterministic approach. Deterministic filters add further accuracy, in addition to the probability threshold set by the probabilistic algorithm.

To the best of our knowledge, this is a rare example of integrating data between different government entities that are both co-owners and co-managers of the project and, in this way, jointly pursue the common objective of enriching the comprehension of a complex phenomenon. Hence, beyond the common purpose, this joint project requires trusted and shared methodology and algorithms for data linkage. In this work, this is guaranteed by an open-source solution for record linkage, designed and maintained by Istat with the purpose of allowing Italian public entities to apply top-level linkage methodologies even without long experience in record linkage and familiarity with sophisticated statistical modelling. Open-source solution guarantees shareability and, most of all, trustworthiness of procedures and results, particularly relevant in this context, due to the double ownership of the involved data.

The application of record linkage techniques between Istat and Inail archives is useful to join pieces of information, providing added value, enhancing the potentialities of data with different origin and filling information gaps. It points out aspects that only through a joint analysis can stand out and that can be essential for injury prevention. The joint analysis can extract additional information to better characterise the accidental event and infer potential associations with risk factors (Bruzzone *et al.*, 2021; Gariazzo *et al.*, 2021; Pireddu *et al.*, 2021).

This characterisation of the accidents and the related work profiles can help to define risk prevention programmes with a resulting mitigation of the phenomenon and its impact on society and public health.

References

Brusco, A., A. Bucciarelli, M. Bugani, C. Gariazzo, C. Giliberti, M. Marinaccio, S. Massari, A. Pireddu, L. Veronico, G. Baldassarre, S. Bruzzone, M. Scortichini, M. Stafoggia, e S. Salerno. 2019. *Gli incidenti con mezzo di trasporto. Un'analisi integrata dei determinanti e dei fattori di rischio occupazionali*. Roma, Italy: Inail.

Bruzzone, S., A. Altimari, G. Baldassarre, R. Boscioni, L. Veronico. 2021. "Work-related road accidents: an in-depth statistical analysis carried out by two different integrated data sources". In this issue of the *Rivista di statistica ufficiale*, N. 3/2021. Roma, Italy: Istat.

Christen, P. 2012*a*. "The data matching process". In Christen, P. *Data matching. Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*: 23-35. Heidelberg, Germany: Springer, *Data-Centric Systems and Applications*.

Christen, P. 2012*b*. "A Survey of Indexing Techniques for Scalable Record Linkage and Deduplication". *IEEE Transactions on Knowledge and Data Engineering*, Volume 24, Issue 9: 1537-1555.

Fellegi, I.P., and A.B. Sunter. 1969. "A theory for record linkage". *Journal of the American Statistical Association*, Volume 64, Issue 328: 1183-1210.

Gariazzo, C., A. Marinaccio, S. Bruzzone, L. Taiano, and L. Veronico. 2021. "Work-related road accidents: a statistical multivariate analysis in Italy". In this issue of the *Rivista di statistica ufficiale*, N. 3/2021. Roma, Italy: Istat.

Hernandez, M.A., and S.J. Astolfo. 1995. "The merge/purge problem for large databases". *ACM Sigmod Record*, Volume 24, Issue 2: 127-138.

Herzog, T.N., F.J. Scheuren, and W.E. Winkler. 2007. *Data Quality and Record Linkage Techniques*. New York, NY, U.S.: Springer Science+Business Media, LLC.

Istituto Italiano di Statistica - Istat (*Italian National Institute of Statistics - Istat*). "RELAIS (REcord Linkage At IStat)". *Methods and Tools*. Roma, Italy: Istat. https://www.istat.it/en/methods-and-tools/methods-and-it-tools/process/processing-tools/relais.

Jaro, M.A. 1989. “Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida”. *Journal of the American Statistical Association*, Volume 84, N. 406: 414-420.

Pireddu, A., A. Altimari, G. Baldassarre, A. Marinaccio, and L. Taiano. 2021. “An analysis on work-related road injuries by macro-economic sector, road type and Italian territorial divisions”. In this issue of the *Rivista di statistica ufficiale*, N. 3/2021. Roma, Italy: Istat.

Scanu, M. (*a cura di*). 2003. “Metodi statistici per il record linkage”. *Metodi e Norme*, N. 16. Roma, Italy: Istat.

Tuoto, T., S. Bruzzone, L. Valentino, G. Baldassarre, N. Cibella and M. Pappagallo. 2012. “Towards an integrated surveillance system of road accidents”. In Società Italiana di Statistica - SIS. *Atti della XLVI Riunione Scientifica (Roma, 20-22 giugno 2012)*. Padova, Italy: CLEUP.

Tuoto, T., A. Burgio, R. Cotroneo, C. Iaccarino, S. Prati, F. Rinesi, F. Rottino, L. Tosco, e L. Valentino. 2015. “Fecondità e maternità: un sistema integrato per la misurazione di fenomeni sanitari e socio-demografici”. *Rivista di statistica ufficiale*, N. 3/2015: 49-70. Roma, Italy: Istat. https://www.istat.it/it/archivio/187424.

Tuoto, T., L. Corallo, N. Cibella, D. Ichim, V. Mastrostefano, A. Nurra, e M. Rinaldi. 2014. “L’integrazione dei risultati delle indagini sulla tecnologia e l’innovazione nelle imprese: una sperimentazione”. *Rivista di statistica ufficiale*, N. 3/2014: 97-128. Roma, Italy: Istat. https://www.istat.it/it/archivio/152014.

Work-related road accidents: an in-depth statistical analysis carried out by two different integrated data sources

Silvia Bruzzone [1], Antonella Altimari [2], Giordana Baldassarre [1], Roberto Boscioni [2], Liana Veronico [2]

Abstract

The aim of the paper is to provide an integrated overview on data included in Road accidents Istat (Italian National Institute of Statistics) register and Inail (Italian National Institute for Insurance against Accidents at Work) archive of work-related road accidents. The data sources available for the surveillance of road accidents have relevant limits, when taken separately. Data integration, performed by record linkage techniques, focusses on the enrich information associated to each road accident. Mainly, information on the circumstances of accident and characteristic of roads and vehicles are linked to information on professional condition, economic sector, type of injury and severity at individual level. Therefore, the integration of data is essential in order to build up a detailed picture to drive preventive actions.

Keywords: Road accidents, work-related road accidents, record linkage, compensation data.

1 Istituto Nazionale di Statistica (*Italian National Institute of Statistics*) - Istat.

2 Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro (*Italian National Institute for Insurance against Accidents at Work*) - Inail.

In the writing of this article, the authors were supported by all the members of the Inail-Istat collaborative group: Antonella Altimari, Michela Bonafede, Roberto Boscioni, Claudio Gariazzo, Alessandro Marinaccio, Stefania Massari, Antonella Pireddu, Luca Taiano and Liana Veronico (for Inail); Giordana Baldassarre and Silvia Bruzzone (for Istat).
Although this article is the result of all the authors' commitment, the paragraphs are attributed as following: 1 and 2 to Silvia Bruzzone; 3.1 and 4 to Silvia Bruzzone and Liana Veronico; 3.2.1 to Giordana Baldassarre; 3.2.2 to Antonella Altimari, Roberto Boscioni and Liana Veronico; 5 to Giordana Baldassarre and Silvia Bruzzone.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

Road traffic injuries are considered to have a relevant impact on public health. A relevant portion of these road accidents are of occupational origin. Istat (Italian National Institute of Statistics) and Inail (Italian National Institute for Insurance against Accidents at Work) produce the official statistical information on road accidents and work-related road accidents respectively.

Istat is responsible for processing and disseminating the official statistical information on road accidents in Italy. The domain of the survey, carried out by Istat, consists of all road accidents, occurred on the national public roads, resulting in death or injury (deaths within 30 days), detected by the Police and *Carabinieri* authorities.

Inail, based on its own official administrative archives, builds and provides statistical databases and institutional open data on accidents at work. Although these two archives present a reliable picture of both general and work-related road accidents, a complete information is missing. While the Istat archive of general road accidents describes the localisation, road characteristics, vehicles, accident type and weather conditions, the Inail one provides information related to occupational parameters such as the economic sectors, injuries and the length of leave. This means that information about the type of involved vehicle or the type of accident are missing in this archive. The connection between these two types of information would allow further and more detailed analysis of road accidents' cases.

An early pilot study about general and work-related road accidents based on record linkage application was carried out by Istat and Inail using accidents occurred during year 2015 (Brusco *et al.*, 2019). The limited amount of data did not allow having a full description of the subject as well as its time variability.

So far, Istat and Inail disseminated several studies concerning the use of record linkage techniques to add valuable information to the road accident source based on Police records. The main field explored by Istat was the association between the Road Accidents survey and the Causes of Death register at individual level. The aim was to provide a set of integrated information for each dead person in a road accident, such as data on cause and cause of death, the role of the deceased and the dynamic and circumstances of the road accident. In addition, the results highlighted the context factors for the road accident, mainly the social and safety ones, and contributed to the updating of systems of indicators on public health and security (Tuoto *et al.*, 2018 and 2012).

In order to provide a more complete description of work-related road accidents not available in the traditional dissemination of results and enriched in its contents, a record linkage between Istat road accidents data and Inail information on work-related road accidents was performed. The integration of information, mainly, focusses on joining the characteristics of road accidents, dead or injured drivers, passengers or pedestrians and vehicles with data on the professional condition (work/work commuting), economic sector, type of injury and severity at individual level (Taiano *et al.*, 2021)3. The analysis of the record linkage between data sources, in the field of work-related road accidents is included in the Italian National Statistical Programme and authorised by the Data Protection Authority.

3 The classical approach to the record linkage theory is due to Fellegi and Sunter (1969). The *RELAIS* toolkit developed by Istat and used to perform the linkage, applies the Fellegi and Sunter method; the methodological aspects are explained in the article by Taiano, L. *et al.*, "Work-related road accidents: a data linkage procedure applied to assess traffic accidents at work and commuting". In this issue of the *Rivista di statistica ufficiale*, N. 3/2021.

2. Materials and methods

2.1 Integrated databases of work-related road accidents

The present study is based on an integrated work-related road accidents archive obtained from a data linkage procedure better described in an accompanying paper (Taiano *et al.*, 2021), covering events occurring during year 2014-2018. Briefly, starting from the general and work-related road accidents archives, a multiple steps data linkage procedure was applied using deterministic and probabilistic approaches.

Although the full years dataset represents a rich source of information useful to define the time trend of the integrated variables, the study presented in this article concerns a detailed descriptive and in-depth analysis of variables focussed on the year of data 2018 only.

This choice is due to different reasons. First, the data of year 2018 was the last available one, providing updated information, as well as more robustness and completeness, which gradually improved over the time.

The second reason is due to similarity among the distributions across the years, similar in the shape, for the main variables studied, during the five-year period considered. The analysis is mainly based on the percentage frequencies of the linked cases and the most representative variables of the integrated dataset, for each year in the period 2014-2018. A Chi Square (c2) test was applied. The Pearson Test Chi Square (c2), in fact, is the well-known method used to verify the association between the variable "year of event" and the other main variables of the study too. The method states when the observed distribution of data fits with the distribution expected whether the variables are independent or not (See Figure 3.1-3.3, Table 3.1 and Appendix - Table A4).

The application of record linkage techniques between the two databases was particularly useful for enhancing the potential of different types of data and filling some information gaps too. In addition to the statistical examination of the integrated data, picking up the benefit brought by the two sources, a new reading key of the themes was applied too. Finally, the authors introduced a rich appendix supplement.

2.2 Data sources: strengths and limits

The analysis of the integrated information requires a special attention for the results evaluation, in consideration of the differences in definitions and domains between the two examined universes.

The Istat "road accidents survey" collects all road accidents resulting in death (within the 30th day) or injuriy, involving at least a vehicle circulating on the national road net and registered by a Police authority (Convention of Vienna, 1968 and United Nations, 1977; European Union, United Nations, OECD, ITF and Eurostat,, 2019; European Commission, 2021). The survey carried out by Istat, with the cooperation of ACI (Automobile Club of Italy) and other public national institutions, is an exhaustive data collection (included in the National Statistical Programme). The data collection system has been adapted to the local level organisation and needs. Istat adopted a flexible data flow model, through the subscription of special agreements with Regions (NUTS2 level) and Provinces (NUTS3 level) (Eurostat Nomenclature of Territorial Units for Statistics, 2021), to facilitate the local authorities' information needs and to improve the timeliness and quality of data collected.

The new geography of the organisation models, in fact, allowed over the time to have a gradual improving of the coverage of the number of accidents and of the completeness of all information, included the date, the coordinates and the demographic information of individuals, useful to optimise the linkage procedures.

The accidents at work, included the claimed road accidents at work, registered by Inail, consists of all cases occurred during the course of work for violent and external causes that determined temporary (> 3 days) or permanent disability or death (Article 2 DPR n. 1124/1965). Accidents "in commuting", *i.e.* road accidents that occur during the journey between home and the workplace and vice versa are included too (Article 12 DL n. 38/2000). From 12 October 2017, accidents reports include communications made only for statistical and information purposes, with the consequent absence from work of at least 1 day too (Article 18 DL n. 81/2008). Such cases, for the previous years, were included in the deductibles of injury below the 3-day threshold. The reported fatal events, on the other hand, consider all cases for which the death is a consequence of the accident.

The number of injured or deceased successfully linked, from the two Inail and Istat databases, is about 26,000 every year, during the period 2014-2018 examined. This amount represents respectively the 28%, out of the total insurance claims of workers for road accidents during work or work commuting (*in itinere*), and the 10% out of the total number of persons, with injuries or dead, involved in road accidents, detected by Police.

Concerning the percentage of linkage between the sources and the evaluation of the goodness of the integration procedure, it is to notice that the difference in domains dimension of the two data sources implies that not all the accidents, with consequent insurance claims of workers, are included in the general road accidents collection denounced by Police and registered by Istat. The last characteristic is in fact the condition of belonging to the Istat domain. Istat does not count all general road accidents if self-declared by the users to the vehicle's insurance companies without the intervention of a Police authority, even when an injured person is involved.

In addition to that, not all work-related road accidents produce a request of compensation to Inail. These missed registrations occur particularly for commuting road accidents and the amount cannot be counted. Consequently, as for the total amount of the road accidents, we could have an underreport of work-related cases.

In summary, only a portion (28%) of the registered work-related road accidents can be found in the complete road accidents archive, and it represents a portion (injured or fatal events) of the whole domain of road accidents occurred in the country (including accidents with only damages to vehicles and without injured persons). According to the dimension of linked cases (about 26,000 every year in the period considered), we are confident about their ability to describe the studied topics.

3. Statistical analyses of the integrated database of work-related road accidents

3.1 General and work-related road accidents: an overview

As previously stated, the analysis focusses on 2018, the last available for the integrated data. Before introducing the results of the data linkage, and providing a complete picture of the framework, some general information is useful about the total number of the general road accidents and of claims for road accidents at work occurred in 2018.

In 2018, 172,553 road accidents occurred in Italy resulting in death or injury (Istat and Police source), down comparing with 2017 (-1.4%), with 3,334 deaths (within 30 days) and 242,919 injured (-1.6%). The number of deaths decreases to 2017 (-44 units, -1.3%), after the increase detected in the previous year. Among the victims, the number of pedestrians (612, +2%), moped users (from 92 victims up to 108) and trucks occupants (189, +16%) increased. The motorcyclists (687, -6.5%), cyclists (219, - 13.8%) and passenger cars users (1,423, - 2.8%) showed a decrease. Although, the total amount of deaths decreased, the number of casualties on motorways (including ring roads and motorways junctions) increased – from 296 in 2017 to 330 in 2018; +11.5% – due to the accident occurred in Genova on August 14 on the Morandi bridge of the A10 Genova-Savona-Ventimiglia, which involved many vehicles and caused 43 casualties. On rural and built-up-area roads, a decrease of victims was recorded (1,603 deaths; -0.7% and 1,401 deaths; -4.5%). Distraction, failure in observing precedence rules and high speed (40.7% in total) were among the most frequent misbehaviour. The most sanctioned violations of the Road Traffic Act were, indeed, failure to comply with the signs; failures to use safety devices and the use of mobile phone driving as well as high speed (Istat 2019).

Injuries reported to Inail and occurring in 2018, updated to 30 April 2020 (the last date available before the linkage), were just over 645 thousand, slightly down (- 0.2%) compared to 2017. There were 1,264 fatal events reported in 2018, up of 8.9% on 2017. While 14.7% of overall accidents were road accidents, the percentage of fatal accidents was 46.5%.

About 420,000 work-related accidents were positively defined and about 15% were road accidents. Among the 1,264 fatal cases, 60.9% are positively defined and almost 58.2% are fatal events occurring on the road.

The data confirms the high danger potential of the "road risk", the incidence of fatal cases for the accidents at work, out of the total is indeed much higher for road accidents than in non-road accidents, with more serious consequences for the injured too.

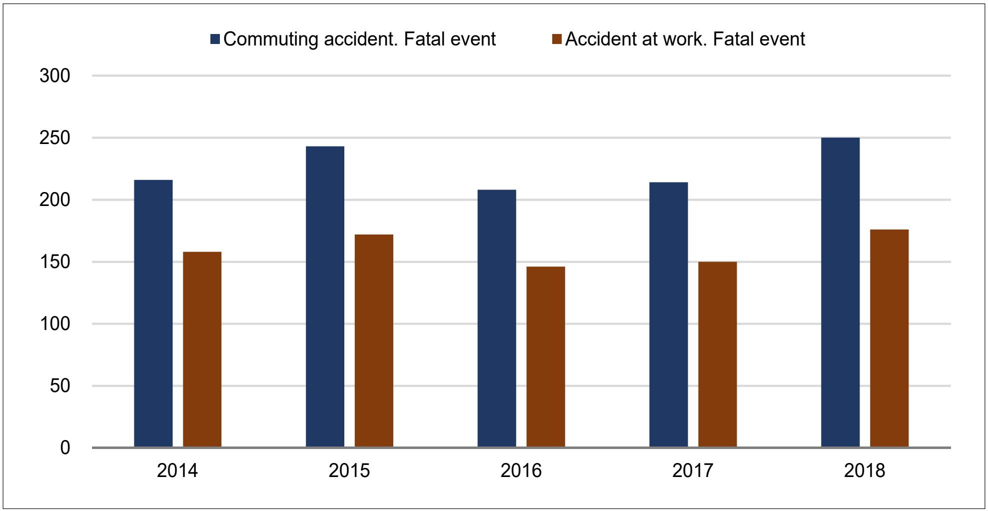
3.2 Work-related road accidents: focus on 2018 integrated data

3.2.1 Description of work-related road accidents by collision variables

The work-related road accidents integrated database was built for five years 2014-2018, where the variable "Year" represents the year in which the accident occurred.

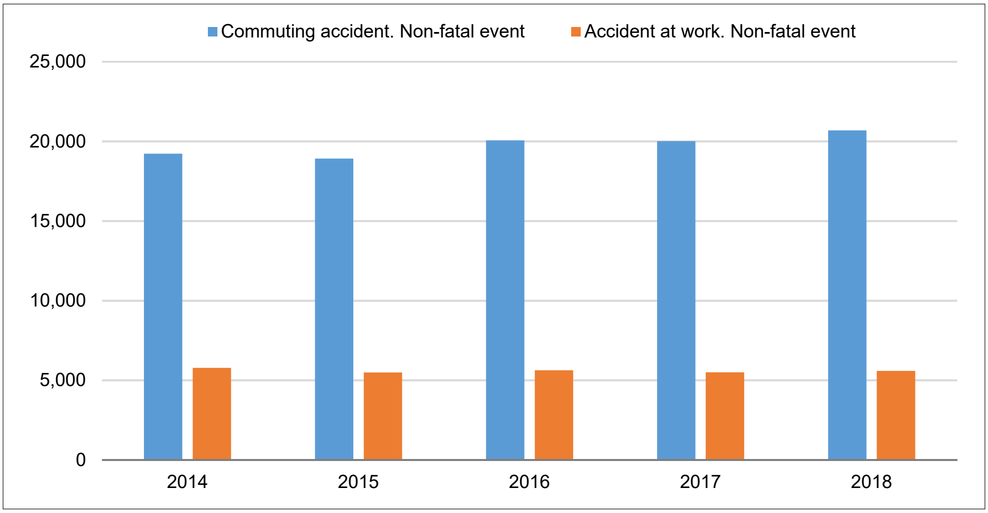
The first figures and tables describe the trend of the whole series, but in the following descriptive analysis, only the last available year, 2018, is considered. (Figures 3.1-3.3, Table 3.1 and Appendix - Table A4).

Figure 3.1 - Fatal injury claims due to road accident by type. Years 2014-2018 (absolute value)



Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Figure 3.2 - Non-fatal injury claims due to road accident type. Years 2014-2018 (absolute value)



Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Trend in the years

The work-related road accidents linked are 26,711 in 2018. This value is stable over the years considered, 2014-2018 with a slight decline in 2015. In 2018, 426 road accidents related to work were fatal and 26,285 were non-fatal (Figure 3.1, Figure 3.2 and Appendix - Table A4).

In the five years considered, 2014-2018, the percentage of non-fatal work-related road accidents has been unvaried, about 98.5%, versus 1.5% for fatal work-related road accidents (Table 3.1).

Table 3.1 - Injury claims due to road accident by event. Years 2014-2018 (absolute value)

Years	Fatal event	Non-fatal event	Total
2014	374	25,009	25,383
2015	415	24,409	24,824
2016	354	25,693	26,047
2017	364	25,508	25,872
2018	426	26,285	26,711

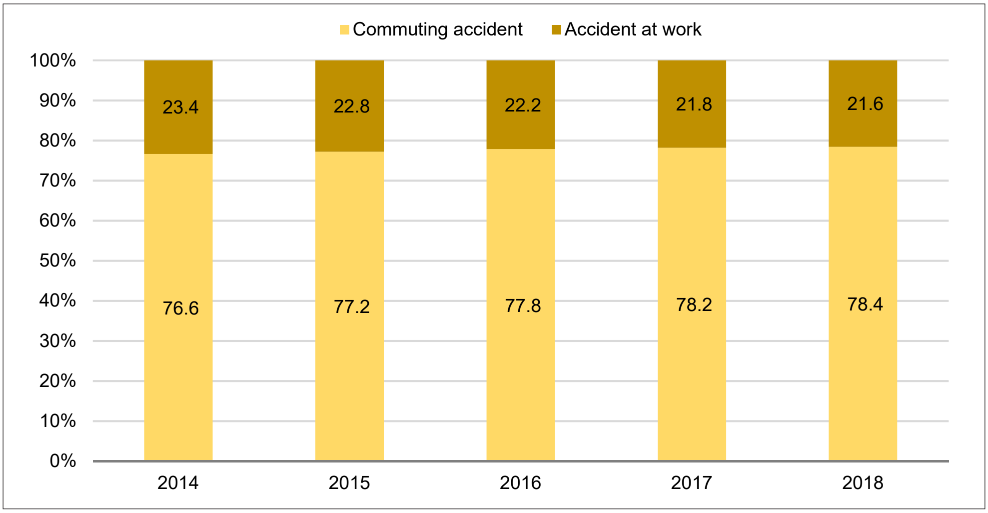
Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Work-related road accidents are divided into two large groups: commuting accidents and accidents at work (Appendix - Table A1).

In 2018, road accidents at work amounted to 5,767 and commuting accident to 20,944.

The injury claims due to road accidents are mostly commuting accidents: they were 78.4% in 2018. Just 21.6% were accidents occurred at work. These values are constant in the recent years: one in five injury claims involves a professional driver (Figure 3.3).

Figure 3.3 - Injury claims due to road accident by type. Years 2014-2018 (percentage value)

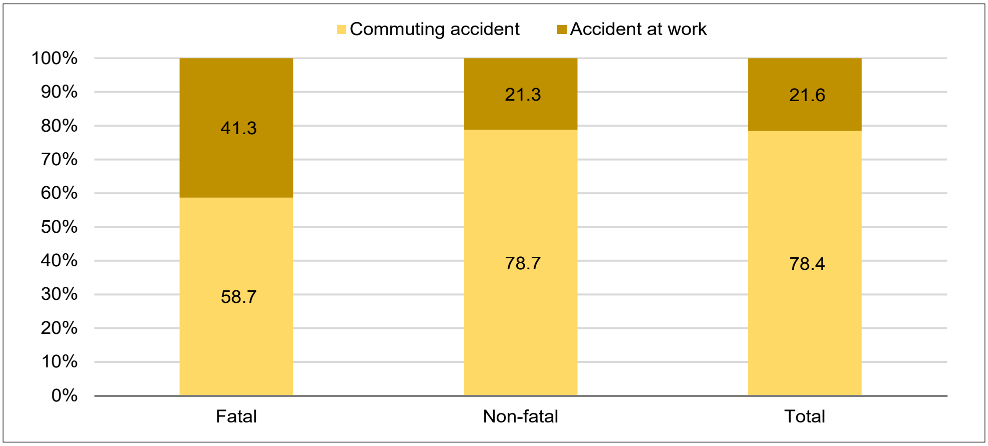


Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Fatal and non-fatal accidents

The 41.3% of fatal accidents occurs at work, so they involve professional drivers (Figure 3.4). Among non-fatal accidents, the percentage of professional drivers is 21.3%. With respect to the total number of accidents, those occurred during work are fatal in 3.1% of the cases, while the percentage of fatal accidents drops to 1.2% if we consider the commuting accidents.

Figure 3.4 - Injury claims due to road accident by event. Year 2018 (percentage values)

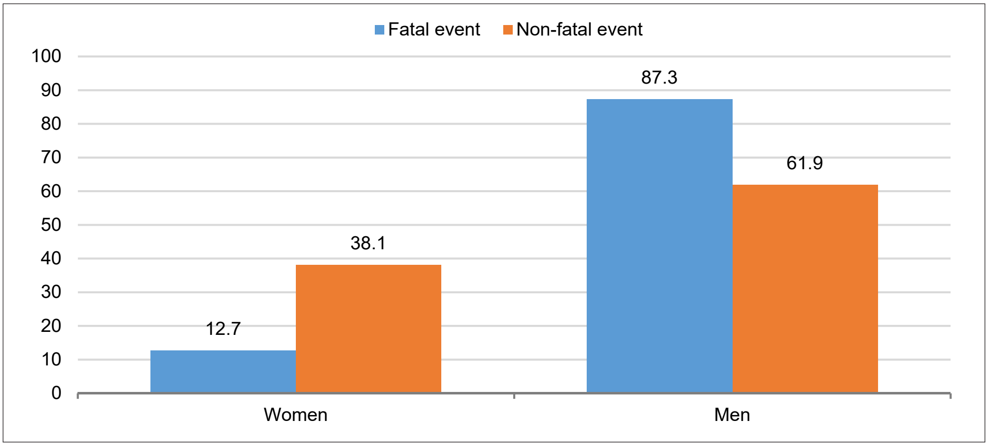


Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Women and men

Injury claims mainly involve men. Work-related road accidents involve in out of 10 cases 6 men (62.3%). This percentage is stable if we consider non-fatal events alone (61.9%) (Figure 3.5). This percentage rises to 87.3% if we consider fatal work-related road accidents, maybe because they are men professional drivers. Non-fatal-work-related road accidents are 38.1% in women.

Figure 3.5 - Injury claims due to road accident by event and gender. Year 2018 (percentage values)



Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Age

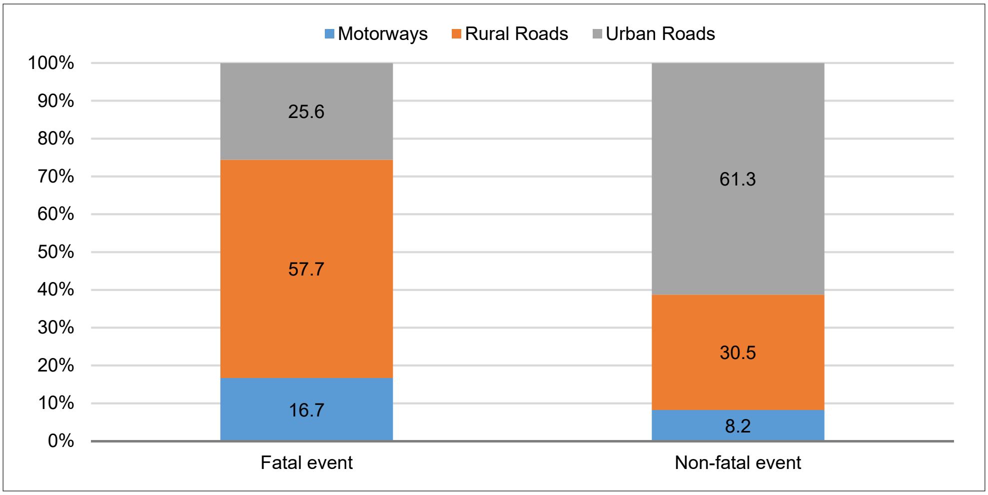
In 2018, 70% of injury claims due to road accident involve people aged 49 and over. This value is the same in non-fatal events. There is a change among the fatal events: 7 out of 10 incidentals occur among people who are 35-64 years old. (Appendix - Table A2).

Road type

Injury claims due to road accident mostly occur on urban roads, about 6 out of 10 cases. In 2018, 16,214 (60.7%) work-related accidents happened on urban roads and 8,270 (31.0%) occurred on rural roads. Considering fatal accidents, the percentage of those happening on rural roads rises to 57.7% (Figure 3.6).

Focussing on the outcome of the accident, non-fatal accidents have the same percentages as overall accidents: 60.7% overall accidents occur on the urban roads, 31.0% on the rural roads and 8.3% on the motorways. There is a difference in fatal accidents. The percentages of accidents on the motorways and on rural roads double considering fatal accidents: 16.7% on motorways and 57.7% on rural roads.

Figure 3.6 - Injury claims due to road accident by localisation of the accident and event. Year 2018 (percentage values)



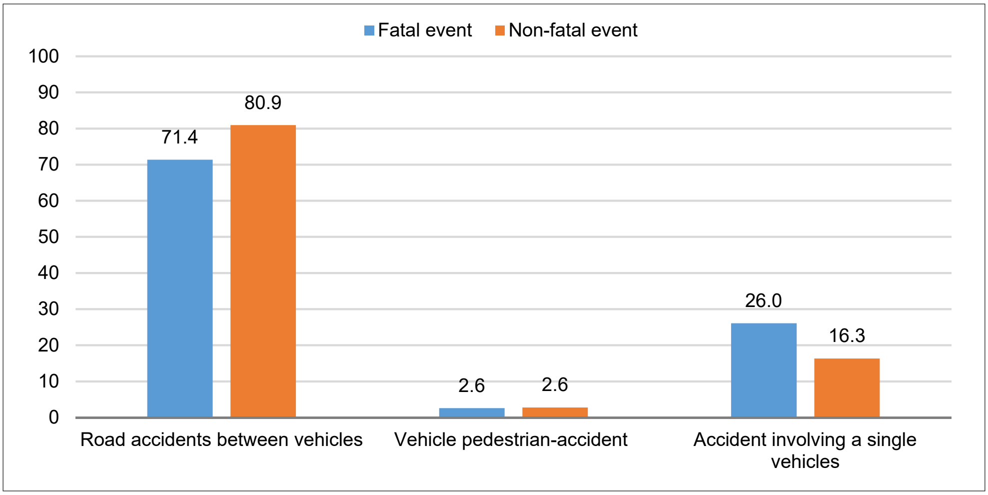
Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Accident type

Work-related road accidents involve at least two vehicles in 80.8% of cases. A very similar value is also found among non-fatal accidents: 80.9% (Figure 3.7).

The percentage of accident involving a single vehicle is 16.5% but among fatal accidents this value is 26.0%.

Figure 3.7 - Injury claims due to road accident by event and road accident group type. Year 2018 (percentage values)

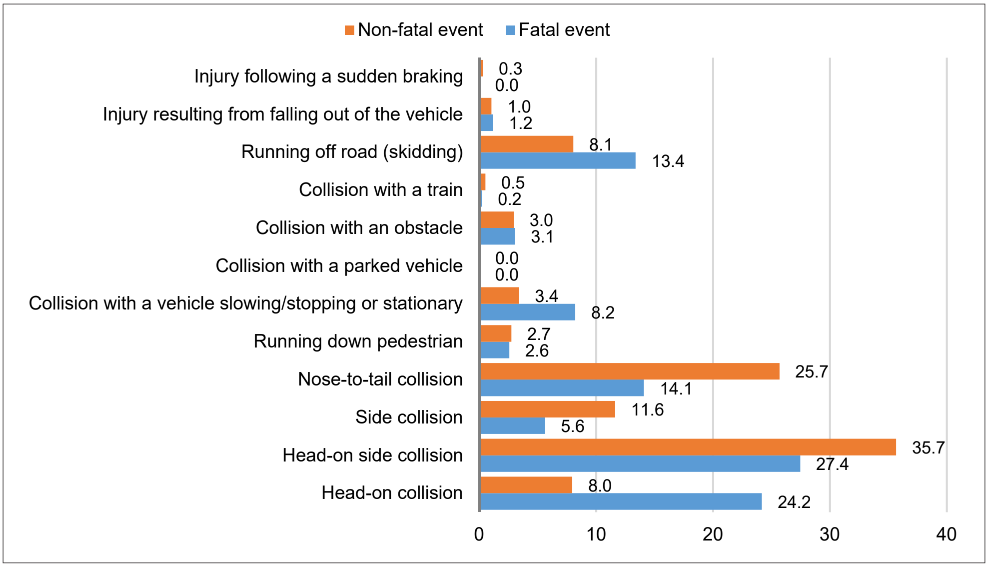


Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

About the road accident type, it is possible to see in detail the various types of accident.

The most common accident type between non-fatal accidents is head-on side collision (35.7%) and nose-to-tail collision (25.7%) (Figure 3.8). Among the fatal accidents the most frequent types are, again, head-on side collision (27.4%) and head-on collision (24.2%). The percentage of fatal accidents due to head-on collision is three times the non-fatal ones. The peculiarity of fatal accidents is that it presents higher percentages than non-fatal accidents for two other types of accident: running off road (skidding) (13.4%) and collision with a vehicle slowing/stopping or stationary (8.2%).

Figure 3.8 - Injury claims due to road accident by event and road accident type. Year 2018 (percentage values)



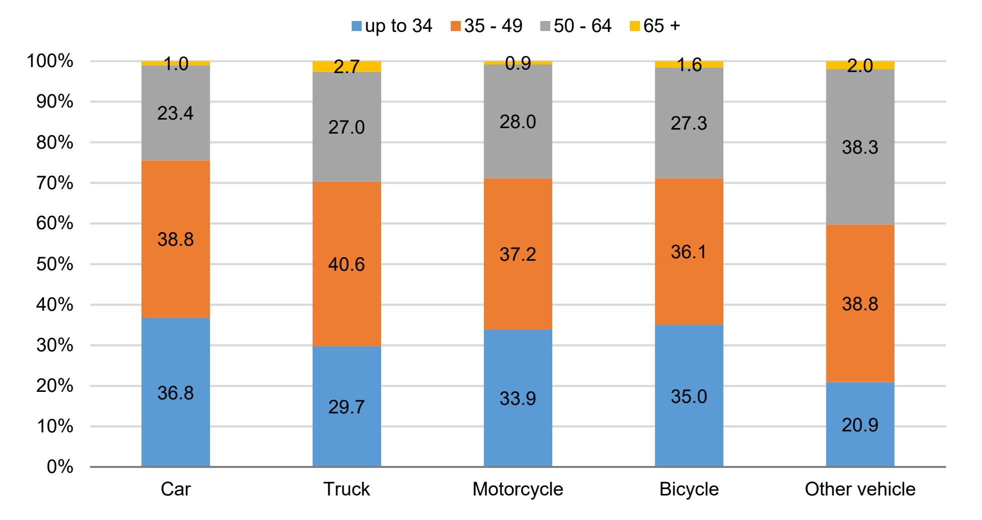
Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Vehicle type

More than half (57%) of injury claims due to road accident occurs to car passengers and 25% to motorcycle passengers (Appendix - Table A3).

People travelling in a car, as a driver or passenger, involved in an injury claims due to a road accident are young with an age of up to 34 years in 36.8% of cases (Figure 3.9). Truck drivers are adults aged between 35 and 49 in 40.6% of cases, like those who travelled on motorcycles (37.2%).

Figure 3.9 - Injury claims due to road accident by age class of users and vehicle type. Year 2018 (percentage values)



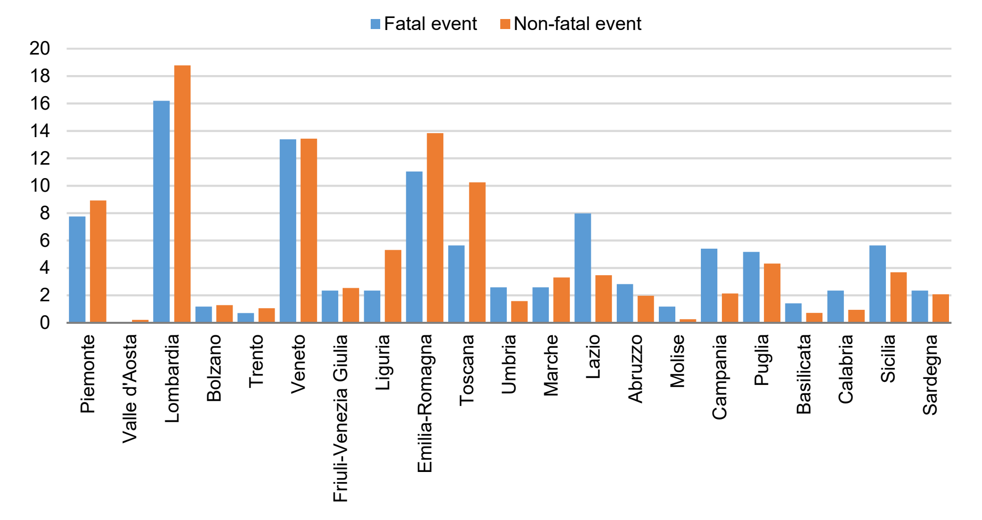
Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Territorial Area

Half of the accidents occur in four Italian regions: Lombardia (18.7%), Veneto (13.4%), Emilia Romagna (13.7%), and Toscana (10.2%) (Figure 3.10).

In the Central and Southern regions, the percentages of fatal accidents are higher than non-fatal ones, with the exception of Toscana and Marche.

Figure 3.10 - Injury claims due to road accident by event type and region. Year 2018 (percentage values)



Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

It is interesting to analyse the situation of work-related road accidents in large geographical areas with respect to the type of traffic signs and the type of road pavement.

In the South and the Islands, the accident rates occurring on roads with absent traffic signs are higher than in the North and Centre. 13.7% in the South and 12.8% in the Islands of accidents occur on unmarked roads compared to 4.3% in the Nord-West, 3.8% in the Nord-East and 6.3% in the Centre (Table 3.2). The presence of double signs, both vertical and horizontal, is lacking in the South and the Islands where only 58.4% of accidents occur on roads with double signs. In the North, the values are higher.

Table 3.2 - Injury claims due to road accident by traffic signs and geographical distribution. Year 2018 (percentage values)

Traffic signs	North-West	North-East	Centre	South	Islands	Italy
Absent	4.3	3.8	6.3	13.7	12.8	6.0
Vertical	4.3	5.5	8.0	18.1	15.7	7.5
Horizontal	8.2	8.7	8.0	8.9	12.0	8.6
Vertical and horizontal	82.6	81.6	76.9	58.4	58.5	77.3
Temporary construction site	0.6	0.4	0.8	0.9	1.0	0.6
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Almost all of the road work-related accidents in Italy occur on paved roads (Table 3.3). It is interesting to see the values, even if low, of the accidents that occur on bumpy paved roads or unpaved roads. In the South, the percentage is higher on bumpy paved roads (0.9%) and in the Centre on unpaved roads (0.4%).

Table 3.3 - Injury claims due to road accident by pavement and geographical distribution. Year 2018 (percentage values)

Paving	North-West	North-East	Centre	South	Islands	Italy
Paved road	99.6	99.6	99.0	99.1	98.9	99.4
Bumpy paved road	0.3	0.3	0.6	0.9	0.8	0.4
Unpaved road	0.1	0.1	0.4	0.0	0.3	0.2
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

3.2.2 Association between road accidents, place and vehicles type, economic categories for compensation and site of injuries

Analysis by vehicle type

An analysis of accidents by type of vehicle involved shows that the majority of fatal events occurred with trucks (3.3%) while all vehicles account in average around 98% in non-fatal events.

In 2018 the accidents involving a private car[4] unfortunately account for the highest number of fatal (over 40%) and non-fatal (around 55%) accidents, followed by those involving a motorbike (around 30% for fatal and just over 24% for non-fatal accidents) (Figure 3.10).

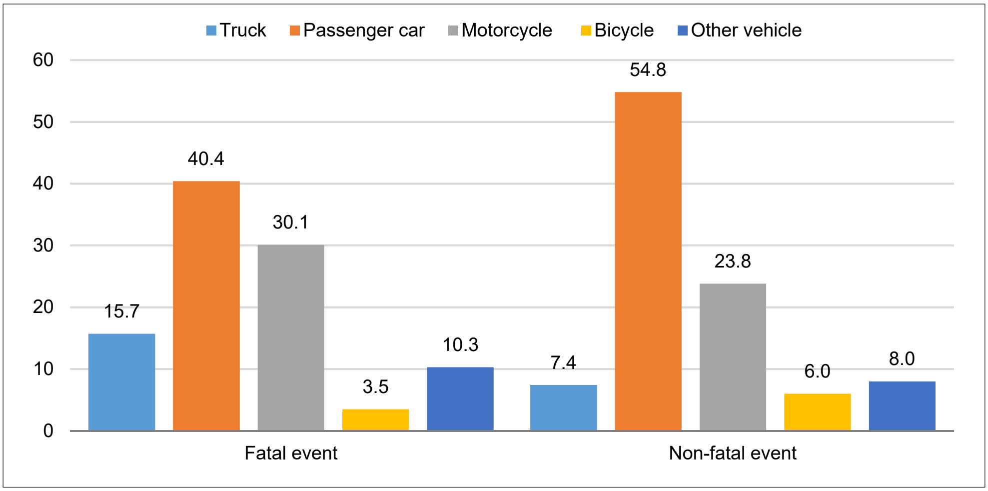
Table 3.4 - Percentage of Injury claims due to road accident by vehicle type and event outcome. Year 2018

Event	Car	Truck	Motorcycle	Bicycle	Other vehicle	Total
Fatal event	40.4	15.7	30.1	3.5	10.3	100.0
Non-fatal event	54.8	7.4	23.8	6.0	8.0	100.0
Total	**54.5**	**7.6**	**23.9**	**5.9**	**8.1**	**100.0**

Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

4 Inail only compensates a commuting accident by private vehicle if the use of the vehicle is necessary, *e.g.* because it is impossible to reach the place of work or it is too far and too long to travel by public transport compared to private transport.

Figure 3.11 - Injury claims due to road accident by event outcome and vehicle type. Year 2018 (percentage values)



Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Analysis by consequences

In the event of an accident at work or occupational disease, Inail provides the worker with the welfare benefits established by law. These benefits are differentiated according to the consequences of the accident in temporary disability, permanent disability or death. In case of absolute temporary inability to work, an indemnity is paid in lieu of salary until clinical recovery; if the accident results in permanent biological damage between 6% and 15% of the psychophysical validity, a capital indemnity is paid for the psychophysical impairment suffered by the worker. If the accident results in permanent biological damage of less than 6%, no compensation will be paid by Inail (so-called excess). If the degree of ascertained impairment of psychophysical integrity is between 16% and 100%, the worker is paid an annuity. Finally, if the worker dies because of an accident at work, an annuity is paid to the survivors. Table 3.5 shows the percentage of compensated injuries by vehicle involved and type of compensation.

Table 3.5 - Compensated injuries due to road accident by vehicle type and type of compensation. Year 2018 (percentage values)

Type of compensation	Car	Truck	Motorcycle	Bicycle	Other vehicle	Total
Lump sum	31.1	6.0	43.8	8.9	10.2	100.0
Survivor's pension	39.3	15.6	26.8	4.0	14.3	100.0
Direct pension	32.9	7.3	43.0	4.2	12.6	100.0
Temporary	58.6	8.9	19.2	5.5	7.8	100.0
None	56.0	5.5	25.1	6.1	7.3	100.0
Total	**54.5**	**7.6**	**23.9**	**5.9**	**8.1**	**100.0**

Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Two out of three of all road accidents are positively recognised and compensated. The number of men (workers) who receive compensation is higher than the number of women (workers) (64% against 36%) regardless of the type of vehicle involved, while women (workers) receive more compensation than their men colleagues when the vehicle involved is a private car (50.5% compared with 49.5%).

Private cars and motorbikes are the means of transport with the highest number of fatal road traffic accidents and consequently the highest number of survivors' pensions (39.3% and 26.8% respectively).

In 43% of motorbike accidents, a permanent disability and hence a direct pension is awarded, even if the disability does not exceed 20%. Some 33% of direct annuities were paid out to workers who were injured while driving a private car.

About 59% of temporary compensations are provided to private car drivers involved in road accidents.

Table 3.6 - Injury claims due to road accident by nature of injury and vehicle type. Year 2018 (percentage values)

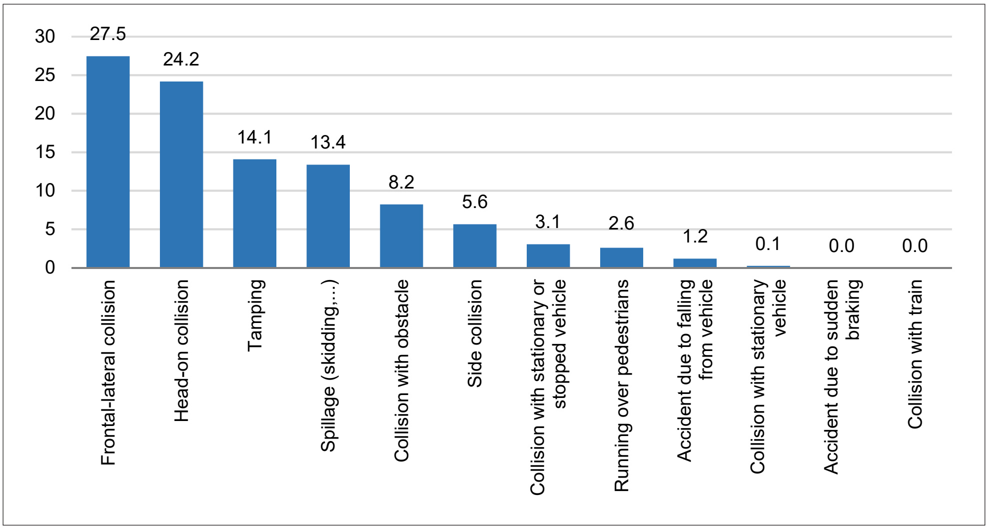
Nature of injury	Car	Truck	Motorcycle	Bicycle	Other vehicle	Total
Bruise	24.7	29.8	41.8	56.3	38.2	32.0
Foreign bodies	-	0.1	-	-	0.1	-
Wound	2.2	3.7	5.4	4.3	4.0	3.3
Fracture	13.0	17.7	39.1	27.1	26.7	21.3
Injuries from other agents	0.3	0.6	0.2	0.2	0.5	0.3
Stress injuries	-	-	-	-	-	-
Dislocation, sprain, distraction	59.8	48.0	13.3	12.1	30.3	43.0
Anatomic loss	-	0.1	0.2	-	0.2	0.1
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

A cross-section of the data by type of vehicle involved and the nature of the injury[5] resulting from the accident shows that the majority of accident events give rise to injuries of the contusion, fracture and dislocation-distortion type. Results are shown in Table 3.6. Motorbikes and velocipedes were the main cause of contusions and fractures (41.8% and 56.3% for the former and 39.1% and 27.1% for the latter), mostly involving the upper and lower limbs, chest, internal organs and in some cases the head. Accidents involving lorries and private cars mainly involved dislocations and sprains (48% and 59.8% respectively), with 3 out of 4 cases involving the spinal column and a smaller percentage involving the head.

Looking at the "nature of the accident" and the damage resulting from it (Figure 3.12), it emerges that the most dangerous accidents, which account for 87% of all fatal accidents, are mainly due to: running off road (skidding) (13.4%), head-on collision (24.2%), frontal-lateral collision (27.5%), rear-end collision (14.1%) and collision with an obstacle (8.2%).

Figure 3.12 - Injury claims due to road accident by event outcome and vehicle type. Year 2018 (percentage values)



Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

5 Generic term for any change, in a pathological sense, in the structure and function of a tissue or organ, irrespective of the causes that may have produced it (*e.g.* mechanical, thermal, chemical, bacterial, *etc.*).

In Table 3.7 five types of accident are analysed in relation to the physical injuries sustained. As shown, injuries to the skull (58.5%) have the highest incidence for the cause of death, followed by injuries to the heart and mediastinal organs (16.1%), the ‘chest wall’ (10.7%) and those to the abdominal wall and organs (9.2%).

Injuries to the skull occur most frequently in accidents involving head-on side collision (71.0%), running off road (skidding) (63.9) and collision with an obstacle (61.1%), which are clearly due to the presence of older vehicles without side safety systems. The lowest value is found in head-on collision (40.0%), where frontal airbags are also present in older vehicles. Damage to the heart and organs of the mediastinum occurred most frequently in the head-on collision (28.3%) and in the rear-end collision (17.1%), which is clearly due to the vehicle coming to a violent stop. Damage to the abdominal wall and organs and to the thoracic wall generally occurs in almost all types of accident with values ranging from 5.6 per cent to 17.1 per cent.

Table 3.7 - Injury claims due to road accident by site of injury and type of accident. Year 2018 (percentage values)

Injury site	Running off road (skidding)	Head-on collision	Head-on side collision	Rear-end collision	Collision with obstacle	Other	Total
Neck	2.8	10.0	1.4	2.4	11.1	2.8	4.6
Skull	63.9	40.0	71.0	51.2	61.1	66.7	58.5
Heart and mediastinal organs	13.9	28.3	7.2	17.1	11.1	16.7	16.2
Abdominal wall and organs	8.3	11.7	10.1	9.8	11.1	2.8	9.2
Thoracic wall	11.1	8.3	10.1	17.1	5.6	11.0	10.7
Other location	-	1.7	-	2.4	-	-	0.8
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: Authors’ processing on integrated Istat and Inail work-related road accidents’ archive

By location of injury and driver/passenger role (Table 3.8), the passenger seat is more dangerous than the driver seat, demonstrated by the fact that for almost all injury locations, there are higher values for the passenger than for the driver.

Table 3.8 - Injury claims due to road accident by site of injury and role. Year 2018 (percentage values)

Injury site	Role			
	Driver	Passenger	Pedestrian	Total
Neck	4.3	5.9	9.1	4.6
Skull	57.7	58.8	72.7	58.4
Heart and mediastinal organs	16.4	17.6	9.1	16.2
Abdominal wall and organs	9.1	11.8	9.1	9.2
Thoracic wall	12.1	-	-	10.8
Other location	0.4	5.9	-	0.8
Total	**100.0**	**100.0**	**100.0**	**100.0**

Source: Authors' processing on integrated Istat and Inail work related road accidents' archive

Analysis by economical activity

Interesting analyses can be carried out by breaking down the accident cases by work activity, in particular by large tariff group[6].

The territorial distribution of accidents occurring at work in the Industry and Services sector (Table 3.9) shows that 60% of accidents in Miscellaneous (mainly services) occurred in the North, 60% in Construction in the North and 21.3% in the Centre, 34.5% in Agricultural work in the Centre-South and 63.6% in Transport in the North.

Table 3.9 - Injury claims due to road accident occurred at work in Industry and Services by large tariff group and geographical distribution. Year 2018 (percentage values)

Large tariff group	Centre	Islands	North-East	North-West	South	Total
Agricultural processing	28.0	8.5	22.0	24.6	16.9	100.0
Chemistry	20.0	5.0	27.5	40.0	7.5	100.0
Construction	21.3	6.3	26.7	33.1	12.6	100.0
Electricity	25.5	11.8	23.5	21.6	17.6	100.0
Wood and related	29.5	6.8	29.6	22.7	11.4	100.0
Metallurgy	18.9	3.7	34.7	32.3	10.4	100.0
Mining	29.4	5.9	41.2	23.5	0.0	100.0
Textiles and clothing	22.4	1.7	25.9	31.0	19.0	100.0
Transport	18.0	5.2	30.1	33.6	13.1	100.0
Miscellaneous	18.9	7.9	29.0	30.6	13.6	100.0
Total	**19.3**	**6.5**	**29.6**	**31.5**	**13.1**	**100.0**

Source: Authors' processing on integrated Istat and Inail work related road accidents' archive

6 The large tariff grouping that groups the tariff items, which associate the work with the premium rate.

The distribution of accidents by hour (Table 3.10) shows that most accidents occur between 7 and 9 a.m. (22.9%), between 1 and 3 p.m. (15.4%) and between 5 and 7 p.m. (13.5%). The tariff groups most affected are those in Miscellaneous (mainly services). As regards night-time accidents (between midnight and 5 a.m.), the prevalence is also in Miscellaneous (52%), followed by Transport and Agriculture.

Table 3.10 - Injury claims due to road accident occurred at work in Industry and Services by large tariff group and hour of the event. Year 2018 (percentage values)

Large tariff group	1-6	7-12	13-18	19-24	Total
Agricultural processing	19.4	36.5	34.4	9.0	100.0
Chemistry	12.7	36.0	41.4	9.9	100.0
Construction	9.6	41.7	41.1	7.5	100.0
Electricity	3.5	56.6	36.3	3.5	100.0
Wood and related	8.6	47.1	38.6	5.7	100.0
Metallurgy	9.4	39.9	43.2	7.5	100.0
Mining	19.4	36.7	37.4	6.5	100.0
Textiles and clothing	8.6	44.1	42.2	5.1	100.0
Transport	15.4	35.1	37.5	10.6	100.0
Miscellaneous	7.0	44.9	36.8	10.5	100.0
Total (a)	**9.0**	**42.1**	**38.2**	**10.0**	**100.0**

Source: Authors' processing on integrated Istat and Inail work related road accidents' archive
(a) The total contains the undefined cases.

4. Discussion

The analysis of the integrated data of the general road accidents collected by Istat and the work-related ones registered by Inail, allows focussing on new interesting data.

The record linkage between different sources has been cited in other articles in literature too (Boufous *et al.*, 2006 and 2009; Adminaite *et al.*, 2017).

The findings contribute to cover the gap in knowledge in the area of work-related road accidents and highlight the most serious injuries and the consequent risk of permanent disability and death. In the study emerges that the injury claims due to work-related road accidents mainly refers to the commuters (about 78%), consequently, in average, only one in five injury claims involves an on-duty driver. The private cars and motorbikes represent the means of transport with the highest number of fatal road traffic accidents and consequently the highest number of survivor claims (39.3% and 26.8% respectively). Especially for the Powered Two Wheels, vehicles without external protections, in 43% of accidents, a permanent disability and hence a direct pension is awarded. Important joined information, not currently available, but contained in the integrated dataset, are the type and the body region of the injury connected with the type of vehicle and the dynamic of the accident (Schick *et al.*, 2019).

Finally, the proportion of commuting road accidents (78%) and those occurring on duty (22%), in the integrated database, follows the same percentages recorded in the original Inail archives. This is an encouraging result for the validity of the results.

The study indeed demonstrates the value of record linkage techniques in addressing some of the limitations of work-related data systems and road accidents archives and in providing a more complete picture of the circumstances of occupational road accidents.

Another important element emerged is that men (workers) receive more compensation (64%) whatever the type of vehicle involved, while women (workers) receive more compensation than their men colleagues, when the vehicle involved is a passenger car (50.5% compared with 49.5%). Besides, a cross-section of the data by type of vehicle involved and the nature of the injury resulting from the accident, results show that the majority of accident events give rise to injuries of the contusion, fracture and dislocation-distortion type. Users of motorcycles and bicycles had, as a main cause of injury, contusions and fractures (respectively 41.8% and 56.3% for motorcyclists and 39.1% and 27.1% for bike-riders), mostly involving the upper and lower limbs, chest, internal organs and the head.

Accidents involving trucks and passenger cars mainly involved dislocations and sprains (48% and 59.8% respectively), with 3 out of 4 cases involving the spinal column and a smaller percentage involving the head.

Looking at the “nature of the accident” and the damage emerges that the most dangerous accidents, which account for 87% of all fatal accidents, are mainly due to: running off the road (13.4%), head-on collision (24.1%), frontal-lateral collision (27.5%), rear-end collision (14.1%) and collision with an obstacle (8.2%).

The analysis of injury claims occurred at work, specifically in Industry and Services, the main group represented, provide interesting results if read by pricing risk large group and geographical distribution.

In particular, the road accidents at work in “Agricultural processing” and in “Electricity” group occurs mainly in the Centre (respectively 28% and 25.5% out of total of the pricing group). Besides, the “Construction” pricing risk group, “Metallurgy”, “Mining”, “Textiles, and clothing” record the highest number of road accidents at work in the North-East area (26.7%, 34.7%, 41.2% and 25.9%). For “Chemistry” and “Transport”, the peak of the distribution is detected for the North-West (40% and 33.6%). In the South and Islands, the percentages of work-related road accidents are lower than in the other territorial areas.

As regards the choice not to use, in this article, indicators and relative indices (*e.g.* accidents/number of employees) by sector of economic activity and groups of tariffs, it is due to the nature of the denominators available, consisting of estimated values and which could introduce a bias.

The Inail statistical archive contains, indeed, only the employee-year information (units of work per year) for Industry and Services, which is based on estimation of the incomes that the employer declares to pay, with reference to the work performed. In particular, employee’s amount, is obtained by the ratio between the declared incomes and the average daily salary for 300 (theoretical number of working days per year in 52 weeks, excluding Sundays and national holidays). Self-employed workers (owners, partners and housekeeper), on the other hand, are given by the "insured heads". The number of Employees-year does not include those categories of workers (*e.g.* apprentice artisans and non-artisans, members of cooperative, *etc.*) for which salaries are not recognised, due to the "insurance bonus" not linked to them; "temporary" workers (former temporary workers) are also excluded from the count. Furthermore, with regard to multi-location companies (*e.g.* large companies with several workplaces), the employees are generally attributed to the parent office. The ATECO classification of the economic activities is attributed to the company as a whole, regardless of the work actually carried out by the workers.

5. Conclusions

Linked administrative databases offer, in general, a powerful resource for studying important issues. Specifically, the integration of the databases on road accidents and work-related road accidents provides an enriched base of new elements not available otherwise to address policies and interventions useful for both general road safety and safety at work.

The goal of data integration underlying the descriptive analysis presented in this work is the enrichment of existing public statistics. By putting together two data sources that have in common the accidents (with deaths and injuries) occurring on public roads and with means of transport, it is possible to complete the information on the person involved. Information relating to the accident and to the work aspect can be attributed to the person involved in the accident, whether injured or deceased, enriching the knowledge on road accidents related to work allows to work on aspects that were not planned neither on the general accident side nor on the work-related one.

For public statistics, the possible integration of data that can be done with other sources is an opportunity not to be missed.

Future work developments may involve other data sources. It could be thought of connecting data sources that contain detailed information on the vehicle involved in the accident as age and maintenance. The data linking considered in this work is done on individual data. For better understanding the context in which related work accidents occur, it might be useful to include contextual data in the study. Aggregate context data could also be considered such as the territorial working fabric, the offer of urban mobility, road maintenance. Methods developed and implemented in several dominions have achieved high-quality linkages for conducting health, safety and social research and to enhance the potential of the two individual sources.

The performance of the applied record linkage has also led to satisfactory results, such as to imply a replicability of the experience also for the years to come.

Finally, the use of these techniques is in line with the strategy implemented by the main official statistical Institutes, especially, in recent years, for the enhancement of existing administrative archives, aimed at both reducing the statistical burden on respondents, and eliminating redundancies in data dissemination.

Appendix

Table A1 - Injury claims due to road accidents by event and type. Years 2014-2018 (absolute value)

Type of accident	Fatal event	Non-fatal event	Total
	2014		
Commuting road accident *in itinere*	216	19,230	19,446
Accident occurred in the course of work	158	5,779	5,937
Total	**374**	**25,009**	**25,383**
	2015		
Commuting road accident *in itinere*	243	18,918	19,161
Accident occurred in the course of work	172	5,491	5,663
Total	**415**	**24,409**	**24,824**
	2016		
Commuting road accident *in itinere*	208	20,066	20,274
Accident occurred in the course of work	146	5,627	5,773
Total	**354**	**25,693**	**26,047**
	2017		
Commuting accident - Fatal event	214	20,012	20,226
Accident occurred in the course of work - Non-fatal event	150	5,496	5,646
Total	**364**	**25,508**	**25,872**
	2018		
Commuting road accident *in itinere*	250	20,694	20,944
Accident occurred in the course of work	176	5,591	5,767
Total	**426**	**26,285**	**26,711**

Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Table A2 - Injury claims due to road accidents by event and age classes. Year 2018 (%)

Age (years)	**Fatal event**	**Non-fatal event**	**Total**
Up to 34	27.5	34.8	34.6
35 - 49	35.0	38.4	38.4
50 - 64	34.3	25.6	25.8
65 +	3.1	1.2	1.2
Total	**100.0**	**100.0**	**100.0**

Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Table A3 - Claimed road accidents by vehicle type and event outcome. Event year 2018 (absolute value)

Event	Truck	Passenger car	Motorcycle	Bicycle	Other vehicle
Fatal event	67	172	128	15	44
Non-fatal event	1,960	14,395	6,247	1,567	2,116
Total	**2,027**	**14,567**	**6,375**	**1,582**	**2,160**

Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Table A4 - Contingency tables analysis and statistics on the association between the main categorical variables of the database and the year of event. Years 2014-2018

Statistical Test and variables	Degrees of Freedom	Value	Probability
Statistics for Type of event - Table of by Year			
Chi-Square	4	117198	0.0196
Likelihood Ratio Chi-Square	4	116593	0.0201
Mantel-Haenszel Chi-Square	1	0.0017	0.9668
Phi Coefficient		0.0095	
Contingency Coefficient		0.0095	
Cramer’s V		0.0095	
Statistics for “Commuting”/“Accident in the course of work” by Year			
Chi-Square	4	324192	<.0001
Likelihood Ratio Chi-Square	4	323245	<.0001
Mantel-Haenszel Chi-Square	1	312294	<.0001
Phi Coefficient		0.0159	
Contingency Coefficient		0.0159	
Cramer’s V		0.0159	
Statistics for Age by Year			
Chi-Square	336	7467411	<.0001
Likelihood Ratio Chi-Square	336	7572584	<.0001
Mantel-Haenszel Chi-Square	1	421660	<.0001
Phi Coefficient		0.0761	
Contingency Coefficient		0.0759	
Cramer’s V		0.0381	
Statistics for Gender by Year			
Chi-Square	4	83020	0.0811
Likelihood Ratio Chi-Square	4	83043	0.0810
Mantel-Haenszel Chi-Square	1	60110	0.0142
Phi Coefficient		0.0080	
Contingency Coefficient		0.0080	
Cramer’s V		0.0080	
Statistics for Tariff categories for compensation by Year			
Chi-Square	952	17298349	<.0001
Likelihood Ratio Chi-Square	952	17320166	<.0001
Mantel-Haenszel Chi-Square	1	92500	0.0024
Phi Coefficient		0.1162	
Contingency Coefficient		0.1154	
Cramer’s V		0.0581	

Source: Authors’ processing on integrated Istat and Inail work-related road accidents’ archive

Table A4 continued - **Contingency tables analysis and statistics on the association between the main categorical variables of the database and the year of event. Years 2014-2018**

Statistical Test and variables	**Degree of Freedom**	**Value**	**Probability**
Statistics for Localisation of the accident by Year			
Chi-Square	36	2144386	<.0001
Likelihood Ratio Chi-Square	36	2137126	<.0001
Mantel-Haenszel Chi-Square	1	302096	<.0001
Phi Coefficient		0.0408	
Contingency Coefficient		0.0408	
Cramer's V		0.0204	
Statistics for Table of Geographical Macro-region by Year			
Chi-Square	16	727546	<.0001
Likelihood Ratio Chi-Square	16	735020	<.0001
Mantel-Haenszel Chi-Square	1	162837	<.0001
Phi Coefficient		0.0238	
Contingency Coefficient		0.0238	
Cramer's V		0.0119	
Statistics for Table of Month by Year			
Chi-Square	44	2187552	<.0001
Likelihood Ratio Chi-Square	44	2193183	<.0001
Mantel-Haenszel Chi-Square	1	70587	0.0079
Phi Coefficient		0.0412	
Contingency Coefficient		0.0412	
Cramer's V		0.0206	
Statistics for Nature of accident by Year			
Chi-Square	44	1694103	<.0001
Likelihood Ratio Chi-Square	44	1663928	<.0001
Mantel-Haenszel Chi-Square	1	30231	0.0821
Phi Coefficient		0.0363	
Contingency Coefficient		0.0362	
Cramer's V		0.0181	
Statistics for Nature of injury by Year			
Chi-Square	32	872429	<.0001
Likelihood Ratio Chi-Square	32	884873	<.0001
Mantel-Haenszel Chi-Square	1	201434	<.0001
Phi Coefficient		0.0280	
Contingency Coefficient		0.0280	
Cramer's V		0.0140	

Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive

Table A4 continued - **Contingency tables analysis and statistics on the association between the main categorical variables of the database and the year of event. Years 2014-2018**

Statistical Test and variables	Degree of Freedom	Value	Probability
Statistics for Time (hour) by Year			
Chi-Square	96	2131427	<.0001
Likelihood Ratio Chi-Square	96	2136345	<.0001
Mantel-Haenszel Chi-Square	1	18593	0.1727
Phi Coefficient		0.0407	
Contingency Coefficient		0.0406	
Cramer's V		0.0203	
Statistics for Role of injured/dead by Year			
Chi-Square	8	127114	0.1222 (a)
Likelihood Ratio Chi-Square	8	126120	0.1259
Mantel-Haenszel Chi-Square	1	40153	0.0451
Phi Coefficient		0.0099	
Contingency Coefficient		0.0099	
Cramer's V		0.0070	
Statistics for Injury Site by Year			
Chi-Square	260	3475359	0.0002
Likelihood Ratio Chi-Square	260	3499807	0.0002
Mantel-Haenszel Chi-Square	1	282354	<.0001
Phi Coefficient		0.0559	
Contingency Coefficient		0.0558	
Cramer's V		0.0279	
Statistics for Compensation type by Year			
Chi-Square	16	1524017	<.0001
Likelihood Ratio Chi-Square	16	1516563	<.0001
Mantel-Haenszel Chi-Square	1	361649	<.0001
Phi Coefficient		0.0344	
Contingency Coefficient		0.0344	
Cramer's V		0.0172	
Statistics for Vehicle type by Year			
Chi-Square	76	2200422	<.0001
Likelihood Ratio Chi-Square	76	2226930	<.0001
Mantel-Haenszel Chi-Square	1	52514	0.0219
Phi Coefficient		0.0421	
Contingency Coefficient		0.0420	
Cramer's V		0.0210	

Source: Authors' processing on integrated Istat and Inail work-related road accidents' archive
(a) All the main variables of the database show an independence in the distribution, with respect to the year of the event. Only for the variable "Role of injured/dead" in the accident (Driver, Passenger, Pedestrian), the chi-square test leads to a probability associated to the corresponding p-value equal to 0.122 (Low significance level, if it is considered the hypothesis H0 of independence of the variables in the statistical test). The final decision of the authors was to focus the analysis, however, only on the year of event 2018, as representative of the whole period, due to the result of the mentioned chi-square test is conditioned by a slight difference in the proportion of passengers, observed in 2014.

References

Adminaite, D., G. Jost, H. Stipdonk, and H. Ward. 2017. "Tapping the potential for reducing work-related road deaths and injuries". *PIN Flash Report 33*. Bruxelles, Belgium: European Transport Safety Council - ETSC.

Boufous, S., and A. Williamson. 2009. "Factors affecting the severity of work related traffic crashes in drivers receiving a worker's compensation claim". *Accident Analysis & Prevention*, Volume 41, Issue 3: 467-473.

Boufous, S., and A. Williamson. 2006. "Work-related traffic crashes: a record linkage study". *Accident Analysis & Prevention*, Volume 38, Issue 1: 14-21.

Brusco, A., A. Bucciarelli, M. Bugani, C. Gariazzo, C. Giliberti, M. Marinaccio, S. Massari, A. Pireddu, L. Veronico, G. Baldassarre, S. Bruzzone, M. Scortichini, M. Stafoggia, e S. Salerno. 2019. *Gli incidenti con mezzo di trasporto. Un'analisi integrata dei determinanti e dei fattori di rischio occupazionali*. Roma, Italy: Inail. https://www.inail.it/cs/internet/comunicazione/pubblicazioni/catalogo-generale/pubbl-incidenti-con-mezzo-di-trasporto.html

European Commission, Directorate-General for Mobility and Transport, CARE Team. 2021. *CARE DATABASE. CADAS - Common Accident Data Set. Reference Guide Version 3.8 - June 2021*. Bruxelles, Belgium: European Commission.

European Union, United Nations Economic Commission For Europe - UNECE, International Transport Forum - ITF, and Eurostat. 2019. "Glossary for transport statistics. 5th Edition, 2019". *Manuals and Guidelines*. Luxembourg: Publications Office of the European Union.

Eurostat. 2021. *NUTS - Nomenclature of Territorial Units for Statistics*. Luxembourg: Eurostat. https://ec.europa.eu/eurostat/web/nuts/nuts-maps.

Istituto Nazionale di Statistica - Istat (*Italian National Institute of Statistics - Istat*). 2021. *I.Stat. Your direct access to the Italian Statistics - Health Statistics, Road Accidents*. Roma, Italy: Istat. http://dati.istat.it/?lang=en&SubSessionId=dc3e67e3-ac7f-4262-818c-568171c99fdc.

Istituto Nazionale di Statistica - Istat (*Italian National Institute of Statistics - Istat*). 2021. "Road Accidents. Year 2020". *Press Release*. Roma, Italy: Istat. https://www.istat.it/en/archivio/259830.

Istituto Nazionale di Statistica - Istat (*Italian National Institute of Statistics - Istat*). 2019. "Road Accidents. Year 2018". *Press Release*. Roma, Italy: Istat. https://www.istat.it/en/archivio/232376.

Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro - Inail (*Italian National Institute for Insurance against Accidents at Work - Inail*). 2021. *Banca dati statistica prevenzionale*. Roma, Italy: Inail. https://www.inail.it/cs/internet/attivita/dati-e-statistiche/banca--dati-statistica.html.

Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro - Inail (*Italian National Institute for Insurance against Accidents at Work - Inail*). 2019. Relazione Annuale 2018. Roma, Italy: Inail.

Schick, S., S. Piantini, M. Wisch, and J. Brown. 2019. "Comparing consequences of using two different definitions for body regions for the improvement of personal protective equipment for powered two-wheelers". *Traffic injury prevention*, Volume 20, Issue supplement 2: S182-S185.

Taiano, L., S. Massari, T. Tuoto, L. Valentino, S. Bruzzone, and L. Veronico. 2021. "Work-related road accidents: a data linkage procedure applied to assess traffic accidents at work and commuting". In this issue of the *Rivista di statistica ufficiale*, N. 3/2021. Roma, Italy: Istat.

Tuoto, T., S. Bruzzone, L. Valentino, G. Baldassarre, N. Cibella, and M. Pappagallo. 2012. "Towards an integrated surveillance system of road accidents". In Società Italiana di Statistica - SIS. *Atti della XLVI Riunione Scientifica (Roma, 20-22 giugno 2012)*. Padova, Italy: CLEUP.

Tuoto, T., D. Moretti, C. Orsi, G. Baldassarre, G. Di Fraia, e S. Bruzzone. 2018. "Le vittime in incidenti stradali: una esperienza di record linkage tra diverse fonti informative". In Osservatorio Nazionale sulla Salute nelle Regioni italiane. *Rapporto Osservasalute 2017. Stato di salute e qualità dell'assistenza nelle regioni italiane - Approfondimenti*. Roma, Italy: Università Cattolica del Sacro Cuore.

United Nations. 1977. “CHAPTER XI - TRANSPORT AND COMMUNICATIONS. B. Road Traffic. 19. Convention on Road Traffic. Vienna, 8 November 1968”. *Treaty Collection*. New York, NY, U.S.: United Nations.

United Nations Economic Commission For Europe - UNECE, International Transport Forum - ITF, and Eurostat. 2019. “Glossary for transport statistics. 5th Edition, 2019”. *Manuals and Guidelines*. Luxembourg: Publications Office of the European Union.

Work-related road accidents: a statistical multivariate analysis in Italy

Claudio Gariazzo [1], Alessandro Marinaccio [1], Silvia Bruzzone [2], Luca Taiano [1], Liana Veronico [1]

Abstract

Despite the relevance of road accidents and their impact on social and health care costs, their work-related component is seldom studied. The available data about road accidents lack of occupational parameters that allow studying the phenomenon. We produced an integrated archive linking, at individual level, road accidents data with compensation for occupational injuries data referring to accidents occurred in Italy from 2014 to 2018. Data were statically analysed by frequency, time series, cluster and multiple correspondence analysis to describe their characteristics, time evolution, and to highlight the most representative road accidents among fatal and non-fatal events at work and during commuting. Results indicate a higher occurrence of accidents during commuting and the importance of miscellaneous economic activities as well as the transport and warehouse economic sectors.

Keywords: Road accidents, occupational injuries, compensation data, data linkage, time series.

1 Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro (*Italian National Institute for Insurance against Accidents at Work*) - Inail.

2 Istituto Nazionale di Statistica (*Italian National Institute of Statistics*) - Istat.

In the writing of this article, the authors were supported by all the members of the Inail-Istat collaborative group: Antonella Altimari, Michela Bonafede, Roberto Boscioni, Claudio Gariazzo, Alessandro Marinaccio, Stefania Massari, Antonella Pireddu, Luca Taiano and Liana Veronico (for Inail); Giordana Baldassarre and Silvia Bruzzone (for Istat).
Although this article is the result of all the authors' commitment, the paragraphs are attributed as following: 1, 2.2, 3, 4 and 5 to Claudio Gariazzo and Alessandro Marinaccio; 2.1 to Silvia Bruzzone, Liana Veronico and Luca Taiano.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

Road traffic injuries represent a relevant public health problem. According to the WHO, road traffic accidents account for almost 1.3 million deaths a year around the world, and between 20 and 50 million victims sustain non-fatal injuries (WHO, 2018). In Italy, the Italian National Institute of Statistics - Istat registered about 170,000 road accidents during the year 2019, for which 3,173 persons died and 241,384 were injured. A 30% reduction is estimated for the year 2020 due to lockdown for COVID-19 pandemic (Istat, 2020). A substantial fraction of these road accidents have an occupational origin. Workers use vehicles either for commuting (home-work travelling routes) or for activities at work (*e.g.* in the transport sector). According to data collected by the Italian National Institute for Insurance against Accidents at Work (Inail), the occupational accidents occurring using a transport vehicle represent about 14% of the total registered occupational injuries, of which 11% are related to commuting and 3% to activities at work (Inail, 2020).

Risk factors for work-related road accidents have been investigated by some authors. Weather is considered to be an important factor particularly for those activities involving heavy truck vehicles. Driving in windy, rainy and snowy conditions was found to be associated with a higher risk (Ahmed *et al.*, 2018; Moomen, Rezapour, and Ksaibati, 2019; Naik *et al.*, 2016; Uddin and Huynh, 2020). Extreme temperatures were also found to be associated with risk of road accidents (Gariazzo *et al.*, 2021; Wu, Zaitchik, and Gohlke, 2018). Additional risk factors are fatigue, stress and sleepiness (Öz, Özkan, and Lajunen, 2010; Robb *et al.*, 2008). Other studies addressed factors like driving behaviours (Mitchell, Bambach, and Friswell, 2014), age of drivers (Newnam *et al.*, 2018), exposure (Pei, Wong, and Sze, 2012), and scheduling issues as well as physical constrains at work (Fort *et al.*, 2010) as determinants for work-related road accidents.

The analyses of road accidents and their determinants are often limited by the amount and quality of data, particularly for work-related road accidents. Road accidents data are routinely collected by national or local Authorities at the time of the event, as in Italy, but the information about the occupational origin is rarely or partially collected and factors useful for an insight analysis are missing. To overcome these limitations, some authors used either case-control studies or cross-sectional studies based on trauma registry, selfadministered postal questionnaire or interviews, in a restricted sample of road accidents. In this way they could analyse the occupational characteristics of work-related road accidents and obtain better knowledge of the types of occupational groups and industry sectors involved (Fort *et al.*, 2010; Hours *et al.*, 2011). Other studies are focussed on particular categories of workers, like truck drivers, to collect information on risk factors (Uddin and Huynh, 2020). The use of workers compensation data, linked with the police road accidents' records, allows increasing the effectiveness of the epidemiological analysis, adding supplementary occupational information otherwise not available (Boufous and Williamson, 2006). The efficiency of data linkage between the two archives often limits its application, as recording errors in the two registration systems, the different sources of registration (police intervention *vs.* occupational accident compensation claims), and missing declarations of occupational accidents, might have produced missed matches. A recent study in Italy linked the two archives for the year 2015 finding that only 23% (20,941) of individuals who claimed for compensation were linked to the general road accidents archive provided by Istat (Brusco *et al.*, 2019). As most of these events caused injury to the involved subjects, the number of fatal events were not sufficient to be analysed. In addition, time series analysis was not possible using this time-restricted dataset.

Due to the above restrictions, the studies about the occupational component of road accidents are rather limited and focussed either on the determinants of road accidents or on their occupational characteristics, but not both at the same time. Consequently, there is a need for combining and analysing road accidents and occupational data as a whole to obtain detailed information about which type of road accidents is more involved by workers characteristics and economic sector.

The aim of the present study is to provide, by means of a multi-year data archive and statistical analysis procedures, a more complete picture of the circumstances of work-related road accidents occurring in Italy, including both fatal and non-fatal events, getting a new insight in the occupational risk factors of road accidents.

2. Materials and methods

2.1 Work-related road accidents data archive

In Italy data about road accidents, in which an injury or fatality occurred, are collected by Istat on the basis of data recorded from different national and local Authorities. Although such data should contain information about the occupational condition of the driver involved in an accident, this is rarely available. Consequently, the only established observatory of work-related road accidents is that provided by Inail in the frame of occupational injuries. The Inail archive (Inail, 2019) covers about 80% of the Italian workforce. It receives compensation claims for occupational injuries over the whole national territory, regarding all workers, except for some categories (armed forces, firefighters and police workers, air transport personnel, autonomous tradespeople and professionals with VAT registration). Inail collects all requests of compensation for accidents occurred during the course of work for violent and external causes with injuries that determined temporary inability (higher than 3 days), permanent inability or death. In case of road accidents, it classifies them either as during commuting (home-work journey) or on-duty (involved vehicle is used for work activity). The requests of compensation are processed to verify if they are work-related, and to determine the type of indemnity assigned (temporary; annuity to survivors; direct annuity; capital account) and, if the case, the duration of leave. Inail process all these events and produces statistical reports about this phenomenon.

We selected from the Inail archive the compensation claims for injuries occurring with the use of a vehicle during either commuting or activities at work. The collected data include date/time and location of accident; demographic variables (gender, age at injury); modality of occupational accident (commuting; on-duty); economic sector of activity derived from the Inail classification of tariff group[3]; information on the gravity of the injury, measured as the duration of leave (including those events below the allowed threshold of 3 days, which were not compensated); and degree of impairment. All the above information are available at individual level. No information

3 The tariff group refers to an aggregation of industrial processes that associate the production sector with the premium rate.

is provided about the characteristics of the road accident. Such information could be gathered from the Istat archive, through a proper individual record linkage between the two archives.

The accident data collected by Istat contain information about some characteristics of the road accident, such as: information about date and localisation (built-up area or outside built-up areas and subcategories); type of road and weather conditions; type of junction; number and types of vehicles involved; road accident type (among moving vehicles, between a moving vehicle and pedestrian, between a moving vehicle and a stationary one or other obstacle, moving vehicle without collision and subcategories), the role of person involved in the accident (driver, passenger, pedestrian), names of individuals involved in the road accident. Istat and Inail archives were linked at individual level using a combined deterministic and probabilistic procedure (Taiano *et al.*, 2021). Briefly, as the data collected by Istat refer to accidents, possibly involving one or more individuals, this archive were first transformed into n records, each one representing a person involved in the accident. Then we linked this archive to the compensation claims one provided by Inail using common variables such as name, surname and age of involved individual, date and location of event. Afterwards, to increase the number of linked data, a probabilistic procedure were applied to residual data, in which matching variables ranged in a predefined window size and a linkage probability is assigned using a sorted neighbourhood algorithm. The data linkage project is included in the Italian National Statistical Programme (PSN) and allowed by the Data Protection Authority. Sensitive data were anonymised after the data linkage for further use.

The data linkage of the two archives produced a unique dataset including subjects injured or death in work-related road accidents occurred in Italy from 2014 to 2018. For the statistical analysis, a few categorical variables were re-aggregated to obtain both a synthesis of the information and an increase in the number of occurrence. As for the economic sectors, the Inail codes of tariff group was used and recoded in macro-economic categories. Table 2.1 lists some variables selected for this study with a short description.

Table 2.1 - List of variables of work-related road accidents

Variable	Value	Description
Age class	<23; 23-40; 41-60; 61+	
Gender	Men; Women	
Modality of occupational accident	Commuting	Road accident occurring during home-work journey
	On-duty	Road accident using a vehicle for working reasons
Road accident type	Collision with obstacle or parked vehicle;	Impact with stationary vehicle; Impact with parking vehicle; Impact with an obstacle
	Nose-to-tail accident;	Nose-to-tail collision
	Collision among moving vehicles;	Frontal collision; Frontal/lateral collision; Lateral collision;
	Running down pedestrian;	Impact with pedestrian
	Moving vehicle without collision	Off-road vehicles; Sudden braking; Dropped from vehicle
Type of vehicle	Private car;	Private passenger car; Trailer passenger car; Four-cycle
	Agriculture machine;	Road tractor; Agricultural tractor; Animal traction vehicle; other agricultural vehicle
	Heavy vehicle;	Private coach; Truck; Articulated lorry; Special vehicle
	Motorcycle;	Moped; Motorcycle one person; Motorcycle with passenger; Three-wheeler
	Public vehicle;	Urban public service bus; Rescue or police car; Tram; Public car; Public coach
	Bike	Bike
Junction	Straight; Crossroad; Bend; Other	
Localisation of the accident	Motorway; Suburban road; Urban road	
Severity	0; 1-14; 15-30; 31-90; >90	Number of days of leave
Macro-economic sector	Agriculture	Agricultural or forestall workers employed in cultivation, forestry and animal husbandry
	Public employees/Students	Employees of central state administration; public educational employees; students
	Servants	
	GG0 - Various activities	Sale; Reception and catering; Health and social services; Cleaning, sanitation and disinfestation; Cinematography, entertainment, cultural and sports activities; Education, scientific research, survey and prospecting; Various activities
	GG1 - Agricultural and food processing	Agricultural mechanical processing; foods processing; animals slaughter
	GG2 - Chemical, paper, leathers	Plastics and rubber; paper and polygraph; leather
	GG3 - Buildings and installation	Construction; Road and rail constructions and maintenance; Construction of urban lines and pipelines; civil and industrial plant engineering
	GG4 - Energy and communication	Production, transmission, conversion and distribution of energy, gas; water, vapour and conditioning installations
	GG5 - Wood and related	Wood manufacturing; Carpentry and restoration works
	GG6 - Metals and machinery	Metallurgy and works connected; Metals transformation; construction, transformation and repairing of machineries
	GG7- Mining, rocks and glass	Research, extraction and processing of minerals and rocks; production and processing of glass
	GG8 - Textiles and packaging	Processing of textile yarns; packaging of clothing, furniture and shoes; laundries, ironing, dry cleaners
	GG9 - Transport and warehouses	Transports; upload, download, porterage of goods and materials; warehousing and goods handling

Source: Authors’ processing on Inail and Istat datasets

2.2 Statistical analysis

Data of subjects involved in work-related road accidents occurred in Italy from 2014 to 2018 were first analysed in a descriptive and univariate analysis for their occurrences according to accident and occupational variables. Then a time series analysis was carried out to study possible trends and seasonality. Afterwards, two statistical multivariate analyses were applied to classify groups of data and to reduce dimensionality of the studied topic: cluster analysis and multiple correspondence analysis. The above statistical techniques are described in details in the following paragraphs. All analyses have been performed separately for fatal and non-fatal accidents.

2.2.1 Time series analysis

In order to investigate about possible trend and seasonal components in the observed number of subjects involved in work-related road accidents, a decomposition of the time components was carried out. A Seasonal and Trend decomposition using Loess (STL) was applied (Cleveland *et al.*, 1990). STL is a filtering procedure for decomposing a time series into trend, seasonal, and remainder components. STL has a simple design that consists of a sequence of applications of the loess smoother. It allows the seasonal component to change over time with a user-controlled rate of change, as well as to control for the smoothness of the trend-cycle. The two main parameters to be chosen when using STL are the trend-cycle window and the seasonal window. These parameters control how rapidly the trend-cycle and seasonal components can change. The former is the number of consecutive observations to be used when estimating the trend-cycle, the latter is the number of consecutive years to be used in estimating each value in the seasonal component. In this analysis, the option "periodic" was used for the seasonal window parameter (*i.e.* identical across years), which means that smoothing is effectively replaced by taking the mean, while the trend-cycle window parameter was left to be calculated by the internal STL code. The robust fitting was used in the loess procedure. The analysis was implemented by using the STL code implemented in R package.

2.2.2 Cluster analysis

In order to group and order homogenous data of subjects involved in work-related road accidents, data were analysed by cluster analysis using the Partitioning Around Medoids (PAM) and Clara methods depending on the amount of data to be clustered (Kaufman and Rousseeuw, 2005).

The main objective of clustering is to find groups of cases (*e.g.* road accidents) which show a high degree of similarity (a cluster), being as dissimilar as possible to those belonging to other clusters. The PAM method is an iterative procedure to search for k representative objects (Medoids) which represent various aspects of the structure of the data. The quality of resulting medoids is measured by the average dissimilarity between every object in the entire dataset and the medoid of its cluster. The metric used for calculating dissimilarities between observations is the Euclidean distance. The Clara method is an extension of PAM methods and it is suitable to deal with large amount of cases in order to reduce computation time and memory usage. Instead of using the full dataset to cluster data, it samples a chosen amount of cases to be clustered, and applies the PAM algorithm to generate an optimal set of medoids for the sample.

Data about subject involved in work-related road accidents were stratified for type of outcome (fatal or injured) and modality of occupational accident (commuting and on-duty) and cluster analysed separately. This choice is motivated by the different data structures involved when the types of outcome and occupational accidents are both taken into account. As an example, we expect a large spread in the economic sectors involved in accidents during commuting, while a more focussed number of sectors for those during on-duty activities. The following categorical variables were selected for clustering: road accident type, type of vehicle involved, localisation of the accident, junction, age class of involved person, macro-economic sector, all as defined above. In a few analyses some modalities of the macro-economic sector were further aggregated to increase the number of occurrences (*e.g.* public employees, servant and agriculture; wood and related, mining, rocks and glass, textiles and packaging).

In this study, the fatal datasets (commuting and on duty) were clustered by means of PAM method, while non-fatal road accidents datasets (commuting and on-duty) were clustered using the Clara method using 50,000 and 10,000 sample sizes, respectively. The optimal number of clusters were chosen by iteratively applying the PAM and Clara method from 2 to 20 number of clusters and by calculating the silhouette width at each step, which is an aggregated measure of how similar is an observation to its own cluster compared its closest neighbouring cluster. Their highest values determine the optimal number of clusters.

As a result, the clustering analysis assigns each case to a specific cluster. The frequency of the each modality of the selected variables has then been calculated by cluster, to describe the cluster's accident and occupational characteristics (fingerprint). Given variable categories, the lesser the spread of the frequencies among its modalities, the more the cluster is defined for that variable. The modalities with the largest frequency are usually close to medoids descriptors.

2.2.3 Multiple Correspondence Analysis

A Multiple Correspondence Analysis (MCA) has been performed to detect and represent the underlying structures of data, searching for the basilar dimensions. MCA can be seen as a generalisation of principal component analysis when the variables to be analysed are categorical instead of quantitative (Abdi and Williams, 2010). It provides the associations between variable categories, whose distance between any points gives a measure of their similarity or dissimilarity.

The MCA was applied to analyse the association among the following variable: type of accident; modality of occupational accident (commuting or on duty); accident severity according to classes of days of leave (0; 1-14, 15-30; 31-90; >90) and macro-economic sectors. As for the latter, a different classification was adopted to get better information on some subsectors of the GG0-various activities sector. This sector contains a significant portion of the collected work-related road accidents. In particular, the following sectors were used in the MCA analysis, as the most contributing ones: GG01-Sales; GG02 - Reception and Catering; GG03 - Health and social services; GG07 - Various activities; GG3 - Buildings and installation; GG6 - Metals and machinery; GG9 - Transport and warehouses; Other.

The above variables allow for a combined analysis that would have been impossible to carry out using only the respective data archives they were taken from.

3. Results

3.1 Statistical description and time series analysis of work-related road accidents

Table 3.1 shows a statistical description of the subjects involved in work-related road accidents registered by Inail and successful linked to Istat archive. In total 128,795 subjects involved in work-related road accidents were identified during the years 2014-2018. About 25,000 accidents were registered each year with a little yearly variation. Northern regions of Italy show the largest contribution with about 82,000 accidents (64%), followed by the Central regions with 25,900 (20.2%). Less than 12,000 are observed in Southern Italy and Islands. Injuries make the most relevant outcome (98.5%) in occupational road accidents, and men are more involved than women (62 *vs.* 38% respectively). Workers with age between 23 and 60 are the most contributing (89%), while younger ones (<23 years old) are found in 7% of the total cases. The greater number of accidents occurs during commuting (77.7%), while road accidents during the use of a vehicle for on-duty account for 22.3% of cases. The collision among moving vehicles represents the largest type (54.8%), followed by the nose-to-tail ones (26%). Private cars and motorcycles are the two most involved types of vehicle (53.5 and 25.9% respectively). Road accidents mostly occur in urban and suburban roads (62.4 and 29.4% respectively), mainly in straight sections (46.1%) and in crossroads (36.1%).

As far as the macro-economic section is concerned, the various activities group (GG0) largely contributes to the total number of work-related road accidents (53.6%). This is a miscellanea group of different activities like sales, reception and catering, health and social services, cleaning services, entertainment, technical services and other activities. Originally, these activities were grouped to avoid an excessive granularity of cases. Much lower contributions are found for the remaining macro-economic sectors. The second largest contribution is ascribed to the metals and machinery sector (8.8%), closely followed by the transport and warehouses (8.3%). Buildings and installation sector accounts for about 6% of cases, while public employees sector contributes for about 5%.

Table 3.1 - Frequencies of work-related road accidents by variables

Work-related road accidents	Number of subjects involved	%
Overall	128,795	100.0
Year		
2014	25,372	19.7
2015	24,813	19.3
2016	26,041	20.2
2017	25,863	20.1
2018	26,706	20.7
Macro-region		
North-East (Emilia-Romagna, Friuli-Venezia Giulia, Trentino-Alto Adige, Veneto)	40,905	31.8
North-West (Liguria, Lombardia, Piemonte, Valle d'Aosta)	42,007	32.6
Centre (Toscana, Umbria, Marche, Lazio)	25,985	20.2
South (Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria)	12,890	10.0
Islands (Sicilia, Sardegna)	7,008	5.4
Outcome		
Dead	1,891	1.5
Injured	126,904	98.5
Gender		
Men	79,875	62.0
Women	48,920	38.0
Age class		
<23	9,456	7.3
23-40	54,247	42.1
41-60	60,475	47.0
61+	4,617	3.6
Modality of occupational accident		
Commuting	100,033	77.7
On-duty	28,762	22.3
Road accident type		
Moving vehicle without collision	13,060	10.1
Running down pedestrian	3,700	2.9
Collision among moving vehicles	70,627	54.8
Nose-to-tail accident	33,298	25.9
Collision with obstacle or parked vehicle	8,110	6.3
Type of vehicle		
Private car	69,217	53.7
Agriculture machine	164	0.1
Heavy vehicle	12,263	9.5
Motorcycle	33,419	25.9
Public vehicle	2,075	1.6
Bike	7,113	5.5

Source: Authors' processing on Inail and Istat datasets

Table 3.1 continued **- Frequencies of work-related road accidents by variables**

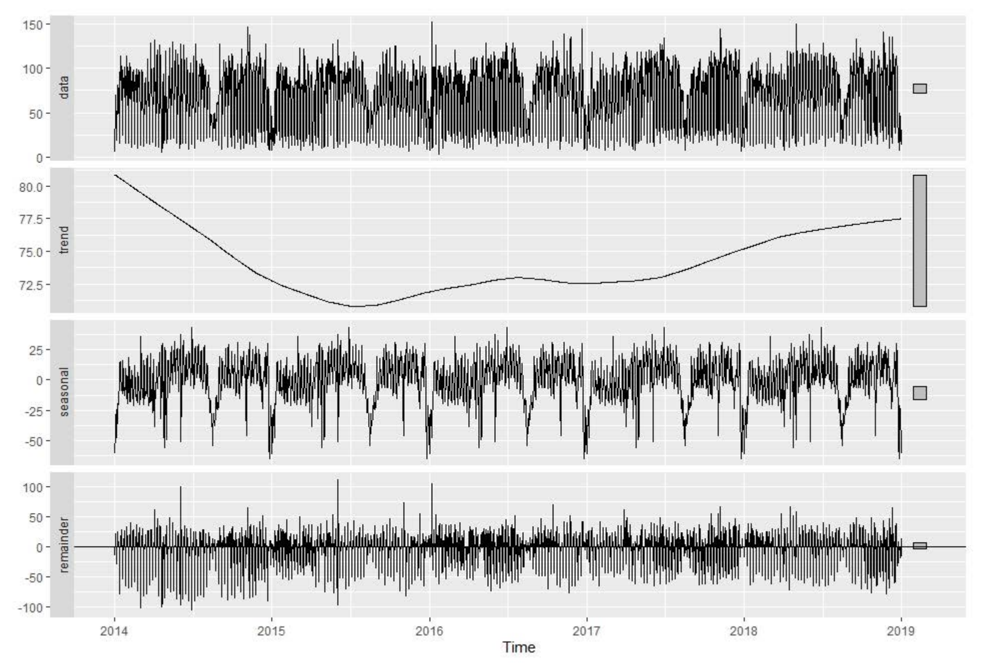
Work-related road accidents		Number of subjects involved	%
Localisation of the accident			
Motorway		10,635	8.3
Suburban road		37,836	29.4
Urban road		80,324	62.4
Junction			
Straight		59,359	46.1
Crossroad		46,467	36.1
Bend		14,277	11.1
Other		8,692	6.7
Macro-economic sector			
Agriculture		2,823	2.2
Public employees/Students		7,084	5.5
Servants		1,497	1.2
GG0 - Various activities		69,022	53.6
	GG01 - Sales	11,421	
	GG02 - Reception & Catering	9,165	
	GG03 - Health & social services	9,958	
	GG04 - Cleaning, sanitation and disinfestation	4,946	
	GG05 - Cinematography, entertainment, cultural and sport activities	813	
	GG06 - Educational, scientific research, survey and prospecting	1,574	
	GG07 - Various activities	28,928	
	GG08 - undefined	1,919	
GG1 - Agricultural and food processing		2,821	2.2
GG2 - Chemical, paper, leathers		2,866	2.2
GG3 - Buildings and installation		8,210	6.4
GG4 - Energy and communication		631	0.5
GG5 - Wood and related		1,037	0.8
GG6 - Metals and machinery		11,305	8.8
GG7- Mining, rocks and glass		759	0.6
GG8 - Textiles and packaging		2,379	1.8
GG9 - Transport and warehouses		10,715	8.3
n.a.		7,646	5.9
Severity [days of leave]			
0		38,799	30.1
1-14		32,833	25.5
15-30		18,691	14.5
31-90		24,097	18.7
>90		14,375	11.2

Source: Authors’ processing on Inail and Istat datasets

Figure 3.1 shows the time series of daily number of subjects involved in work-related road accidents occurred in Italy from 2014 to 2018. Up to 150 involved subjects are observed on a daily bases, with a minimum of about 25. A strong weekly and seasonal variability is also detected, with a significant decrease in the number of events during weekends and holidays, as well as in the month of August, usually used for summer holidays in Italy. A plot of the total monthly road accidents is shown in Figure A1 in the Appendix. The positive association between extreme temperature and road accidents occurrence has been previously estimated (Gariazzo *et al.*, 2021), with an increase of risk of work-related ones for both hot and cold.

The trend and seasonal components are also shown in Figure 3.1. The trend exhibits a decrease up to the middle of 2015, to increase gently from that year on. The seasonal component ranges from -50 to 25 on a daily bases. The lowest values are observed during summer and Christmas holidays when workers have days off. The remainders, the number of events after removing trend and seasonal components, ranges between -75 and 50 events per day, with a significant weekly pattern. Based on a measure of strength of trend and seasonality defined by Wang *et al.* (Wang, Smith, and Hyndman, 2006), a value of 0.99, in a range 0-1, was obtained for both trend and seasonality.

Figure 3.1 - Decomposition of time components of subjects involved in work-related road accidents



Source: Authors' processing on Inail and Istat datasets. Number of subjects involved in daily accidents (upper figure), trend component (higher middle figure), seasonal component (middle figure), reminder component (bottom figure). Boxes in the right side represent the relative scale of components

3.2 Cluster analysis results

The cluster analysis identifies 13 clusters for both on-duty and commuting non-fatal work-related road accidents. Nine and ten clusters were obtained for on-duty and commuting fatal road accidents respectively. As described above, the cases belonging to each cluster were grouped to calculate the frequencies of modalities of each selected variable. Based on these frequencies, heat maps were then plotted to show the accident and occupational characteristics of each identified cluster. Figure 3.2 shows the heat maps of the frequencies of the modalities of each selected variable by cluster, obtained for on-duty and commuting in non-fatal road accidents. Figure A2 in the Appendix shows the correspondent plot for fatal accidents.

As for non-fatal on-duty results, we found five clusters (CL1-4 and CL11) which identify the private car as vehicle involved, with events mostly occurring in urban and suburban roads and an accident type involving moving vehicles or nose-to-tail accident in straight or crossroads, out of those represented by cluster 11, which occur in motorways. These events involve workers at age 23-60 employed in various activities (GG0) (CL1-3, CL11) and transport, warehouses (GG9), as well as the metals and machinery sector (GG6) (CL4). All these car-related events contribute for about 41% of the total on-duty non-fatal road accidents. Another significant portion of road accidents is ascribed to those occurring using a heavy vehicle (27%) (CL6, CL7, CL10, CL12, CL13). They mainly occur in suburban roads and motorways, with the exception of cluster 12, which takes place in urban roads, and are characterised by an accident type involving prevailing moving vehicles or nose-to-tail accident with some cases (5%, CL10 and CL13) by moving vehicle without collision. The macro-economic sectors mostly involved in these heavy vehicle related accidents are the transport and warehouses sector (GG9), the buildings and installation sector (GG3) and the various activities sector (GG0). The remaining clusters (CL5, CL8, and CL9) are represented by different type of vehicles with a small prevalence of motorcycles, occurring in urban roads. The involved workers have age class either 23-40 or 41-60, and are employed in the various activities sector (GG0) or in the transport and warehouses macro-economic sector (GG9).

As for non-fatal commuting results, the use of a private car is found in 8 of 13 clusters. These clusters contribute for about 62% of the total non-fatal commuting road accidents. They mainly occur in urban and suburban roads in proximity of straights or crossroads and involving an accident type among moving vehicles or nose-to-tail accident. The most representative age classes are the middle age classes (23-40, 41-60) which cover almost the entirety of the working period. The most frequent macro-economic sector involved in this kind of accidents is the various activities sector (GG0) with an additional contribution from other sectors like the transport and warehouses (GG9), the metal and machinery (GG6) and the building and installation (GG6), all ascribed to cluster 3 and 9. Four clusters are related with the use of a motorcycle, accounting for about 33% of the total non-fatal commuting work-related road accidents. The urban roads are the most frequent type of road where they take place, mainly involving collision among moving vehicles, as accident type, in proximity of straights or crossroads. For these accidents the various activities economic sector (GG0) contributes for 20% of the total non-fatal commuting accidents, followed by the metal and machinery sector (GG6), the transport and warehouses sector (GG9), and the building and installation sector (GG6), which contribute with one cluster (CL13) for about 9% of the total. The remaining cluster (CL3 in Figure 3.2, bottom) is associated with other macro-economic sectors with about 3% of the cases. Finally, we found a cluster which is characterised by the use of a bike (CL10 in Figure 3.2, bottom), and contributes for about 5% of the non-fatal commuting road accidents. Its profile is described by occurring in urban road, with a collision among moving vehicles, as accident type, in proximity of straights or crossroads and involves workers of middle age classes employed in the various activities sector (GG0).

The cluster analysis results of fatal on-duty work-related road accidents (Figure A2 of Appendix, upper) show similar characteristics as those of non-fatal ones. The only remarkable differences are the larger contribution of heavy vehicles compared to private cars (45 *vs.* 26%), followed by accidents involving motorcycles (22%) with two clusters, and the additional contribution of accidents involving pedestrians, not accounted in the non-fatal analysis, with 6.7% of the cases. In addition, motorways appear with higher frequencies in two clusters. The various activities sector (GG0) and the transport and warehouses sector (GG9) are confirmed to be the most frequent macro-economic sectors involved. As for fatal commuting road accidents, the main characteristics of clusters obtained for non-fatal accidents are also confirmed (Figure A2 of Appendix, bottom). Workers employed in the various activities sector (GG0) involve nine of ten identified clusters. Private cars and motorcycles are confirmed as the two most involved vehicles in fatal commuting road accidents. A remarkable difference with respect to non-fatal commuting road accidents is the higher frequency of suburban roads compared to urban ones.

Figure 3.2 - Heat map of distribution of characteristics of on-duty (upper) and commuting (bottom) work-related non-fatal road accidents by cluster

Source: Authors' processing on Inail and Istat datasets

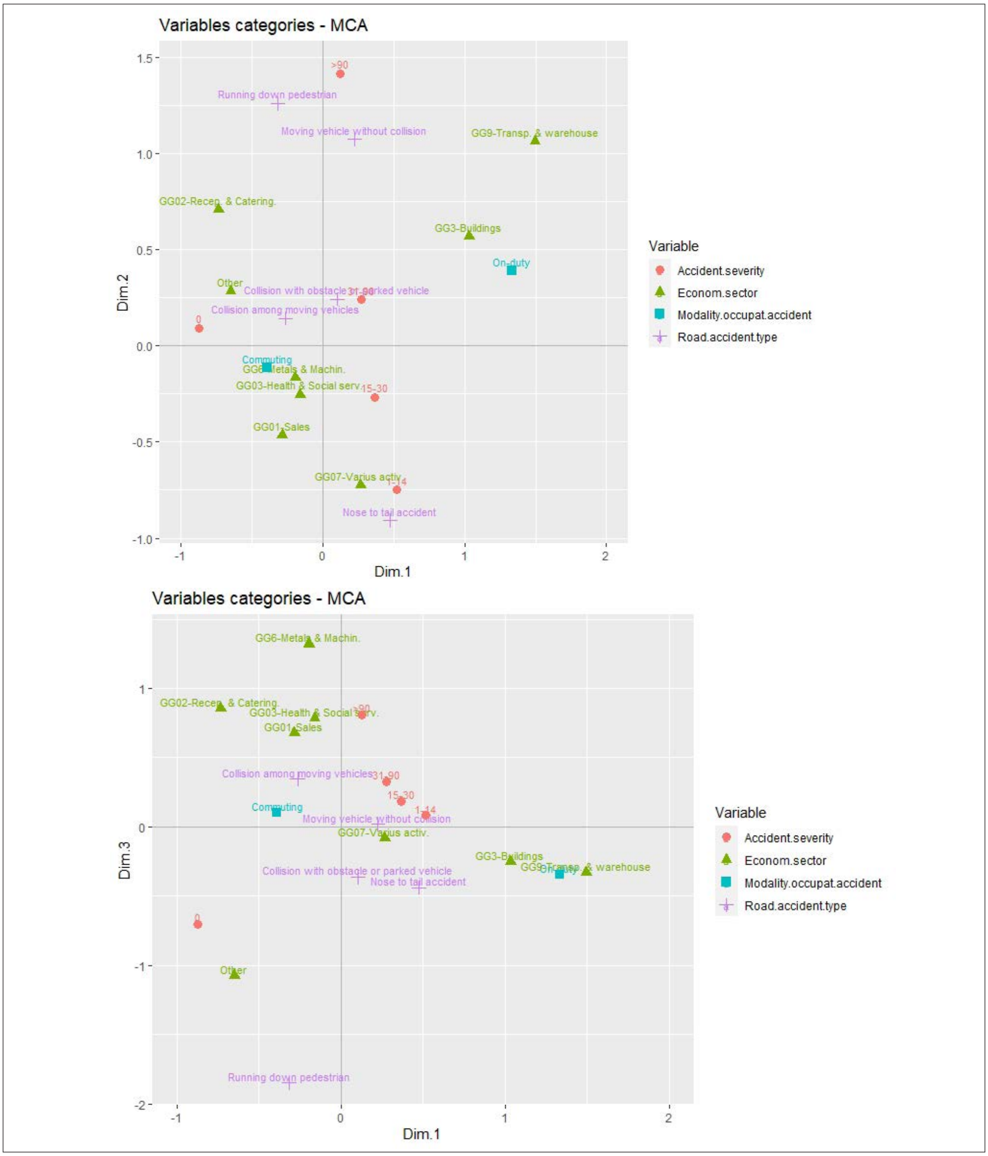
3.3 Multiple correspondence analysis results

MCA results indicate that more than ten dimensions are needed to explain about 67% of the observed variance. Figure A3 in the Appendix shows a scree plot of explained variance by each MCA dimension. It can be seen that the three principal components explain about 23% of the observed variance. Figure A4 in the Appendix shows the correlation between variables and the MCA principal dimensions. It can be seen that the modality of the occupational accident is correlated with dimension 1 and the type of accident with dimension 2. Figure 3.3 shows two plots of coordinates of each variable categories with dimensions 1 and 2 (upper figure) and dimensions 1 and 3. In these plots, variable categories with a similar profile are grouped together and negatively correlated variable categories are positioned on opposite sides of the plot origin. Furthermore, the distance between category points and the origin measures the quality of the variable category on the factor map. Category points that are away from the origin are well represented on the factor map.

Figure 3.3 shows that the two modalities of occupational accidents (on-duty and commuting) are inversely correlated, as well as the nose-to-tail accident with the others types of accidents. The three variable categories On-duty, GG3-Buildings and GG9-transports and warehouse seem to group, suggesting that these economic sectors are related with on-duty accidents. Another group is formed by accidents during commuting, occurring in GG6-Metals and machinery sector, GG03-Health and social services sector and GG01-Sales economic sector with a severity of 15-30 days of leave. The "Running down pedestrian" and "Moving vehicle without collision" types of accidents are close to the 'more than 90 days of leave' category, meaning that a greater severity is associated to these kind of accident. Finally, the "Nose-to-tail accident" likely groups with accidents occurring in the GG07-Various activities economic sector, and with a severity of 1-14 days of leave. It confirms that this kind of accident is not related with a particular economic sector but spreads over miscellanea of them.

The contributions of variables categories to MCA principal dimensions is shown in Figure A5 of Appendix. The categories On-duty, GG9-Transports & warehouse, Nose-to-tail accident, the severities with days of leave >90, 0 and 1-14 days, GG07-Various activities sector and commuting for the modality of occupational accidents, are the most important in the definition of MCA dimensions 1 and 2. However, since a low percentage of variance is explained by the first two principal dimensions (16%), more dimensions would be needed to fully describe the contribution by variables categories.

Figure 3.3 - Coordinates of each variables categories in each dimension. Dim. 1 *vs.* Dim. 2 (upper figure), Dim. 1 *vs.* Dim. 3 (bottom figure)



Source: Authors' processing on Inail and Istat datasets

4. Discussion

Work-related road accidents are known to be a significant portion of the registered occupational accidents (WHO, 2018). Their contribution to the total road accidents is unknown as the registration systems often fail in collecting information about the possible occupational origin of these events. The requests for compensation for occupational injuries are consequently the most reliable available data sources to study the phenomenon. Such archives contain information about the individuals claiming for compensation, the municipality where the accident occurred, the health consequences, the economic sector in which the workers was employed, the duration of leave and the degree of impairment, but nothing about the characteristics of the road accident which determined the request for compensation. To overcome this limitation, this work developed a data linkage procedure which combined at individual level the data provided by the Italian road accidents archive, with the data of requests for compensation for occupational injuries. This is the first multi-year combined work-related road accidents data archive ever provided in Italy. It allows, for the first time, to study not only the occupational characteristics of this phenomenon, but also the descriptors of road accidents.

The main findings of this study are the statistical description of the dataset, the results of the time series analysis, the classification of the accidents in terms of road, nature, localisation and involved macro-economic sectors parameters, as well as the reduction of dimensionality to represent the underlying structures of data.

The statistical description of this dataset highlights the greater occurrence of commuting road accidents with respect to those occurring during on-duty activities (77 *vs.* 23%). These results are well known from a number of statistical reports published by Inail about work-related road accidents (Inail, 2020; Brusco *et al.*, 2019). Men are more represented than women (62 *vs.* 38%) and this result reflects what was reported in the last annual report of Inail (Inail, 2020). A similar result is obtained about the disproportion between fatal and non-fatal events (1.5 *vs.* 98.5%). Consequently, as far as the occupational parameter are concerned, the dataset used in this study is consistent with those reported by Inail in its annual reports about occupational accidents. The economic macro-sector ascribed to various activities (GG0) is found to be the most involved one (54%). It covers a number of miscellaneous activities composed by different type of services and reflects the tertiary component of the Italian economic structure. Minor contributions are detected for the transport and warehouses sector, the metals and machinery sector and the building and installation sector. A novel additional information provided by this study is about the road accidents characteristics. Collisions due to moving vehicles or nose-to-tail ones are the two most frequent types of work-related accidents (55 and 26% respectively). The accidents involve mainly private cars and motorcycles (54 and 26% respectively), with an additional contribution of heavy vehicles (9%). As for motorcycles during the last decade, the traffic congestion occurring in metropolitan areas has increased the use of such travelling mode (Automobile Club d'Italia 2021). This has dramatically increased the number of accidents involving motorcycles with a correspondent increase in number of motorcyclists dead or injured. The result is corroborated by the findings of this study, which addressed the urban roads as the most frequent location where accidents occur (62%), followed by the suburban roads (29%) with straights and crossroads as the two most frequent types of localisation where accidents occur.

These results confirm those obtained by both national reports (Istat, 2020) and published research papers (Eboli, Forciniti, and Mazzulla, 2020) obtained for indistinct road accidents.

The work-related road accidents were found to have strong weekly and seasonal components. The former mainly depends on the organisation and scheduling of work activities, which is distributed in working days and reduced in weekends. The latter component has a periodical behaviour with a minimum in August and during holidays periods as well as an increasing monthly trend from winter to middle of summer. The number of yearly road accidents was found almost constant (about 25,000 per year) in compliance with findings in France (Charbotel, Martin, and Chiron, 2010), but a slightly variation in the long-term trend was detected.

A novel result of this study was the classification of work-related road accidents in terms of accident and occupational parameters. To account for health outcome, representative road accidents were identified by clustering data for fatal and non-fatal outcome and each of them by accidents occurring during commuting or on-duty activities. As for non-fatal on-duty events we found five clusters, accounting for 41% of all events, which describe road accidents occurring using a private car in a urban road, driven by workers at age 23-60 and involving the various activities macro-economic sector, the transport and warehouses sector, and the metals and machinery sector. Others types of non-fatal accidents on-duty were those involving heavy vehicles, involving five clusters and contributing for 27% of these kind of events. The descriptors of these accidents are suburban roads or motorways, moving vehicles or nose-to-tail accident, as accident type, and involve mainly the transport and warehouses sector, the building and installations sector, and the various activities economic sector. The results about heavy vehicles are consistent with those obtained in Australia (Boufous and Williamson 2006) and France (Charbotel, Martin, and Chiron, 2010; Hours *et al.*, 2011). As for non-fatal road accidents during commuting, we found two large groups of clusters. The first one is characterised by accidents occurring in urban or suburban roads in proximity of straights or crossroads, and involving private cars. This group of clusters contributes for about 62% of this kind of events and involve primarily the various activities sector and secondarily the transport and warehouses sector, the metal and machinery sector and the building and installation sector. The second group is composed by clusters involving a motorcycle in an accident occurring in an urban road due to a collision among moving vehicles. They account for about 33% of the non-fatal commuting events and involve mainly the various activities economic sector with smaller contributions from the metal and machinery sector, the transport and warehouses sector and the building and installation one. With respect to on-duty accidents, commuting ones show a larger spread in the involved economic sectors, mainly for both the larger number of workers involved and the number of economic sectors.

The results of the cluster analysis for fatal work-related road accidents show similar characteristics as those about non-fatal ones, with a few remarkable differences like the involvement of pedestrians and a larger contribution of heavy vehicles compared to private cars in on-duty accidents.

The multiple correspondence analysis provided results indicating how the different types of accidents group when the modality of occupational accidents, economic sector and severity of events (measured as number of days of leave) are taken into account. The large heterogeneity of data did not allow for describing their variance with a few principal components. Using three dimensions, the MCA analysis was able to explain about 23% of the total variance. Results showed that on-duty accidents grouped with the “Transports and warehouse” sector, as well as with the “Building “economic sector. It confirms the results of cluster analysis for the on-duty accidents datasets. These economic sectors are known to be exposed as transports of goods. The accidents occurring during commuting were found by MCA analysis to group in the first two dimensions with the “Metal & machinery” sector, the “Health & social services” sector, and the “Sales” economic sector, as well as with the 15-30 days of leave group. On a lesser extent, the GG07-Various activities sector was found also linked with accidents during commuting. Such results partially confirm those obtained by the cluster analysis, but also allow us to distinguish the contribution of the “Health and social services” sector and the Sales sector, which were aggregated in the larger GG0-various activities economic sector when cluster analysed. Conversely, to on-duty accidents, such commuting events seem to be related with multiple economic sectors, confirming their heterogeneous origin.

Based on the above findings, we can consider that work-related road accidents are mainly involved in non-fatal events occurring during commuting and in a limited extent by events during on-duty activities. The cluster results provided the occupational and accident characteristics of these events, identifying key descriptors and consequently possible directions of prevention measures aimed to reduce the occurrence of such events. The home-work journey by car or motorcycle is found at risk particularly in urban and sub-urban areas where both traffic and intersections are quite high. These commuting events involve mainly the service industry sector in which miscellanea of different activities can be found. Among the possible measures to reduce the incidence of such commuting events, we should consider to limit the use of private vehicles by increasing the public transports particularly during rush hours in urban areas. An additional measure could be to increase the number of workers who could work at home at least for those kind of jobs in which this work activity can be applied. Such measure can be applied particularly in the service industry sector, which has been identified in this study as that mainly involved in such events. A recent national report (Inail, 2021) address the smart working during the COVID-19 as one of the key factor of the observed reduction in the work-related road accidents. The road-accidents occurring during on-duty activities were found to involve specific economic sectors such as transport and warehouses, service sector, as well as building and installation. To reduce such events, the limitation in the use of vehicle is not a reliable solution. General prevention measures, such as the limitation of exposure by reducing the number of working hours and the introduction of rest time, as well as training and information sessions about road safety, could help in reducing such on-duty events. Fatigue, stress and sleepiness are other factors to be considered in work-related accidents and prevention measures could be identified to limit them.

This work has provided useful information in characterising and classifying the work-related road accidents phenomenon occurring in Italy. However, it contains a number of limitations. We could not relate the accidents with both the individual aspects of involved drivers and the information about internal and external concurring factors linked with the accident. As for the individual aspects important features like fatigue, scheduling issues (organisational, procedural and timely measures), physiological conditions of drivers and its driving performance or drug assumptions were not available. Other concurring circumstances like traffic, speed of involved vehicles were not available. All this information is impossible to be retrieved at an individual level. The lack of this information does not allow us to investigate about the reasons of these accidents, and whether they are related to job tasks. However, such studies can be carried out at individual levels for specific type of workers, and it was out of the scope of this work, which was focussed on a more general picture of work-related accidents. In addition, there is an unknown number of work-related road accidents which were missing in this archive. First, some categories of workers were missing in the compensation data (about 20% of workforce), although a few of them are not expected to largely contribute to occupational road accidents (eg. armed forces, firefighters, police workers and air transport personnel) and consequently affect the findings of this study. However, autonomous tradespeople and professionals with VAT registration are missing in the Inail compensation archive, but could be included in the ISTAT one. They represent about 21% of the total workforce. In case of road accidents, these types of job can be hardly classified as commuting or on-duty events, as far as the heterogeneities of these kind of jobs are concerned. The lack of their inclusion in this study can be considered a possible limitation, but they should not modify the findings presumably. Second, some missing events come from the efficiency of the linkage procedure, which depends on the correctness of the information contained in both the accidents and the compensation claims archives. Uncertainty in the linkage process could come from the probabilistic component of the linkage procedure, as for the deterministic pass, accuracy is enforced by the nature of the variables selected. The inclusion of additional filters allows accounting for most of the linked pairs (more than 90%). More than 95% of those pairs are localised in the same province or municipalities geographically close. This gives us enough confidence on the accuracy of the algorithm. In addition, the equality of the identifying fields make extremely narrow the chances that they can refer to distinct accidents (Taiano et. al., 2021). Consequently, uncertainties regarding the linkage process could not potentially influence the results. Other missing data could come from the lack of registration in one or both archives. As an example, a road accident cannot be registered in case of a missed intervention of local authorities or a registered accident could lack of a request for compensation and not be registered as an occupational injuries. For these reasons, the actual dimension of the phenomenon cannot be figured. We think that the combined archive is sufficiently representative of the work-related accidents phenomenon, as demonstrated by the consistency of other results with those obtained from this archive.

At the same time, this study addressed some of the limitations of work-related data systems in providing a more complete picture of the circumstances of occupational road accidents.

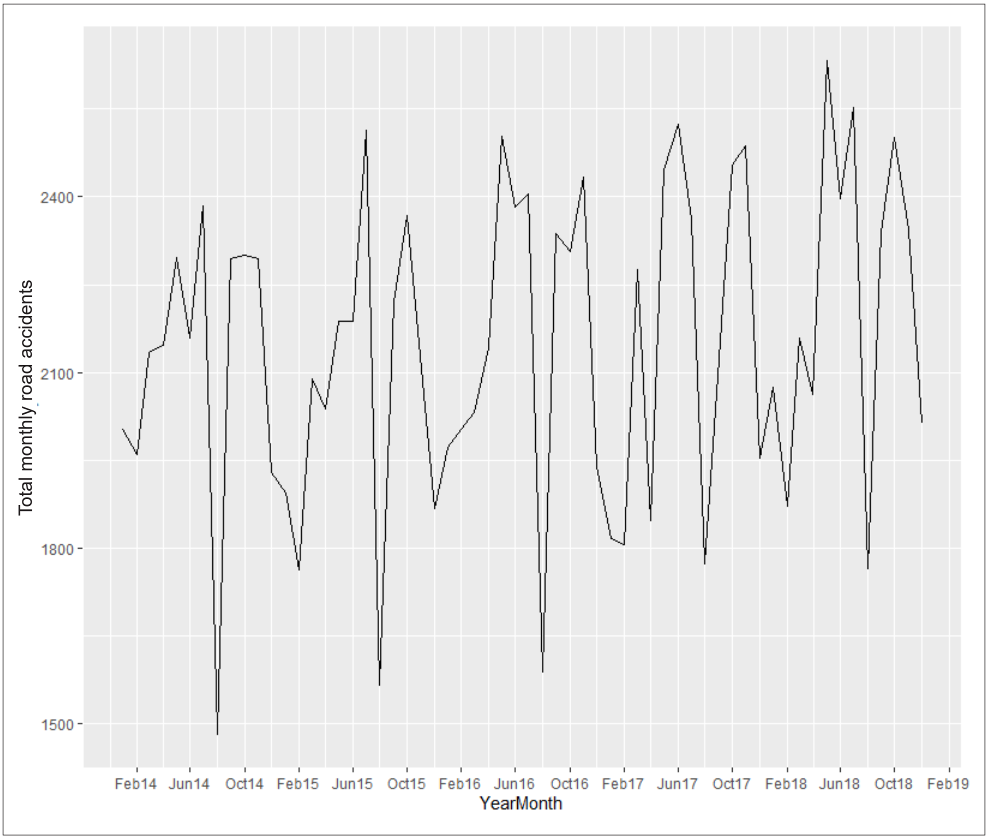
5. Conclusions

The work-related road accidents represent a significant portion of occupational accidents. The intrinsic limitations in assessing and monitoring such phenomenon and in identifying accident and occupational determinants where faced by combining linking the road accidents archive and the compensation claims for occupational injuries archive at individual level. Data were statistically analysed to extract combined information otherwise not available by evaluating them separately. Results address the urban and suburban areas as the most dangerous places where these events occur. Different types of collision are found as the origin of accidents, using private cars and motorcycles as preferred vehicle. Time series analysis highlights a strong weekly and seasonal variation. The cluster analysis allows to identify typical road accidents profiles, including both accident parameters and involved economic sectors, by health outcomes and modality of occupational accident. Both cluster and MCA analysis found that road accidents during on-duty service are dominated by the “Transports and warehouse” sector and by the “Buildings” economic sector, while commuting accidents can be ascribed to more heterogeneous sectors.

Based on these findings, prevention and reduction policies are needed to limit the impact on health and social costs of road accidents in general and of work-related ones in particular.

Appendix

Figure A.1 - Time series of the total monthly work-related road accidents occurring in Italy from 2014 to 2018

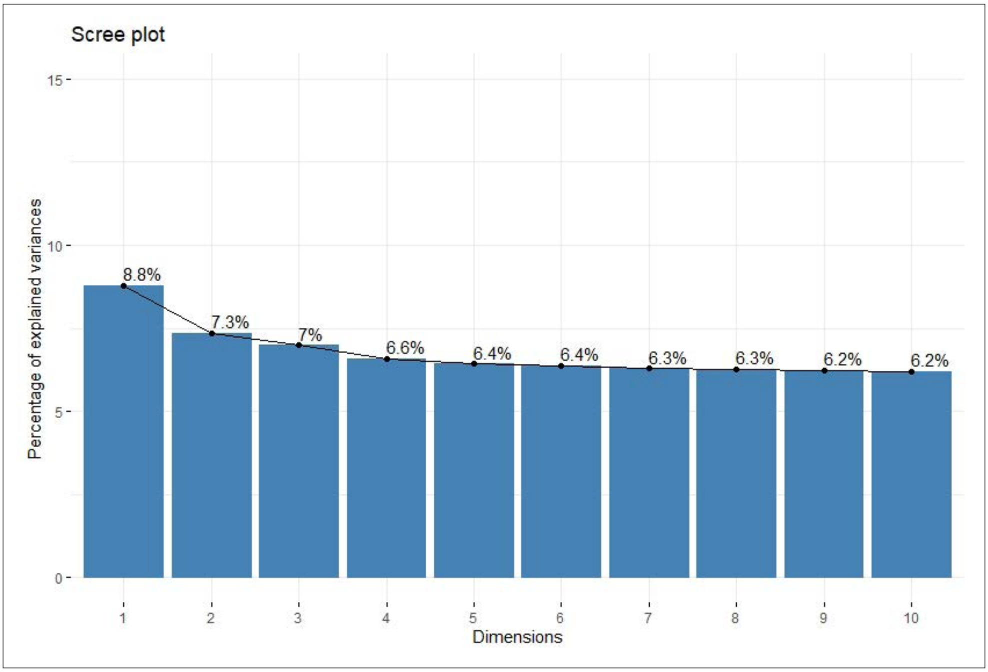


Source: Authors' processing on Inail and Istat datasets

Figure A.2 - Distribution of characteristics among clusters in fatal work-related road accidents

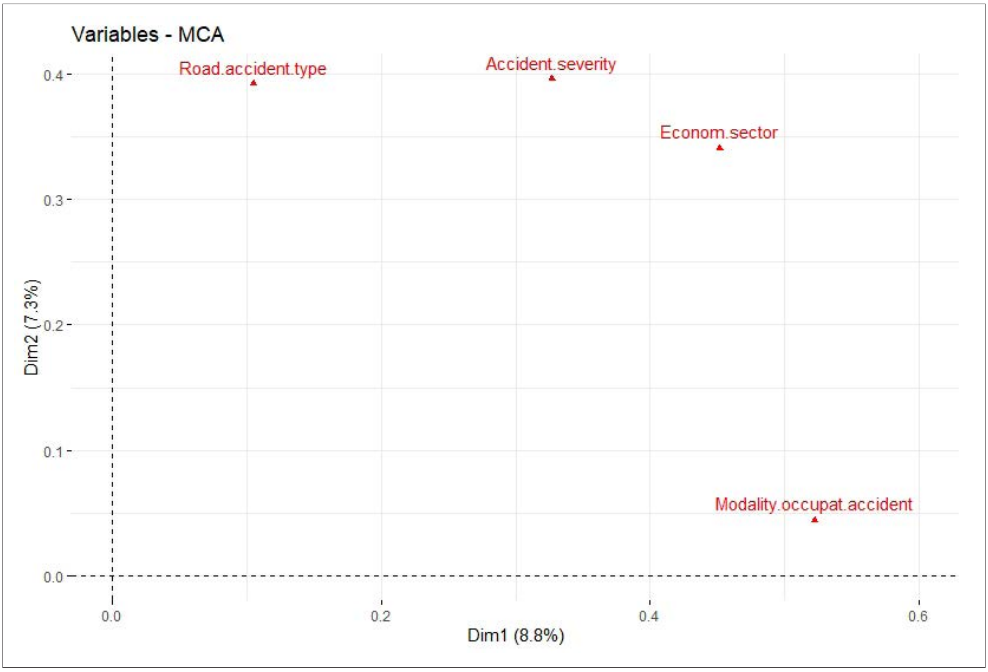
Source: Authors' processing on Inail and Istat datasets

Figure A.3 - Percentage of explained variances by each MCA dimension



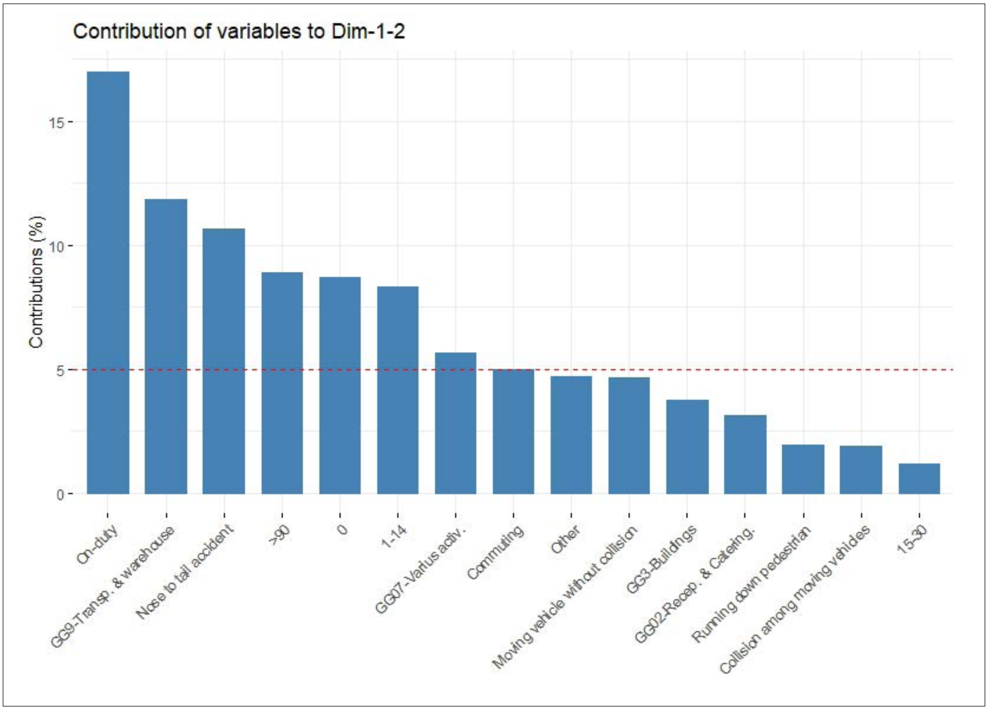
Source: Authors' processing on Inail and Istat datasets

Figure A.4 - Correlation between variables and MCA principal dimensions



Source: Authors' processing on Inail and Istat datasets

Figure A.5 - Contribution of variable categories to MCA principal dimensions. The red dashed line on the graph above indicates the expected average value, if the contributions were uniform



Source: Authors' processing on Inail and Istat datasets

References

Automobile Club d'Italia - ACI (Italian Automotive Club). 2021. Annuario Statistico 2021. https://www.aci.it/laci/studi-e-ricerche/dati-e-statistiche/annuario-statistico/annuario-statistico-2021.html.

Abdi, H., and L.J. Williams. 2010. "Principal Component Analysis". *WIREs Computational Statistics*, Volume 2, Issue 4: 433-459.

Ahmed, M.M., R. Franke, K. Ksaibati, and D.S. Shinstine. 2018. "Effects of truck traffic on crash injury severity on rural highways in Wyoming using Bayesian binary logit models". *Accident Analysis & Prevention*, Volume 117: 106-113.

Boufous, S., and A. Williamson. 2006. "Work-related traffic crashes: a record linkage study". *Accident Analysis & Prevention*, Volume 38, Issue 1: 14-21.

Brusco, A., A. Bucciarelli, M. Bugani, C. Gariazzo, C. Giliberti, M. Marinaccio, S. Massari, A. Pireddu, L. Veronico, G. Baldassarre, S. Bruzzone, M. Scortichini, M. Stafoggia, e S. Salerno. 2019. *Gli incidenti con mezzo di trasporto. Un'analisi integrata dei determinanti e dei fattori di rischio occupazionali*. Roma, Italy: Inail.

Charbotel, B., J.L. Martin, and M. Chiron. 2010. "Work-related versus non-work-related road accidents, developments in the last decade in France". *Accident Analysis & Prevention*, Volume 42, Issue 2: 604-611.

Cleveland, R.B., W.S. Cleveland, J.E. McRae, and I. Terpenning. 1990. "STL: A Seasonal-Trend Decomposition Procedure Based on Loess". *Journal of Official Statistics - JOS*, Volume 6, Issue 1: 3-73.

Eboli, L., C. Forciniti, and G. Mazzulla. 2020. "Factors influencing accident severity: an analysis by road accident type". In *Transportation Research Procedia*, Volume 47: 449-456.

Fort, E., L. Pourcel, P. Davezies, C. Renaux, M. Chiron, and B. Charbotel. 2010. "Road accidents, an occupational risk". *Safety Science*, Volume 48, Issue 10: 1412-1420.

Gariazzo, C., S. Bruzzone, S. Finardi, M. Scortichini, L. Veronico, and A. Marinaccio. 2021. "Association between extreme ambient temperatures and general indistinct and work-related road crashes. A nationwide study in Italy". *Accident Analysis & Prevention*, Volume 155, 106110.

Hours, M., E. Fort, B. Charbotel, and M. Chiron. 2011. "Jobs at risk of work-related road crashes: an analysis of the casualties from the Rhône Road Trauma Registry (France)". *Safety Science*, Volume 49, Issue 8-9: 1270-1276.

Istituto Nazionale di Statistica - Istat (*Italian National Institute of Statistics - Istat*). 2021. "Road Accidents. Year 2020". *Press Release*. Roma, Italy: Istat. https://www.istat.it/en/archivio/259830.

Istituto Nazionale di Statistica - Istat (*Italian National Institute of Statistics - Istat*). 2020. "Road Accidents. Year 2019". *Press Release*. Roma, Italy: Istat. https://www.istat.it/en/archivio/245981.

Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro - Inail (*Italian National Institute for Insurance against Accidents at Work - Inail*). 2021. Andamento degli infortuni sul lavoro e delle malattie professionali. 2021. Roma, Italy: Inail. https://www.inail.it/cs/internet/docs/alg-dati-inail-2021-giugno-luglio-pdf.pdf.

Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro - Inail (*Italian National Institute for Insurance against Accidents at Work - Inail*). 2020. Relazione Annuale 2019. Roma, Italy: Inail.

https://www.inail.it/cs/internet/comunicazione/pubblicazioni/rapporti-e-relazioni-inail/relazione-annuale-anno-2019.html.

Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro - Inail (*Italian National Institute for Insurance against Accidents at Work - Inail*). 2019. *Banca dati statistica prevenzionale*. Roma, Italy: Inail. https://www.inail.it/cs/internet/attivita/dati-e-statistiche/banca--dati-statistica.html.

Kaufman, L., and P.J. Rousseeuw. 2005. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ, U.S.: John Wiley & Sons, *Wiley Series in Probability and Statistics*.

Mitchell, R.J., M.R. Bambach, and R. Friswell. 2014. "Work and non-work-related vehicle crashes: the contribution of risky driving practices". *Safety Science*, Volume 68: 65-72.

Moomen, M., M. Rezapour, and K. Ksaibati. 2019. "An investigation of influential factors of downgrade truck crashes: A logistic regression approach". *Journal of Traffic and Transportation Engineering (English Edition)*, Volume 6, Issue 2: 185-195.

Naik, B., L.-W. Tung, S. Zhao, and A.J. Khattak. 2016. "Weather impacts on single-vehicle truck crash injury severity". *Journal of Safety Research*, Volume 58: 57-65.

Newnam, S., D. Blower, L. Molnar, D. Eby, and S. Koppel. 2018. "Exploring crash characteristics and injury outcomes among older truck drivers: An analysis of truck-involved crash data in the United States". *Safety Science*, Volume 106: 140-145.

Öz, B., T. Özkan, and T. Lajunen. 2010. "Professional and non-professional drivers' stress reactions and risky driving". *Transportation Research Part F: Traffic Psychology and Behaviour*, Volume 13, Issue 1: 32-40.

Pei, X., S.C. Wong, and N.N. Sze. 2012. "The roles of exposure and speed in road safety analysis". *Accident Analysis & Prevention*, Volume 48: 464-471.

Robb, G., S. Sultana, S. Ameratunga, and R. Jackson. 2008. "A systematic review of epidemiological studies investigating risk factors for work-related road traffic crashes and injuries". *Injury Prevention*, Volume 14, Issue 1: 51-58.

Taiano, L., S. Massari, T. Tuoto, L. Valentino, S. Bruzzone, and L. Veronico. 2021. "Work-related road accidents: a data linkage procedure applied to assess traffic accidents at work and commuting". In ths issue of the *Rivista di statistica ufficiale*, N. 3/2021. Roma, Italy: Istat.

Uddin, M., and N. Huynh. 2020. "Injury severity analysis of truck-involved crashes under different weather conditions". *Accident Analysis & Prevention*, Volume 141, 105529.

Wang, X., K. Smith, and R. Hyndman. 2006. "Characteristic-Based Clustering for Time Series Data". *Data Mining and Knowledge Discovery*, Volume 13, Issue 3: 335-364.

World Health Organization - WHO. 2018. *Global Status Report on Road Safety 2018*. Geneva, Switzerland: WHO.

Wu, C.Y.H., B.F. Zaitchik, and J.M. Gohlke. 2018. "Heat waves and fatal traffic crashes in the continental United States". *Accident Analysis & Prevention*, Volume 119: 195-201.

An analysis of work-related road injuries by macro-economic sector, road type and Italian territorial divisions

Antonella Pireddu [1], Antonella Altimari [1], Giordana Baldassarre [2], Alessandro Marinaccio [1], Luca Taiano [1]

Abstract

Information about accidents is essential for studying road safety: work-related journeys constitute a significant part of traffic problems. By linking data from the archives of both the Italian National Institute for Insurance against Accidents at Work - Inail and the Italian National Institute of Statistics - Istat, we were able to analyse approximately 129,000 work-related road accidents that occurred in Italy during the period 2014-2018.

The nationwide study investigated on injury rates by road type and economic sector, otherwise not available in the two archives separately.

Injury rates within and between Italian territorial divisions were analysed by spatial, descriptive analysis and analysis of variance - ANOVA. Maximum injury rates were found in the sales, metals and machinery, transport and warehouses, health and social services, construction and plant sectors. Indexation using the ratio of accident frequencies and the number of employees was useful for expressing a dynamic assessment of how many work-related road injuries "derive" from the total number of workers employed in a sector. ANOVA analysis showed that injury variances on rural and urban roads, as well as in macro-economic sectors differed in Italian

1 Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro (*Italian National Institute for Insurance against Accidents at Work*) - Inail.

2 Istituto Nazionale di Statistica (*Italian National Institute of Statistics*) - Istat.

In the writing of this article, the authors were supported by all the members of the Inail-Istat collaborative group: Antonella Altimari, Michela Bonafede, Roberto Boscioni, Claudio Gariazzo, Alessandro Marinaccio, Stefania Massari, Antonella Pireddu, Luca Taiano and Liana Veronico (for Inail); Giordana Baldassarre and Silvia Bruzzone (for Istat).
Although this article is the result of all the authors' commitment, the paragraphs are attributed as following: Antonella Pireddu: Conceptualisation, Data curation, Methodology, Writing; Alessandro Marinaccio: Methodology; Antonella Altimari, Giordana Baldassarre, Luca Taiano: Data curation.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

territorial divisions. This study is also intended for local public health authorities responsible for assessing and managing risks in specific contexts.

Keywords: Work-related road injuries, territorial division, macro-economic sector, road type, ANOVA.

1. Introduction

Data about accidents provide essential information for road safety (Hollò *et al.*, 2010). The causality of work-related road (W-RR) accidents is essential since the journey from home to work (commuting) accounts for a significant part of urban and rural traffic.

A recent study analysed 20,941 W-RR accidents by linking data concerning general road accidents registered by the Italian National Institute of Statistics (Istat) and occupational accidents registered by the Italian National Institute for Insurance against Accidents at Work (Inail) that occurred in Italy in 2015 (Brusco *et al.*, 2019).

The disparity between the two integrated archives, which emerged from the same study, was related to the incomplete nature of the records. Over the years, the addition of more detailed data has reduced this gap. A further reason for a lack of correspondence between the two archives was attributable to a failure to collect information on the injured person and on the real reasons for his/her journey.

Despite these limitations, for the first time in Italy the 2015 study assembled in a single dataset both occupational information on road user injuries and the context of the road accident. Currently, no other archives containing occupational and contextual information on accidents are available in Italy. To fill this gap, a new integrated dataset of W-RR accidents, based on a longer period of observation and containing about 129,000 records on collisions resulting in death or injury, has been made (Bruzzone *et al.*, 2021; Taiano *et al.*, 2021).

We conducted an analysis designed to evaluate the variance of approximately 129,000 injuries by road type (Legislative Decree 30 April 1992 n. 285/*Decreto legislativo* n. 285/1992) and macro-economic sector in Italian territorial divisions. This included an analysis of road and accident data, a thorough spatial descriptive analysis of W-RR injuries by macro-economic group, road type and location, integrated with an analysis of variance (ANOVA), which provided variance associated the two accident parameters (road type and macro-economic sector), within and between and the 5 Italian territorial divisions. This study is also intended for professionals in the territorial divisions who are responsible for assessing and managing risks in specific contexts.

2. Data source

2.1 The Italian situation

The Italian territory extends for 302,073 sq. km (Istat, 2018), and is divided into five territorial areas: the North West (approx. 58,000 sq. km), the North East (62,000 sq. km), the Centre (58,000 sq. km), the South (74,000 sq. km) and the Islands (50,000 sq. km). Roads, widely distributed over these five Italian divisions (*OpenStreetMap* data) stretch for about 128,000 km in the North West (2.2 km of infrastructure per sq. km), 131,000 km in the North East (2.11), 114,000 km in the Centre (1.96), 147,000 km in the South (2.01) and 76,000 km in the Islands (1.52). The number of "estimated workers per year"[3] was over 17,000,000 in 2018 (6,237,773 in the North West, 4,047,193 in the North East, 3,958,636 in the Centre, 2,163,973 in the South and 979,553 in the Islands). In 2018, over 172,000 road accidents in Italy resulted in 3,334 deaths (within 30 days) and over 240,000 injuries (source Istat and the Police force). During the five-year period 2014-2018, Istat recorded a total number of 874,847 road accidents involving about 1,253,715 persons injured or dead (254,528 persons in 2014 and 246,253 in 2018), since the single accident, both work-related and non-work-related, certainly, is associated with one or more injured or dead persons (Istat, 2021*a*).

Approximately 15% of the over 600,000 work accidents reported to Inail in 2018 were road accidents. During the five-year period 2014-2018, 465,943 work-related road injuries were recorded by Inail (93,056 in 2014 and 94,553 in 2018). According to Inail data, each injured person was associated with only one traffic accident.

Between these two data sources, there is an "intersection set" based on road injuries present in both the Istat and the Inail archives.

3 Inail data. For Industry and Services in both ATECO and LARGE TARIFF GROUPS (macro-economic sector), Inail estimates employees per year as the ratio between declared wages and the average daily wage per 300 (theoretical number of working days per year excluding holidays). Self-employed workers, apprentices, craftsmen and non-craftsmen, and members of porters' or fishermen' cooperatives are excluded from this total as are undefined categories. Furthermore, in companies with branches all over Italy, all employees are generally assigned only to the headquarters. Finally, in Italy, the ATECO classification assigns the same Class to each worker of each company. Data available on the URL: https://www.inail.it/cs/internet/attivita/dati-e-statistiche/banca--dati-statistica.html. 2018, Updated to 31/10/2020.

2.2 The Inail and Istat archives

RELAIS (REcord Linkage At IStat) integration of the data linkage between the Inail archives on work-related accidents and the Istat archives concerning road accidents resulting in death or injury revealed 128,837 injuries (drivers, passengers or pedestrians, injured or dead within 24 hours or in the first 30 days following the accident) that occurred in Italy in the period from 2014 to 2018. The integrated dataset encompasses complementary information, as the former expresses occupational parameters while the latter refers to road accident parameters.

The Istat archives contain, indeed, complete data on all reported collisions (Istat data warehouse I.Stat. 2014-2018, 2019) on Italian roads involving at least one vehicle. The Istat archives contain the most complete and accurate information available about road accidents in Italy and include accident parameters such as road type, road user or driver, vehicle type and environment as well as the geographic coordinates of the road point where the collision occurred (Pireddu and Bruzzone, 2021).

The Inail archives contain the most complete and accurate information available about work-related road accidents in Italy. The archives cover about 80% of the Italian workforce (Inail, 2019), but do not include those employed in the armed forces, fire brigades, the police force, air transport personnel, independent traders and professionals with VAT registration (Gariazzo *et al.*, 2021). The Inail archives include the social (injury claims due to work-related road accidents and compensation data for injured workers or their families in case of death) as well as the occupational parameters of workers and the relevant economic sector of activity, in addition to information on the involvement or non-involvement of a vehicle at the time of the accident, which are classified as injuries *at work* and during *commuting*.

The occupational variables, such as the activity in which workers are engaged, are classified into two groups: ATECO (Istat) and LARGE TARIFF GROUP[4] or macro-economic sector (Inail). The former is the Italian standard

4 Large tariff group refers to the highest class aggregating items of the 4 tariffs applied to the management of industry and services (industry, crafts, tertiary, other activities). There are 10 classes that conglomerate the processes of each large tariff group (macro-economic sector) according to homogeneous production sectors (1. agricultural processing and food; 2. chemical, paper and leather; 3. construction and plants; 4. energy and communications; 5. wood and the like; 6. metals and machinery; 7. mining, minerals and glass; 8. textiles and

classification of productive economic activities derived from the *Nomenclature générale des Activités économiques dans les Communautés Européennes* (Vicari *et al.*, 2007). Some critical issues are connected with ATECO on account of the significant number of undefined cases. Furthermore, according to the ATECO classification, the geographic location of each workplace frequently indicates the location of the company headquarters and not the place where the worker actually commutes from home to work or drives around during on-duty service. For each company in Italy, the ATECO classification assigns the same code to all its workers, regardless of the activity actually carried out by each employee. These aspects could cause distortions that affect spatial analysis indexing[5] and the calculation of rates by sectors if several units are distributed all over the country.

On the other hand, the classification of work activity by macro-economic sector (see Footnote 1) is more consistent with the work actually carried out by the injured person and with the location where the worker commutes from home to work or where the accident takes place. Although this classification is partially incomplete due to undefined cases, it was more suitable for our analysis. Therefore for the purposes of this study, only the Inail classification by macro-economic sector was used.

packaging; 9. transport and warehouses; 10. various activities). The large tariff group may contain different items for different tariffs. Technical reference: [Inail-010b, 11]. Url: https://dati.inail.it/opendata/default/Infortuni/index.html. For the purpose of this study "macro-economic sector" is used instead of "Large tariff group".

5 The W-RR injury index by province and macro-economic sector, based on the estimated number of employees (see footnote 1) is calculated using the formula (1) Index = N° W-RR Injuries/ N° Employees by macro-economic sector.

3. Methods

Data concerning approximately 129,000 W-RR accidents that occurred in Italy in the period from 2014 to 2018 were analysed. The study included a descriptive and detailed spatial analysis (Cima *et al.*, 2014; Costabile *et al.*, 2012) of accidents aggregated by economic sector, location and road type and a One Way Analysis of Variances (ANOVA) of W-RR injuries by road type and macro-economic sector (see Footnote 1), between and within the 5 aforementioned Italian territorial divisions. ANOVA is an inferential technique which, when applied to two or more groups of data, enables us to compare internal variability with variability between the groups. This technique assumes that variance can be divided into two components: within groups and between groups. The null hypothesis (H_0) commonly assumes that the data of all groups have the same stochastic distribution and that the observed differences between the groups are due to randomness. This division of variance is based on the hypothesis that the differences in W-RR injuries can be explained by the characteristics of the group to which they belong (*i.e.* Italian territorial divisions). When variance between groups contributes significantly to total variance, it can be accepted that the observed difference is linked to the characteristics of the group. Vice versa, when the variance within groups contributes significantly to the total variance, the observed differences can be considered to be linked to the specific case.

The variables analysed were: the macro-economic sector (sales; reception and catering; health and social services; cleaning, sanitation and disinfestation; cinematography, culture and sport; education, research and survey; various and undefined activities; agriculture and food processing; chemical, paper and leather; construction and installation; energy and communications; wood and the like; metals and machinery; mining, minerals and glass; textiles and packaging; transport and warehouses; not classified; public administration; service personnel and students); the road type (motorway; rural road; urban road); the Italian provinces and territorial divisions (the North West; the North East; the Centre; the South; the Islands).

Among the former we analysed and indexed (Hollò *et al.*, 2010) the most relevant in terms of frequency and means *i.e. GG01 sales, GG03 health and social services, G3 construction and installation, GG6 metals and machinery, GG9 transport and warehouses*. The indicators used for the analysis were based on the ratio between the number of accidents and the estimated number of employees in each macro-economic sector (1) Number of W-RR Injuries / Estimated number of employees.

Inail calculates the number of employees using the ratio of the total amount of wages declared and the average daily wage per 300 (theoretical number of working days per year, excluding holidays). For our analysis, the estimated number of employees used in formula (1) was based on the aforementioned Inail criterion. This estimate excluded certain types of workers and undefined cases (see Footnotes 1 and 3).

Our study included a spatial analysis and mapping referring to the macro-economic sector, suitable for a preliminary data investigation. ANOVA (Lix *et al.*, 1996) was applied to the same variables used for descriptive analysis: road type, macro-economic sector and territorial area. The absolute and relative W-RR injury variances were calculated by road type and macro-economic sector between and within the 5 Italian territorial divisions (Table 4.1). At this point, the spatial and descriptive analyses were completed using ANOVA and the Kruskall-Wallis test. The findings are reported below. Spatial analysis was performed using *Quantum Geographic Information System* (QGIS version 3.20.3), whereas ANOVA was performed using *Rstudio* (version 3.18.3) and *PSPP* (version 0.10.5).

4. Results

4.1 Descriptive analysis

The absolute frequencies of W-RR injuries were 42,017 in the North West and 40,920 in the North East, totalling 82,937 units. The Centre (25,996), the South (12,894) and the Islands (7,010) registered approximately 45,900 cases (Tables 4.1 and A1 in Appendix A).

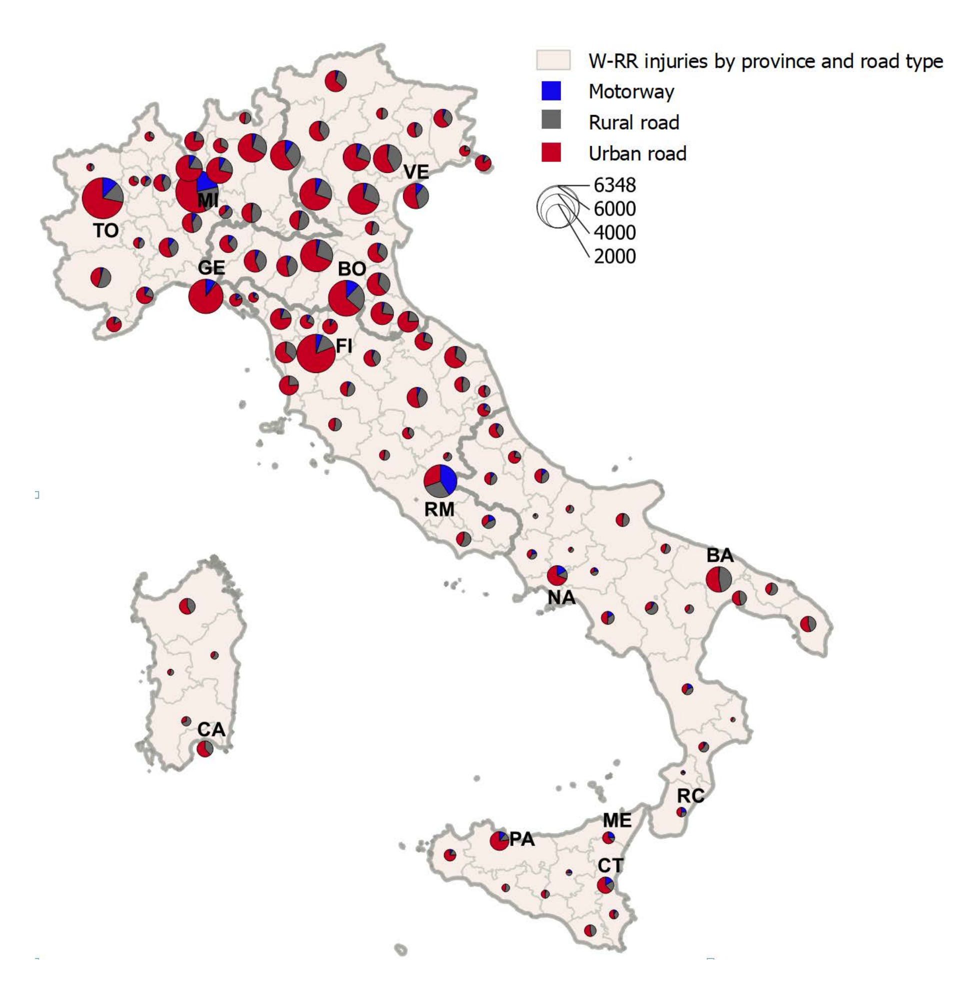
Table 4.1 - Injury frequencies by macro-economic sector, territorial division and road type (absolute values). Five macro-economic sectors. Italy. 2014-2018

Macro-economic sector	Territorial division	Motorway	Rural road	Urban road	Total
GG01 Sales	Centre	215	677	1,459	2,351
	Islands	62	245	513	820
	North East	170	1,007	2,209	3,386
	North West	358	941	2,413	3,712
	South	74	482	601	1,157
	Italy	879	3,352	7,195	11,426
GG03 Health social services	Centre	156	549	1,264	1,969
	Islands	59	261	443	763
	North East	77	790	1,945	2,812
	North West	203	728	2,251	3,182
	South	72	480	683	1,235
	Italy	567	2,808	6,586	9,961
GG3 Construction and plant	Centre	242	600	906	1,748
	Islands	36	160	206	402
	North East	236	961	1,285	2,482
	North West	450	809	1,438	2,697
	South	108	449	326	883
	Italy	1,072	2,979	4,161	8,212
GG6 Metals and machinery	Centre	132	566	1,141	1,839
	Islands	26	87	157	270
	North East	162	1,432	2,610	4,204
	North West	327	1,127	2,776	4,230
	South	54	367	347	768
	Italy	701	3,579	7,031	11,311
GG9 Transport and warehouses	Centre	476	721	999	2,196
	Islands	72	183	244	499
	North East	641	1,034	1,511	3,186
	North West	688	1,061	1,917	3,666
	South	212	543	417	1,172
	Italy	2,089	3,542	5,088	10,719

Source: Authors' processing on integrated Inail - Istat archive data. Italy 2014-2018

Details about annual injury frequency are reported in an accompanying paper (Bruzzone *et al.*, 2021).

Figure 4.1 - W-RR injuries by road type and province (absolute values). Italy. 2014-2018



Source: Authors' processing on integrated Inail - Istat data. OpenStreetMap data. QGIS

Figure 4.1 illustrates the 107 Provinces (grey), the 5 Italian territorial divisions (dark grey) and the 14 Italian metropolitan areas (initialled - MI, RM, *etc.*). For each province the total absolute W-RR injury frequency rate is symbolised by the pie chart diameter (label), while the three segments of the pie chart show the percentage of injuries on motorways (blue), rural roads (grey) and urban roads (red). In the same Figure, the three segments of the pie charts show the different distribution in the Provinces, and within and between the 5 territorial divisions.

The integrated dataset for 2014-2018 demonstrated that injury rates were 10,639 (6%-19%) for motorways; 37,866 for rural roads (28%-33%) and 80,332 for urban roads (47%-62%). The only exception was represented by the province of Rome (Metropolitan City) where the accident rates were distributed almost equally in the three types of road (Figure 4.1). Previous Inail - Istat record linkage (Brusco *et al.*, 2019) indicated that out of the total number of accidents that occurred in 2015, approximately 3% were on motorways, 19% on rural roads and 78% on urban roads.

Table 4.2 - Descriptive Statistics. W-RR injuries by road type and territorial division (absolute values). (95% Confidence Interval for Mean). Italy. 2014-2018

Road type	Territorial division	Province (N)	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Min.	Max.
Motorway	Centre	22	120.09	326.24	69.56	-24.56	264.74	1	1,556
	Islands	14	37.43	56.34	15.06	4.90	69.96	0	157
	North East	22	105.86	114.18	24.34	55.24	156.49	8	542
	North West	25	165.96	293.55	58.71	44.79	287.13	0	1,363
	South	24	41.46	51.45	10.50	19.73	63.19	0	234
	Italy	107	99.43	215.76	20.86	58.08	140.78	0	1,556
Rural road	Centre	22	340.05	240.97	51.37	233.21	446.88	41	1,087
	Islands	14	161.79	99.95	26.71	104.08	219.49	63	373
	North East	22	561.09	308.38	65.75	424.36	697.82	74	1,125
	North West	25	418.60	344.01	68.80	276.60	560.60	55	1,379
	South	24	221.29	201.41	41.11	136.24	306.34	31	1,044
	Italy	107	353.89	293.19	28.34	297.69	410.08	31	1,379
Urban road	Centre	22	721.50	851.00	181.43	344.19	1,098.81	87	4,213
	Islands	14	301.50	270.68	72.34	145.21	457.79	32	957
	North East	22	1,193.05	773.56	164.92	850.07	1,536.02	220	2,887
	North West	25	1,096.12	1,178.05	235.61	609.85	1,582.39	111	4,249
	South	24	274.50	288.23	58.84	152.79	396.21	15	1,211
	Italy	107	750.77	866.13	83.73	584.76	916.77	15	4,249

Source: Authors' processing on integrated Inail - Istat archive data. Italy 2014-2018

Table 4.3 - Descriptive statistics. W-RR injuries by 5 macro-economic sectors and territorial division (absolute values). (95% Confidence Interval for Mean). Italy. 2014-2018

Macro-economic sector	Territorial division	Province (N)	Freq.	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Min.	Max.
GG01 Sales	Centre	22	2,351	106.86	112.93	24.08	56.79	156.94	20	479
	Islands	14	820	58.57	46.65	12.47	31.64	85.51	8	166
	North East	22	3,386	153.91	90.36	19.26	113.85	193.97	28	337
	North West	25	3,712	148.48	145.59	29.12	88.38	208.58	11	521
	South	24	1,157	48.21	47.28	9.65	28.24	68.17	4	230
	Italy	107	11,426	106.79	107.94	10.43	86.10	127.47	4	521
GG03 Health social services	Centre	22	1,969	89.50	86.68	18.48	51.07	127.93	23	389
	Islands	14	763	54.50	40.99	10.95	30.84	78.16	11	132
	North East	22	2,812	127.82	73.06	15.58	95.43	160.21	34	334
	North West	25	3,182	127.28	121.16	24.23	77.24	177.29	19	472
	South	24	1,235	51.46	44.87	9.16	32.51	70.41	8	209
	Italy	107	9,961	93.09	87.35	8.44	76.35	109.84	8	472
GG3 Construction and plant	Centre	22	1,748	79.45	66.88	14.26	49.80	109.11	19	296
	Islands	14	402	28.71	17.96	4.80	18.35	39.08	7	66
	North East	22	2,482	112.82	65.47	13.96	83.79	141.85	21	259
	North West	25	2,697	107.88	92.70	18.54	69.61	146.15	17	341
	South	24	883	36.79	34.92	7.13	22.05	51.54	0	172
	Italy	107	8,212	76.75	71.78	6.94	62.99	90.5	0	341
GG6 Metals and machinery	Centre	22	1,839	83.59	66.11	14.09	54.28	112.9	8	282
	Islands	14	270	19.29	14.37	3.84	10.99	27.58	2	48
	North East	22	4,204	191.09	137.33	29.28	130.20	251.98	47	460
	North West	25	4,230	169.20	184.83	36.97	92.91	245.49	11	665
	South	24	768	32	29.07	5.93	19.72	44.28	1	102
	Italy	107	11,311	105.71	131.63	12.73	80.48	130.94	1	665
GG9 Transport and warehouses	Centre	22	2,196	99.82	124.31	26.50	44.70	154.93	24	534
	Islands	14	499	35.64	27.09	7.24	20.00	51.28	8	83
	North East	22	3,186	144.82	106.79	22.77	97.47	192.16	19	491
	North West	25	3,666	146.64	153.63	30.73	83.23	210.05	9	626
	South	24	1,172	48.83	43.78	8.94	30.35	67.32	7	213
	Italy	107	10,719	100.18	115.06	11.12	78.12	122.23	7	626

Source: Authors' processing on integrated Inail - Istat archive data. Italy 2014-2018

The descriptive analysis based on "accidents by macroeconomic sector (GG) and territorial division" shown in Table 4.3, provides absolute frequencies, averages, standard deviations, standard errors, lower and upper limits of the confidence intervals, minimum and maximum. In all the sectors examined, the highest frequencies and the highest means were found in the North of the country. In particular, the GG6 metal and machinery sector recorded 4,204 cases in the North East and 4,230 cases in the North West (average 169.20 and mean confidence interval 92.91-245.49). Dispersion peaks (standard deviation) were observed in the North West, while the minimum values for the GG6 metal and machinery sector, the GG9 transport and warehouse sector and the GG01 sales sector were found in the Islands. The minimum standard error and therefore the greatest reliability of data was observed in the Islands for the GG6 metals and machinery sector (3.84) and the GG3 construction and plant sector (4.80).

Table 4.4 - Descriptive statistics. W-RR injuries by macro-economic sector and territorial division (relative values). (95% Confidence Interval for Mean). Italy. 2014-2018

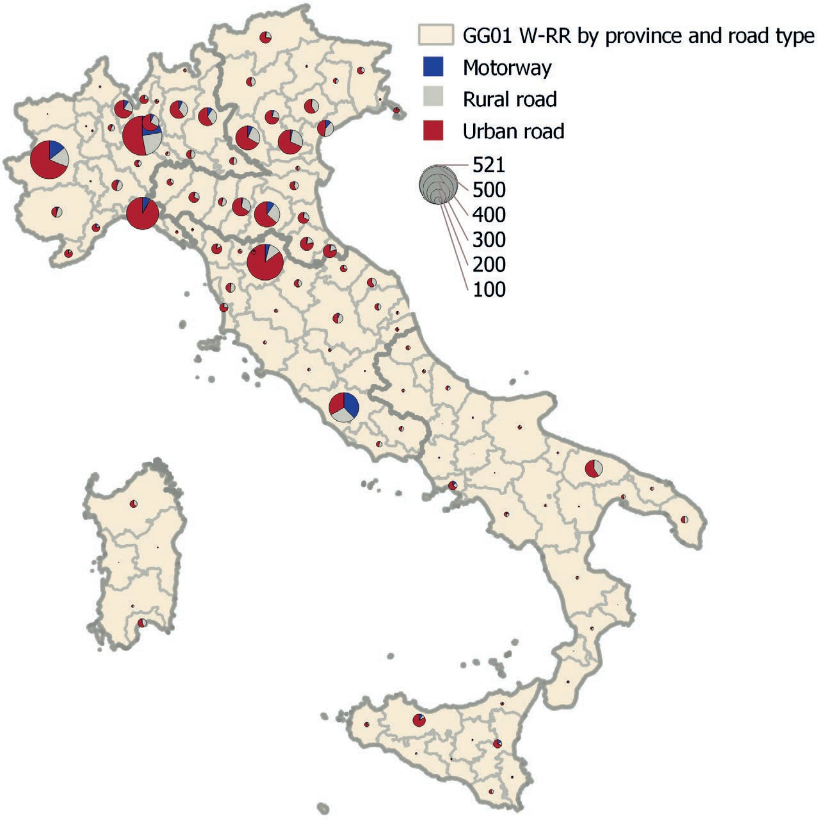
Macro-economic sector	Territorial division	Province (N)	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Min.	Max.
GG01 Sales	Centre	22	0,99	0,31	0,07	0,85	1,12	0,4	1,6
	Islands	14	0,72	0,27	0,07	0,57	0,88	0,3	1,2
	North East	22	1,09	0,39	0,08	0,92	1,26	0,6	1,9
	North West	25	1,11	0,59	0,12	0,87	1,35	0,2	2,5
	South	24	0,54	0,27	0,06	0,43	0,66	0,1	1,1
	Italy	107	0,90	0,45	0,04	0,82	0,99	0,1	2,5
GG03 Health social services	Centre	22	1,25	1,04	0,22	0,79	1,71	0,2	3,9
	Islands	14	0,75	0,70	0,19	0,35	1,15	0,2	2,4
	North East	22	0,89	0,43	0,09	0,70	1,08	0,5	2,1
	North West	25	0,80	0,28	0,06	0,68	0,92	0,3	1,5
	South	24	0,43	0,17	0,04	0,36	0,51	0,2	0,7
	Italy	107	0,82	0,64	0,06	0,70	0,95	0,2	3,9
GG3 Construction and plant	Centre	22	0,72	0,24	0,05	0,61	0,83	0,2	1,3
	Islands	14	0,39	0,12	0,03	0,32	0,46	0,2	0,6
	North East	22	0,76	0,22	0,05	0,66	0,86	0,3	1,2
	North West	25	0,66	0,24	0,05	0,56	0,76	0,4	1,3
	South	24	0,36	0,20	0,04	0,28	0,45	0.0	0,8
	Italy	107	0,59	0,27	0,03	0,54	0,64	0.0	1,3
GG6 Metals and machinery	Centre	22	0,73	0,27	0,06	0,61	0,85	0,4	1,6
	Islands	14	0,42	0,16	0,04	0,33	0,51	0,2	0,8
	North East	22	0,69	0,23	0,05	0,59	0,79	0,3	1,1
	North West	25	0,70	0,37	0,07	0,54	0,85	0,3	1,8
	South	24	0,44	0,27	0,05	0,33	0,56	0.0	1,2
	Italy	107	0,61	0,30	0,03	0,55	0,67	0.0	1,8
GG9 Transport and warehouses	Centre	22	2,14	0,97	0,21	1,71	2,58	0,3	4,4
	Islands	14	0,99	0,20	0,05	0,87	1,11	0,8	1,4
	North East	22	1,82	0,60	0,13	1,55	2,08	0,9	3,2
	North West	25	1,96	0,79	0,16	1,63	2,28	0,5	3,8
	South	24	1,09	0,44	0,09	0,91	1,28	0,3	2.0
	Italy	107	1,65	0,81	0,08	1,49	1,80	0,3	4,4

Source: Authors' processing on integrated Inail - Istat archive data. Italy 2014-2018. Inail archives of employed workers, BDS Inail, Industry and Services Group, Italy 2018. Updated to 31/10/2020 (see Footnotes 1, 2 and 3")

Spatial analysis of W-RR injuries by economic group, province and estimated number of the employed provided absolute and relative values (Figures 4.2 - 4.6).

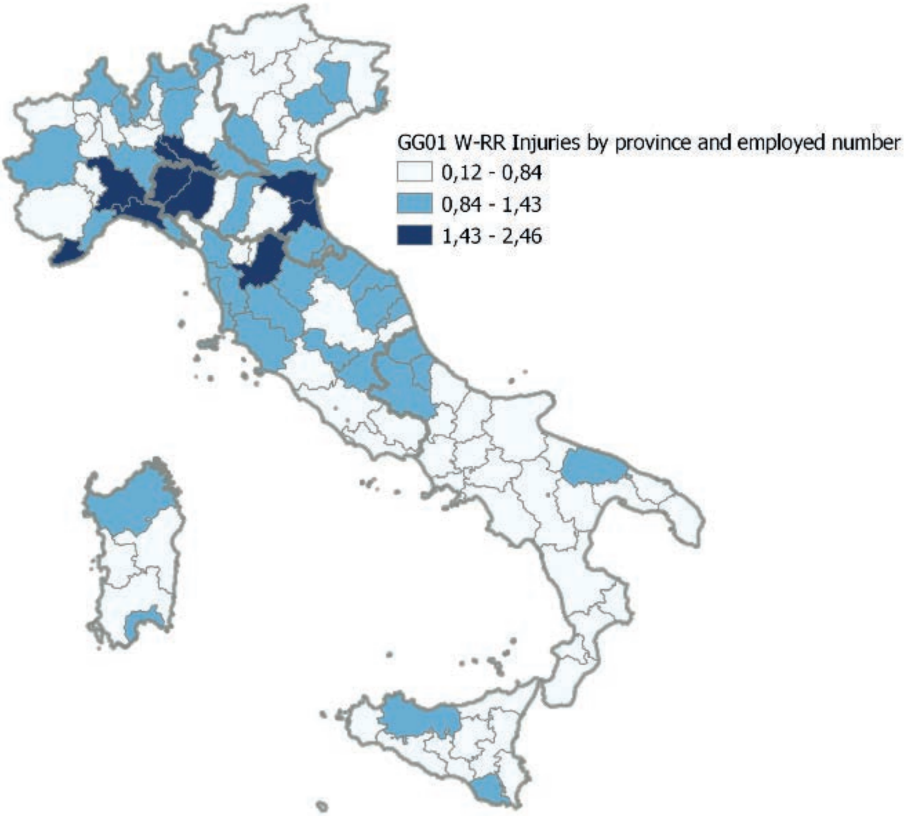
Figure 4.2 - GG01 Sales W-RR injuries. Table and map of absolute values by road type and provinces (bottom) and index map by employees (a) **and province (below). Italy. 2014-2018**

Macro economic sector	Motorway	Rural road	Urban road	Italy
1 GG01 Sales	879	3352	7195	11426



Source: Authors' processing on integrated Inail - Istat Archives (bottom and below). Inail archives of employed workers BDS Inail, Industry and Services Group, Italy 2018. Updated to 31/10/2020 (below). OpenStreetMap data. QGIS
(a) See Footnotes 1, 2 and 3.

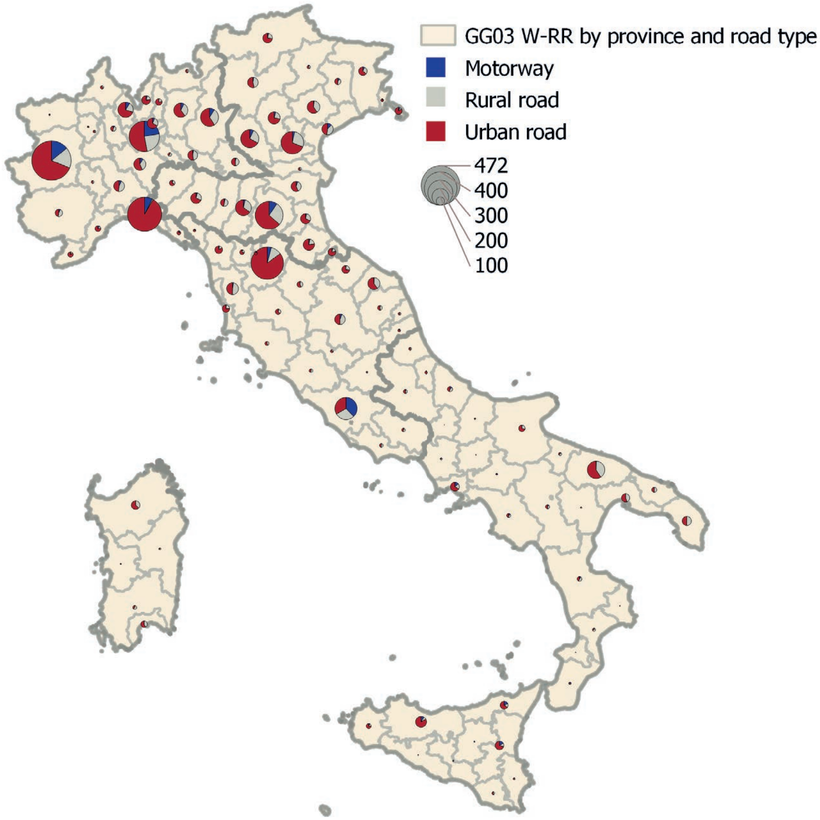
Figure 4.2 continued - **GG01 Sales W-RR injuries. Table and map of absolute values by road type and provinces (bottom) and index map by employees** (a) **and province (below). Italy. 2014-2018**



Source: Authors' processing on integrated Inail - Istat Archives (bottom and below). Inail archives of employed workers BDS Inail, Industry and Services Group, Italy 2018. Updated to 31/10/2020 (below). OpenStreetMap data. QGIS (Jenks)
(a) See Footnotes 1, 2 and 3.

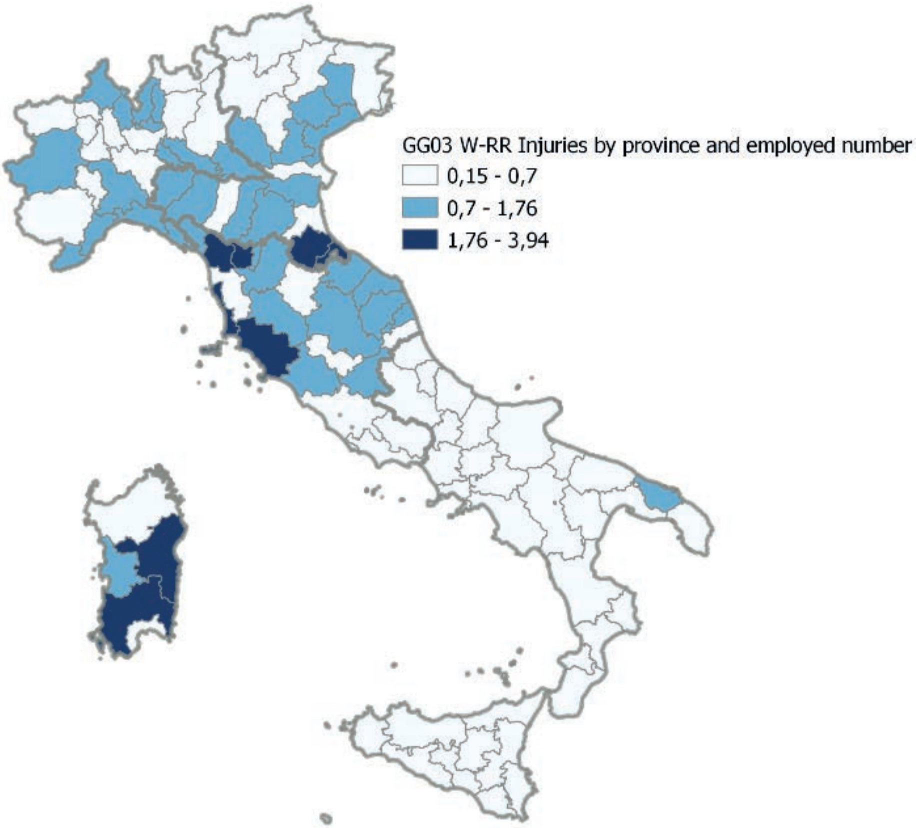
Figure 4.3 - GG03 Health and social services W-RR injuries. Table and map of absolute values by road type and provinces (bottom) and index map by employees and province (below). Italy. 2014-2018

Macro economic sector	Motorway	Rural road	Urban road	Italy
1 GG03 Health & social services	567	2808	6586	9961



Source: Authors' processing on integrated Inail - Istat Archives (bottom and below). Inail archives of employed workers BDS Inail, Industry and Services Group, Italy 2018. Updated to 31/10/2020 (below). OpenStreetMap data. QGIS

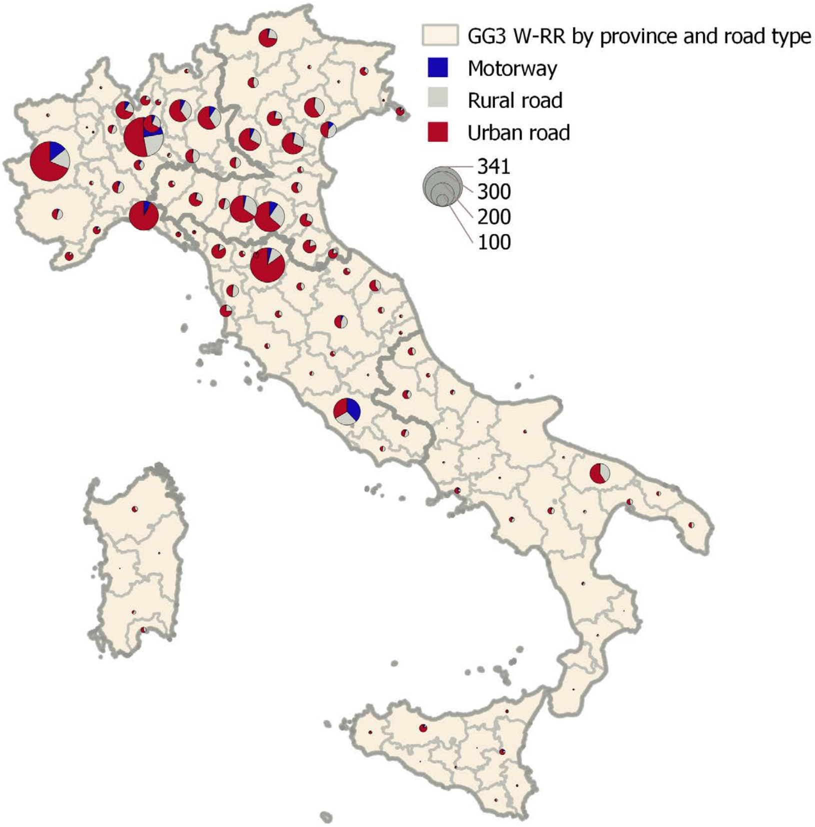
Figure 4.3 continued - **GG03 Health and social services W-RR injuries. Table and map of absolute values by road type and provinces (bottom) and index map by employees and province (below). Italy. 2014-2018**



Source: Authors' processing on integrated Inail - Istat Archives (bottom and below). Inail archives of employed workers BDS Inail, Industry and Services Group, Italy 2018. Updated to 31/10/2020 (below). OpenStreetMap data. QGIS (Jenks)

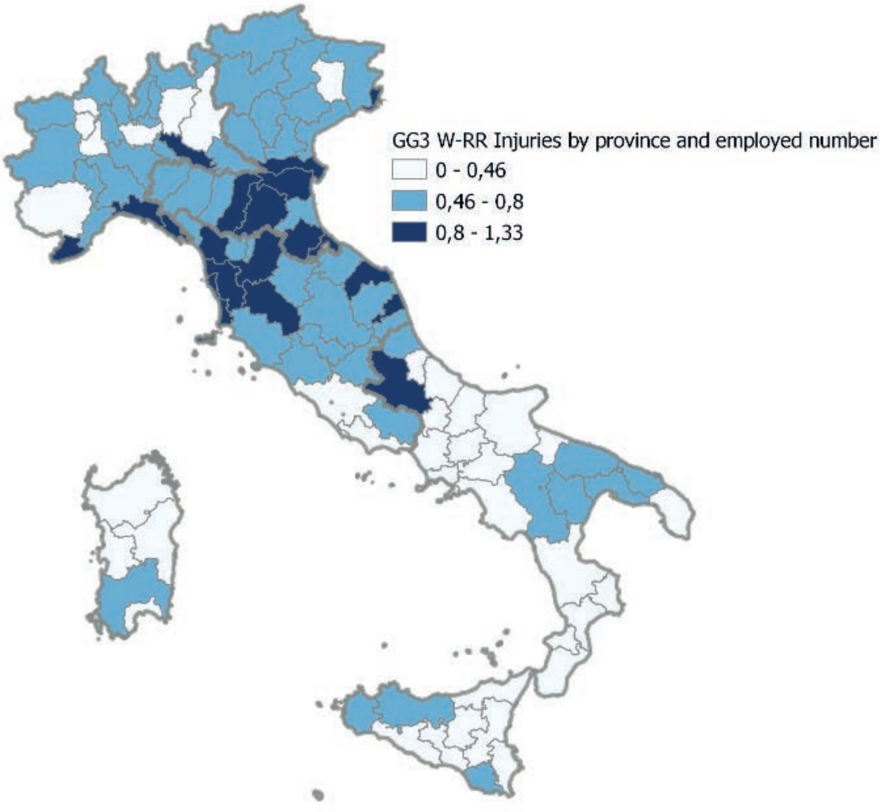
Figure 4.4 - GG3 Construction and plant W-RR injuries. Table and map of absolute values by road type and provinces (bottom) and index map by employees and province (below). Italy. 2014-2018

Macro economic sector	Motorway	Rural road	Urban road	Italy
1 GG3 Construction and plant	1072	2979	4161	8212



Source: Authors' processing on integrated Inail - Istat Archives (bottom and below). Inail archives of employed workers BDS Inail, Industry and Services Group, Italy 2018. Updated to 31/10/2020 (below). OpenStreetMap data. QGIS

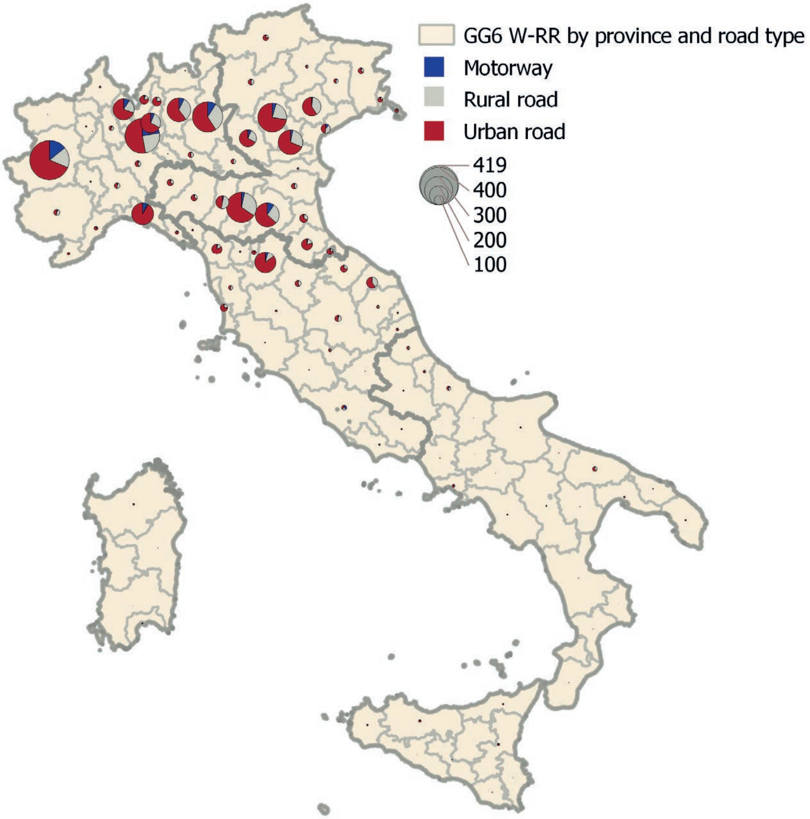
Figure 4.4 continued - **GG3 Construction and plant W-RR injuries. Table and map of absolute values by road type and provinces (bottom) and index map by employees and province (below). Italy. 2014-2018**



Source: Authors' processing on integrated Inail - Istat Archives (bottom and below). Inail archives of employed workers BDS Inail, Industry and Services Group, Italy 2018. Updated to 31/10/2020 (below). OpenStreetMap data. QGIS (Jenks)

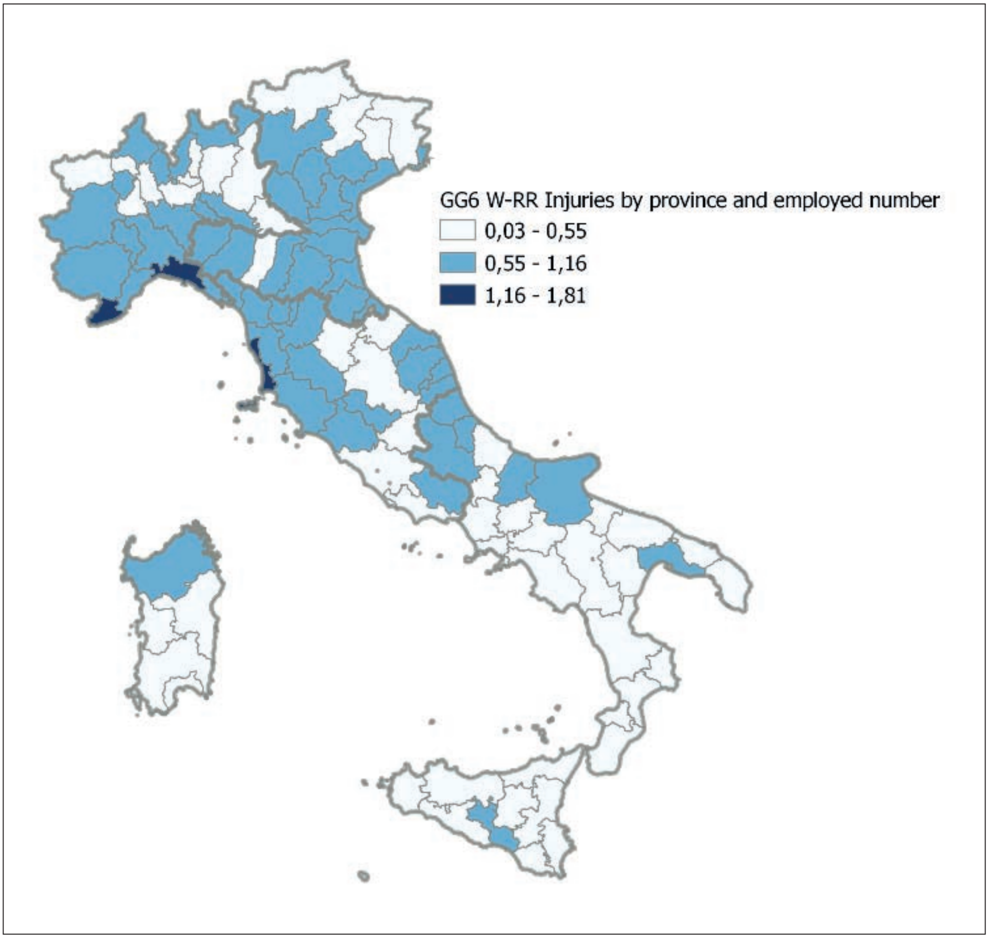
Figure 4.5 - GG6 Metals and machinery W-RR injuries. Table and map of absolute values by road type and provinces (bottom) and index map by employees and province (below). Italy. 2014-2018

Macro economic sector	Motorway	Rural road	Urban road	Italy
1 GG6 Metals and machinery	701	3579	7031	11311



Source: Authors' processing on integrated Inail - Istat Archives (bottom and below). Inail archives of employed workers BDS Inail, Industry and Services Group, Italy 2018. Updated to 31/10/2020 (below). OpenStreetMap data. QGIS

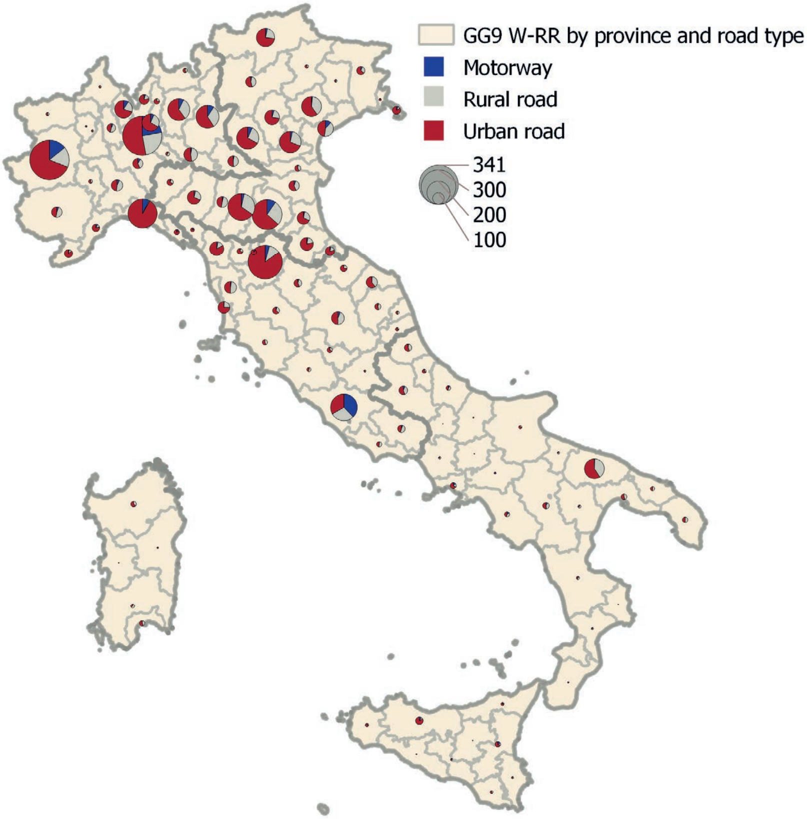
Figure 4.5 continued - **GG6 Metals and machinery W-RR injuries. Table and map of absolute values by road type and provinces (bottom) and index map by employees and province (below). Italy. 2014-2018**



Source: Authors' processing on integrated Inail - Istat Archives (bottom and below). Inail archives of employed workers BDS Inail, Industry and Services Group, Italy 2018. Updated to 31/10/2020 (below). OpenStreetMap data. QGIS (Jenks)

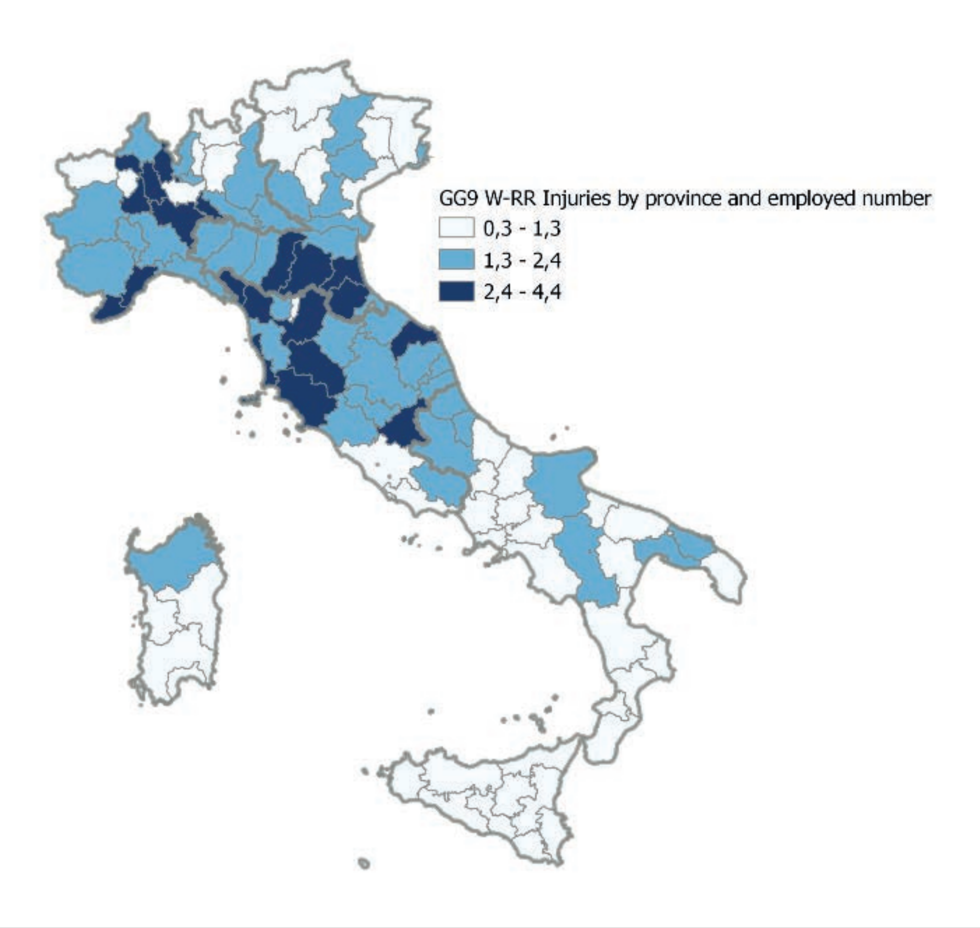
Figure 4.6 - GG9 Transport and warehouses W-RR injuries. Table and map of absolute values by road type and provinces (bottom) and index by employees and province (below). Italy. 2014-2018

Macro economic sector	Motorway	Rural road	Urban road	Italy
1 GG9 Transport and warehouses	2089	3542	5088	10719



Source: Authors' processing on integrated Inail - Istat Archives (bottom and below). Inail archives of employed workers BDS Inail, Industry and Services Group, Italy 2018. Updated to 31/10/2020 (below). OpenStreetMap data. QGIS

Figure 4.6 continued - **GG9 Transport and warehouses W-RR injuries. Table and map of absolute values by road type and provinces (bottom) and index by employees and province (below). Italy. 2014-2018**



Source: Authors' processing on integrated Inail - Istat Archives (bottom and below). Inail archives of employed workers BDS Inail, Industry and Services Group, Italy 2018. Updated to 31/10/2020 (below). OpenStreetMap data. QGIS (Jenks)

As reported in Tables 4.1, 4.3 and 4.4, the GG01 sales sector recorded 11,426 injuries and maximum indexes (1.43-2.46) distributed mainly in the North East, North West and Centre (Figure 4.2); the GG6 metals and machinery sector recorded 11,311 injuries and maximum indexes (1.16-1.81), distributed mainly in the North West and Centre along the Tyrrhenian coast (Figure 4.5), while the GG9 transport and warehouse sector recorded 10,719 injuries and maximum indexes (2.40-4.43) distributed mainly in the North West, North East and Centre (Figure 4.6). This was followed by the GG03 health & social services sector that recorded 9,961 injuries and maximum indexes (1.76-3.94), distributed mainly in the North East in the province of Ravenna, in the Centre and in Sardinia (Figure 4.3). On account of its widespread territorial presence, the GG3 construction and plant sector, with 8,212 injuries and maximum indexes (0.80-1.33) distributed mainly in the North East, Centre and to a lesser degree in the North West and South (Figure 4.4) was also included in our observations.

4.2 ANOVA Analysis of variance

ANOVA was performed for two associations: "injuries by road type and territorial division" and "injuries by macro-economic sector (GG) and territorial division". For the former, the p-value (Sig.) was > 0.05 (confidence value limit) for motorways and < 0.05 for rural and urban roads (Table 4.5).

Table 4.5 - ANOVA. W-RR injuries by road type and territorial division (absolute values). Italy. 2014-2018

Road type	Groups or Territorial divisions	Sum of Squares	df	Mean Square	F	Sig.
Motorway	Between Groups	255,433.47	4	63,858.37	1.39	0.242
	Within Groups	4,679,166.76	102	45,874.18		
	Italy	4,934,600.22	106			
Rural road	Between Groups	1,992,042.57	4	498,010.64	7.13	0.000
	Within Groups	7,119,528.09	102	69,799.29		
	Italy	9,111,570.65	106			
Urban road	Between Groups	15,573,684.56	4	3,893,421.14	6.21	0.000
	Within Groups	63,944,958.59	102	626,911.36		
	Italy	79,518,643.16	106			

Source: Authors' processing on integrated Inail - Istat data. Italy 2014-2018

We failed to reject the null hypothesis for motorways, while it was rejected for rural and urban roads. Variance for rural and urban roads (p-value < 0.05) appears to differ in territorial divisions and vice versa to be homogeneous for motorways (Tables 4.5 and B1 in Appendix B).

Table 4.6 - ANOVA. W-RR injuries by 5 macro-economic sectors and territorial division (absolute values). Italy. 2014-2018

Macro-economic sector	Groups or Territorial divisions	Sum of Squares	df	Mean Square	F	Sig.
GG01 Sales	Between Groups	207,210.0	4	51,802.51	5.14	0.001
	Within Groups	1,027,764.0	102	10,076.12		
	Italy	1,234,974.0	106			
GG03 Health social services	Between Groups	118,485.8	4	29,621.45	4.38	0.003
	Within Groups	690,317.3	102	6,767.82		
	Italy	808,803.1	106			
GG3 Construction and plant	Between Groups	123,632.0	4	30,908.00	7.46	0.000
	Within Groups	422,448.2	102	4,141.65.00		
	Italy	546,080.2	106			
GG6 Metals and machinery	Between Groups	506,880.0	4	126,720.00	9.72	0.000
	Within Groups	1,329,848.0	102	13,037.73		
	Italy	1,836,728.0	106			
GG9 Transport and warehouses	Between Groups	219,388.8	4	54,847.19	4.72	0.002
	Within Groups	1,184,005.0	102	11,607.89		
	Italy	1,403,394.0	106			

Source: Authors' processing on integrated Inail - Istat data. Italy 2014-2018

For the five macro-economic sectors, the Sig. value resulted less than 0.05 and the null-hypothesis (homogeneity of variances between territorial divisions) was rejected (Table 4.6).

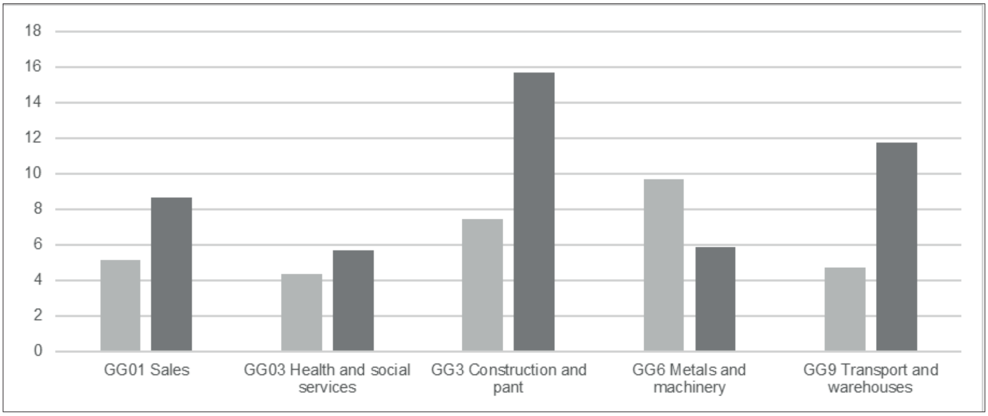
ANOVA performed on indexed frequencies provided overall values of significance < 0.05 (Table 4.7). Consequently, despite the values of significance >0.05 for GG3 construction and plant and GG6 metals and machinery sectors obtained in Table B2 in Appendix B, H_0 was rejected for the five macro-economic sectors analysed. Figure 4.7 illustrates the absolute and relative F values for the 5 macro-economic sectors. As regards absolute frequencies, maximum F values were found for GG6 metals and machinery and minimum values for GG03 health and social services. With regard to indexed frequencies, maximum F values were found for GG3 construction and plant and minimum values for GG03 health and social services, GG6 metals and machinery and GG9 transport and warehouses. The F values for indexed frequencies were higher than F for absolute values. The only exception was the GG6 metals and machinery macro sector.

Table 4.7 - ANOVA. W-RR injuries. Macro-economic sector and territorial division (relative values), Italy, 2014-2018

Macro-economic sector	Groups or Territorial divisions	Sum of Squares	df	Mean Square	F	Sig.
GG01 Sales	Between Groups	5.50	4	1.38	8.66	0.000
	Within Groups	16.20	102	0.16		
	Italy	21.71	106			
GG03 Health social services	Between Groups	7.87	4	1.97	5.68	0.000
	Within Groups	35.38	102	0.35		
	Italy	43.25	106			
GG3 Construction and plant	Between Groups	2.92	4	0.73	15.70	0.000
	Within Groups	4.74	102	0.05		
	Italy	7.66	106			
GG6 Metals and machinery	Between Groups	1.81	4	0.45	5.89	0.000
	Within Groups	7.83	102	0.08		
	Italy	9.64	106			
GG9 Transport and warehouses	Between Groups	21.81	4	5.45	11.74	0.000
	Within Groups	47.37	102	0.46		
	Italy	69.18	106			

Source: Authors' processing on integrated Inail - Istat data. Italy 2014-2018. Inail Archives of Employed Workers, BDS Inail. Industry and Services Group. Italy 2018. Updated to 31/10/2020 (see Footnotes 1, 2 and 3)

Figure 4.7 - ANOVA. W-RR injuries by macro-economic sector. Variance ratio between and within divisions (F). Absolute (grey) and relative values (dark grey). Italy. 2014-2018



Source: Authors' processing on integrated Inail - Istat data. Italy 2014-2018. Inail Archives of Employed Workers, BDS Inail. Industry and Services Group. Italy 2018. Updated to 31/10/2020 (see Footnotes 1, 2 and 3)

5. Discussion

Road accident rates are influenced by many factors such as traffic, road junctions, driving speed and, last but not least, national driving habits and the technical standard of the vehicles used. There is still a lack of knowledge regarding the interaction of the various factors influencing W-RR accident rates (PIARC 2008). Information about accidents and their victims is essential for studying road safety (Hollò *et al.*, 2010), and data concerning W-RR accidents are extremely important since work-related journeys account for a significant part of urban and rural traffic.

Study findings are not consistent between and within Italian territorial divisions for both absolute and relative W-RR injury rates. According to Partyka, simple models of traffic fatalities have been developed using only readily available factors concerning population size and the size of the potential labour force. However, the reader should use caution in interpreting modelling analysis, because the model produces estimates of the effects of variables included in the analysis, but does not imply estimates of the effects of variables omitted in the analysis (Partyka, 1984). According to Hollò, also indicators are unable to provide a full understanding of road safety trends, and, if they are applied generally without the required background information, this may even lead to serious misinterpretation of trends in road casualties (Hollò *et al.*, 2010).

Despite these limitations, the integrated archives on road collisions and injuries, comprehensive of location and economic classification, enabled us to investigate W-RR injuries, nationwide. The descriptive and spatial analysis of the integrated Inail - Istat dataset revealed that the majority of W-RR accidents occurred in the North East, the North West and the Centre (the metropolitan areas of Rome and Florence).

The descriptive analysis of the absolute frequencies of W-RR injuries by road type, macro-economic sector and territorial division (Table 4.1 and Figure 4.1) shows that the urban and rural roads of the North and the Centre are mainly involved in relation to the GG01 sales sector, the GG03 health and social services sector, and the GG6 metals and machinery sector; while in the GG3 construction and plant sector and in the GG9 transport and warehouse sector, both urban and rural roads, and motorways with injury peaks are mainly concentrated in the North and Centre of the country.

Absolute W-RR injury rates by macro-economic sector were highest in the GG01 sales sector, followed by the GG6 metals and machinery sector, the GG9 transport and warehouses sector, the GG03 health and social services sector, the GG3 construction and plant sector, whereas indicators of traffic injuries normalised by the estimated number of workers, were respectively highest in the GG9 transport and warehouses sector, the GG03 health social services sector, the GG01 sales sector, the GG6 metals and machinery sector and the GG3 construction and plant sector (see Tables 4.3, 4.4 and Figures 4.2 - 4.6).

As regards motorways, overall the highest mean was recorded in the North West, while the peak was observed in the Centre (1,556). This last observation confirms the results of a previous study (Brusco *et al.*, 2019) in which the main Rome motorway (the *Grande Raccordo Anulare* - the Ring Road) alone reached the highest accident and injury frequency rate in Italy. On rural roads, the highest peaks and means were recorded in the North of the country, especially in the North East. On urban roads, the means were higher in the North, while the injury peaks (both over 4,200) were in the North West and the Centre (Table 4.2).

Therefore, analysis of macro-economic sector revealed some differences between the absolute (Table 4.3) and indexed rates (Table 4.4). In the GG03 health and social services sector, the indexed values were higher in the Centre and the Islands, while the absolute values were concentrated in the North West. In the GG6 metal and machinery sector, indexed frequencies were highest in the North West and in the Centre, while absolute values were highest in the North. Finally, in the GG9 transport and warehouse sector, the indexed peaks were observed in the Centre and vice versa, the absolute ones were recorded in the North West. A comparison of absolute and indexed data allowed the issue of accidents to be reassessed from a different point of view that took into account also the estimated worker number of each sector (see Table 4.4 and Figures 4.2-4.6).

Indexation using the ratio of two related variables (accident frequencies and the number of employees in each sector) provided a different viewpoint on traffic accidents, useful for expressing a dynamic assessment of how many W-RR injuries “derive” from the total number of workers employed by sector. This information could be useful for analysing W-RR risks by macro-economic sector. The spatial analysis, performed on absolute and normalised injury rates, did not enable us to identify a common thread in frequency and thereby to reach universally interpretable results for the whole territory. ANOVA pointed out information not available by only the spatial and descriptive analysis.

ANOVA, performed on road type, macro-economic group and territorial division (Tables 4.5 - 4.7 and Appendices) demonstrated that homogeneity of variances between territorial divisions was rejected (except for motorways) and the differences in divisions, hypothesised for the 2014-2018 integrated dataset, were confirmed.

Variances "between" were higher than variances "within" divisions for rural and urban roads, but vice versa for motorways (Table 4.5). A possible explanation of this finding could be the limited number of motorways and the low traffic volume recorded on the same.

As regards the association between macro-economic groups and territorial divisions, the homogeneity of variances between the latter was rejected both for absolute and relative injury rates (Tables 4.6 and 4.7). The highest heterogeneity concerned the macro-economic GG3 construction and plant sector, while the lowest was in the GG03 health and social services sector (absolute and relative). In general, the coefficient F for absolute frequencies was greater than F indexed except in the metals and machinery sector (Figure 4.7). This exception, as well as the results obtained in Table B2 in Appendix B, could be linked to distortion in attributing location as described in Footnotes 1, 2 and 3.

Our study puts forward two sets of hypothetical explanations for the heterogeneity of variances in all the divisions. The first concerns the different structural changes involving organisational, technological and occupational aspects effected in each one that influenced the number of estimated workers (2,537,828 in the North West, 1,854,435 in the North East, 1,432,023 in the Centre, 1,127,507 in the South and 485,740 in the Islands (see Footnote 1 for the 5 economic sector investigated). According to Eksler, structural changes in society and the economy are the most important factor behind trends in road fatalities (Eksler, 2009).

A further explanation concerns the quality of the Inail and Istat dataset, in terms of availability, reliability and completeness of data recorded in each division and included in the integrated Inail - Istat archive data.

A recent survey focussed on the percentage of workers who, in the North West of Italy in the five-year period 2014-2018, used the train or other urban means of transport to get to work. The percentage was much higher than in the remaining divisions (especially those of the South and the Islands). In the North West, rail freight traffic was also higher in the same period than elsewhere (Istat, 2021*b*). These last aspects that diverge from the results of our study may be the cause of differentiation in traffic exposure and accident risk.

A recent study analysed the work-related accident trend in transport and highlighted its critical and peculiar aspects but failed to make any reference to the accident sites or to the indexation by the estimated number of employees (Inail, 2019). Another study referring to the ATECO classification (Istat), analysed indexed W-RR injuries that occurred during the five-year period 2014-2018 in the sales, transport, manufacturing and construction sectors (Pireddu *et al.*, 2021), confirming the highest absolute frequencies in the North and Centre and on urban roads and in the metropolitan area of Rome where, for most of the ATECO sectors, accidents were evenly distributed between motorways, extra-urban and urban roads. According to the ATECO classification, the sales, manufacturing and transport sectors recorded higher indexed injury rates. These differences could be traced back to a different redistribution of accidents and employees in the two classifications (macro-economic sector and ATECO). Variables such as traffic, marginally evaluated by the classification based on the road type (Pireddu and Bruzzone, 2021), driving circumstances (European Commission CARE. 2016) and different driving habits throughout Italy were omitted in our analysis, as well as vehicle devices and technical standards. Despite these limitations, to be further analysed, our study points out essential findings and methods for all those who, with a reactive or proactive approach, are required to analyse road safety and risks in a specific context (Legislative Decree 9 April 2008 n. 81 - *Decreto Legislativo* n. 81/2008 - and Directive 2019/1936/ EU), tariff group and territorial division.

6. Conclusions

The nationwide study investigated injury rates by economic sector, road type and territorial division. The integrated Inail - Istat dataset that included 128,837 accidents resulting in death or injury on Italian roads during the 2014-2018 five-year period, enabled us to analyse occupational and accidental parameters that were not available in the two data archives taken separately.

Absolute and relative W-RR injury frequencies were analysed. ANOVA findings applied to road type showed that injury variances between Italian divisions on rural and urban roads (excluding motorways) were heterogeneous. Variances were also heterogeneous for the macro-economic sector and territorial division (absolute and relative values).

Our study provides essential findings and useful methods for analysing mobility and planning intervention to prevent road accidents with a reactive or proactive approach. The study is also intended for local public authorities responsible for assessing and managing risks in specific contexts.

Appendix A.
W-RR injuries by large tariff group (macro-economic sector) and road type. Italy. 2014-2018

Table A1 - Injury frequencies by macro-economic sector, territorial division and road type (absolute values). All the macro-economic sectors. Italy. 2014-2018

Macro-economic sector and road type		Motorway	Rural road	Urban road
Total	128,837	10,639	37,866	80,332
1 - Industry and services				
1-Service personnel		18	192	1,287
1-GG01 - Sales		879	3,352	7,195
1-GG02 - Reception & Catering		291	2,298	6,577
1-GG03 - Health & social services		567	2,808	6,586
1-GG04 - Cleaning, sanitation and disinfestation		250	1,164	3,532
1-GG05 - Cinematography, culture and sport		102	209	515
1-GG06 - Educational, research and survey		98	427	1,064
1-GG07 - Various undefined activities		3,006	7,207	1,8964
1-GG08 - Undefined activities		191	565	1,186
1-GG1 - Agriculture and food processing		128	1,064	1,630
1-GG2 - Chemical, paper, leather		142	1,049	1,675
1-GG3 - Construction and plant		1,072	2,979	4,161
1-GG4 - Energy and communications		51	221	359
1-GG5 - Wood and the like		47	361	630
1-GG6 - Metals and machinery		701	3,579	7,031
1-GG7 - Mining, minerals and glass		35	283	441
1-GG8 - Textiles and packaging		79	722	1,578
1-GG9 - Transport and warehouses		2,089	3,542	5,088
1-Not classified		465	2,494	4,691
2 - Agriculture				
2 - Agriculture and food processing		101	1,592	1,139
3 - Public administration		303	1,540	3,817
4 - Students		24	218	1,186

Source: Authors' processing on integrated Inail - Istat data. Italy 2014-2018

Appendix B.
ANOVA. W-RR injuries by macro-economic sector and road type. Italy. 2014-2018

Table B1 - Test of Homogeneity of Variances. Injuries by road type (absolute values). Italy. 2014-2018

Road type	Levene statistic	df1	df2	Sig.
Motorway injuries	2.54	4	102	0.044
Rural road injuries	5.08	4	102	0.001
Urban road injuries	6.60	4	102	0.000

Source: Authors' processing on integrated Inail - Istat data. Italy 2014-2018

Table B2 - Test of Homogeneity of Variances. Injuries by macro-economic sector (relative values). Italy. 2014-2018

Macro-economic sector	Levene statistic	df1	df2	Sig.
GG01 Sales	4,37	4	102	0,003
GG03 Health social services	9,12	4	102	0.000
GG3 Construction and plant	1.00	4	102	0,412
GG6 Metals and machinery	0,93	4	102	0,450
GG9 Transport and warehouses	6,30	4	102	0.000

Source: Authors' processing on integrated Inail - Istat data. Italy 2014-2018. Inail Archives of Employed Workers, BDS Inail. Industry and Services Group. Italy 2018. Updated to 31/10/2020 (see Footnotes 1, 2 and 3)

References

Brusco, A., A. Bucciarelli, M. Bugani, C. Gariazzo, C. Giliberti, M. Marinaccio, S. Massari, A. Pireddu, L. Veronico, G. Baldassarre, S. Bruzzone, M. Scortichini, M. Stafoggia, e S. Salerno. 2019. *Gli incidenti con mezzo di trasporto. Un'analisi integrata dei determinanti e dei fattori di rischio occupazionali*. Roma, Italy: Inail.

Bruzzone, S., A. Altimari, G. Baldassarre, R. Boscioni, L. Veronico. 2021. "Work-related road accidents: an in depth-statistical analysis carried out by two different integrated data sources". In this issue of the *Rivista di statistica ufficiale*, N. 3/2021. Roma, Italy: Istat.

Cima, V., M. Carroccio, e R. Maseroli. 2014. "Corretto utilizzo dei Sistemi Geodetici di Riferimento all'interno dei software GIS". In *Atti della 18a Conferenza Nazionale ASITA*, 14-16 ottobre 2014, Firenze: 359-363. Milano, Italia: Federazione delle Associazioni Scientifiche per le Informazioni Territoriali e Ambientali - ASITA.

Costabile, S., S. Martini, L. Petriglia, G. Corrarello, e A. Avanzi. 2012. "Geoportale Nazionale. Un approfondimento sulle metodologie di conversione e trasformazione coordinate". *GEOMEDIA*, Volume 16, N. 6: 26-28.

Decreto Legislativo 9 aprile 2008, n. 81 (*Legislative Decree n. 81/2008*). "Attuazione dell'articolo 1 della legge 3 agosto 2007, n. 123, in materia di tutela della salute e della sicurezza nei luoghi di lavoro". Gazzetta Ufficiale della Repubblica Italiana, Serie Generale, n. 101 del 30 aprile 2008, *Supplemento Ordinario* n. 108.

Decreto Legislativo 30 aprile 1992, n. 285 (*Legislative Decree n. 285/1992*). "Testo aggiornato recante il nuovo codice della strada". Gazzetta Ufficiale della Repubblica Italiana, *Serie Generale*, n. 67 del 22 marzo 1994, *Supplemento Ordinario* n. 49.

Directive (EU) 2019/1936 of the European Parliament and of the Council of 23 October 2019 amending Directive 2008/96/EC on *road infrastructure safety management*.

Eksler, V. 2009. "Road Mortality in Europe: A Regional Approach". *Ph.D. Thesis*. Versailles, France: Versailles St-Quentin en Yvelines University.

European Commission, Community Database on Road Accidents - CARE. 2016. "Road Safety: new statistics call for fresh efforts to save lives on EU roads", *Press Release* 31 March 2016. Brussels, Belgium: European Commission.

Gariazzo, C., S. Bruzzone, S. Finardi, M. Scortichini, L. Veronico, e A. Marinaccio. 2021. "Association between extreme ambient temperatures and general indistinct and work-related road crashes. A nationwide study in Italy". *Accident Analysis & Prevention*, Volume 155, 106110.

Glass, G.V., P.D. Peckham, and J.R. Sanders. 1972. "Consequences of non-compliance with the assumptions underlying the analysis of the fixed effects of variance and covariance". *Review of Educational Research*, Volume 42, Issue 3: 237-288.

Harwell, MR., E.N. Rubinstein, W.S. Hayes, and C.C. Olds. 1992. "Summarizing Monte Carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases". *Journal of Educational Statistics*, Volume 17, Issue 4: 315–339.

Hollò, P., V. Eksler, and J. Zukowska. 2010. "Road safety performance indicators and their explanatory value: A critical view based on the experience of Central European countries". *Safety Science*, Volume 48, Issue 9: 1142-1150.

Istituto Nazionale di Statistica - Istat (*Italian National Institute of Statistics - Istat*). 2021*a*.

"Rilevazione degli incidenti stradali con lesioni a persone". *Informazioni sulla rilevazione*. Roma, Italy: Istat. https://www.istat.it/it/archivio/4609.

Istituto Nazionale di Statistica - Istat (Italian National Institute of Statistics - Istat). 2021*b*. *Indicatori territoriali per le politiche di sviluppo*. Roma, Italy: Istat. https://www.istat.it/it/archivio/16777.

Istituto Nazionale di Statistica - Istat (*Italian National Institute of Statistics - Istat*). 2018. *Annuario Statistico Italiano 2018*. Roma, Italy: Istat. https://www.istat.it/it/archivio/225274.

Istituto Nazionale di Statistica - Istat (*Italian National Institute of Statistics - Istat*). *I.Stat, Your direct access to the Italian Statistics: the complete data warehouse for experts*. Roma, Italy: Istat. http://dati.istat.it/Index.aspx.

Istituto Nazionale di Statistica - Istat (*Italian National Institute of Statistics - Istat*). "Survey on Road accidents resulting in death or injury". In *SIQual - Information system on quality of statistical production processes*. Roma, Italy: Istat. http://siqual.istat.it/SIQual/lang. do?language=UK.

Istituto Nazionale di Statistica - Istat (*Italian National Institute of Statistics - Istat*), e Automobile Club d'Italia - ACI. 2019. "Incidenti stradali in Italia. Anno 2018". Comunicato Stampa. Roma, Italia: Istat. https://www.istat.it/it/archivio/232366.

Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro - Inail (*Italian National Institute for Insurance against Accidents at Work - Inail*). 2019. "Andamento degli infortuni sul lavoro e delle malattie professionali". *Dati Inail*, N. 7/2019. Roma, Italy: Inail. https://www.inail.it/cs/internet/comunicazione/news-ed-eventi/news/news-dati-inail-trasporti-2019.html.

Lix, L.M., J.C. Keselman, and H.J. Keselman. 1996. "Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance F Test". *Review of Educational Research*, Volume 66, Issue 4: 579-619.

Partyka, S.C. 1984. "Simple models of fatality trends using employment and population data". *Accident Analysis & Prevention*, Volume 16, Issue 3: 211-222.

Pireddu, A., A. Altimari, G. Baldassarre, M. Bonafede, R. Boscioni, S. Bruzzone, C. Gariazzo, A. Marinaccio, S. Massari, L. Taiano, T. Tuoto, L. Valentino e L. Veronico. 2021. "Gli infortuni nella classificazione ATECO. Analisi territoriale degli infortuni stradali lavoro-correlati. Potenzialità degli archivi integrati". Poster presentato alla *Quattordicesima Conferenza Nazionale di Statistica "Un moderno ecosistema dei dati per una ripartenza inclusiva e sostenibile"*. 31 Novembre - 1 Dicembre 2021. Roma, Italy: Istat. https://www.istat.it/it/archivio/261537.

Pireddu, A., and S. Bruzzone. 2021. "An analysis of the influence of tunnel length and road type on road accident variables". *Rivista di statistica ufficiale*, N. 2/2021: 71-102. Roma, Italy: Istat. https://www.istat.it/it/archivio/262320.

Taiano, L., S. Massari, T. Tuoto, L. Valentino, S. Bruzzone, and L. Veronico. 2021. "Work-related road accidents: a data linkage procedure applied to assess traffic accidents at work and commuting". In this issue of the *Rivista di statistica ufficiale*, N. 3/2021. Roma, Italy: Istat.

The *Rivista di statistica ufficiale* publishes peer-reviewed articles dealing with cross-cutting topics: the measurement and understanding of social, demographic, economic, territorial and environmental subjects; the development of information systems and indicators for decision support; the methodological, technological and institutional issues related to the production process of statistical information, relevant to achieve official statistics purposes.

The *Rivista di statistica ufficiale* aims at promoting synergies and exchanges between and among researchers, stakeholders, policy-makers and other users who refer to official and public statistics at different levels, in order to improve data quality and enhance trust.

The *Rivista di statistica ufficiale* was born in 1992 as a series of monographs titled "*Quaderni di Ricerca Istat*". In 1999 the series was entrusted to an external publisher, changed its name in "*Quaderni di Ricerca - Rivista di Statistica Ufficiale*" and started being published on a four-monthly basis. The current name was assumed from the Issue N. 1/2006, when the Italian National Institute of Statistics – Istat returned to be its publisher.

La Rivista di statistica ufficiale pubblica articoli, valutati da esperti, che trattano argomenti trasversali: la misurazione e la comprensione di temi sociali, demografici, economici, territoriali e ambientali; lo sviluppo di sistemi informativi e di indicatori per il supporto alle decisioni; le questioni metodologiche, tecnologiche e istituzionali relative al processo di produzione dell'informazione statistica, rilevanti per raggiungere gli obiettivi della statistica ufficiale.

La Rivista di statistica ufficiale promuove sinergie e scambi tra ricercatori, stakeholder, policy-maker e altri utenti che fanno riferimento alla statistica ufficiale e pubblica a diversi livelli, al fine di migliorare la qualità dei dati e aumentare la fiducia.

La Rivista di statistica ufficiale nasce nel 1992 come serie di monografie dal titolo "Quaderni di Ricerca Istat". Nel 1999 la collana viene affidata a un editore esterno, cambia nome in "Quaderni di Ricerca - Rivista di Statistica Ufficiale" e diventa quadrimestrale. Il nome attuale è stato scelto a partire dal numero 1/2006, quando l'Istituto Nazionale di Statistica - Istat è tornato a esserne l'editore.

ISSN 1828-1982 € 15,00