DGITM/DMR/TUD-VA
Editor: Elsa LANAUD

July 13, 2022

Safety validation of automated road transport systems: clarification through the analysis of accident data

Study report

Preamble

The safety validation of automated vehicles and transport systems is based on a certain number of methodological bricks (or "frameworks" within the meaning of the national strategy for the development of automated road mobility in France (Ministère de la Transition Ecologique, Direction Générale des Infrastructures, des Transports et de la Mer, 2020)). Among these building blocks, the establishment of one or more reference safety levels appears to be central. To do this, the analysis of accidents on conventional vehicles provides useful, even unavoidable insight, in the absence of feedbacks on automated systems.

This report presents the work carried out during 2021 on accident data in order to feed the work relating to the safety validation of automated road transport systems (ARTS). This work has been shared and discussed with the French ecosystem. It also, on a preliminary basis, contributed to European work as part of the preparatory work for the implementing regulation relating to the type-approval of fully automated vehicles (known as the "shuttle" regulation).

Table of Contents

Key points of the report – assumptions and main results

Working assumptions:

- The reference accident rate of conventional vehicles can be transposed to vehicles equipped with an automated driving system.
- Study area: mainland France (continental territory plus Corsica).
- 2 use cases:
  - driving on motorways for individual mobility (and national roads);
  - urban public transport.
- Study period: 7 or 8 years depending on the use case.
- The results obtained are inseparable from the methodology, the data and their precision.
- The accident rates presented are representative accident rates per kilometer travelled on the network considered and by the type of vehicle considered.

Main results

Table 1: Presentation of the results by use case of the calculated accident rates.

Average accident rate by macro use case	**Mortality rate**	Injury rate	**Fatal accident rate**	Non-fatal accident rate
Motorways and lanes with motorway characteristics (FR) *2012-2018* *All types of vehicles*	**$1.15.10^{-9}$**	$2.64.10^{-8}$	**$1.03.10^{-9}$**	$1.90.10^{-8}$
National roads *2012-2018* *All types of vehicles*	**$5.71.10^{-9}$**	$6.79.10^{-8}$	**$5.07.10^{-9}$**	$4.66.10^{-8}$
Urban public transport (UPT) - Ile-de-France excluded *2012-2019* *Public transport (buses, coaches)*	**$3.96.10^{-8}$**	$1.49.10^{-6}$	**$3.64.10^{-8}$**	$8.31.10^{-7}$

Table 2: Presentation of the results of the average mortality and fatal accident rates for the 27+3 territories considered over the 2012-2019 period for the use case of urban public transport.

Urban public transport (buses, coaches) *Period 2012-2019* *21 main cities ("métropoles")* *6 urban territories ("agglomérations")* *Ile-de-France*	**Mortality**	**Fatal accidents**
Average rates	**$2.92.10^{-8}$**	**$2.81.10^{-8}$**

Introduction

This report presents the methodology used to analyse accident data from the *Observatoire National Interministériel de la Sécurité Routière (ONISR)* database. It provides the results obtained as of January 31, 2022, as part of the work carried out on the safety validation of automated road transport systems (ARTS). Its aim is to integrate and supply the safety validation framework by objectivizing the reference safety levels in terms of accident rates.

More generally, this work presents a French contribution to the methodology for establishing reference safety levels for the validation of automated systems in the context of European works (regulation known as “ADS”). Its purpose is to provide initial reference accident values for systems, whether in the context of vehicle type approval at European level, but also in the context of the commissioning of automated road transport systems at a national level.

The first part describes the data used and presents the general methodology of accident analysis. This part is intended to present the data from which the reference levels have been established at this stage and to familiarize a potential reader with their use and/or exploitation.

The second part presents the context of the safety demonstration of automated road transport systems, in which the determination of reference values of accidentality falls.

The quantitative results of the analyzes are presented next.

Finally, it is important to specify that the analysis described it this document is based on taking into account traffic and accidents involving conventional road vehicles. Determining the cause of the accident is impossible with the data recorded. Unlike other methods of analyzing reference safety levels, the data that emerges from this analysis counts both accidents whose origin can be directly attributed to a technical system failure (in the case of conventional vehicles, we speak of more vehicle), than accidents caused by interactions of road users, called traffic hazards in the scenario-based approach[1] detailed in another document. These notions are clarified in the body of the document.

NB: this report is based, among other things, on the preliminary work carried out within the framework of the national GAME working group (globally at least equivalent) led by the Service Technique des Remontées Mécaniques et des Transports Guidés (STRMTG) (Pierre JOUVE, Florent SOVIGNET) and the working group "Connectivity use cases for the automated vehicle” led by the Direction Générale des Infrastructures, des Transports et des Mobilités (DGITM) (Jérémy DIEZ), which this report completes and details.

[1] Methodology report - Safety demonstration of automated road transport systems: Expected contributions from driving scenarios, DGITM, February 2022

Data presentation and processing

Accident analysis presented in this document and carried out as part of the safety validation for the automated vehicle (AV) were carried out using data from the *Bulletin d'analyse des accidents corporels de la circulation (BAAC)* on a sample of years within the geographical scope of metropolitan France. The sample varies according to both the requirements and expectations of the working group for which the analysis were carried out and both the data available in the parallel databases used to carry out the study.

The data, as far as OpenData is concerned, are available and downloadable from the data.gouv.fr website with free access.

Data available at the following address: https://www.data.gouv.fr/fr/datasets/bases-de-donnees-annuelles-des-accidents-corporels-de-la-circulation-routiere-annees-de- 2005-a-2019/ ,

Uploaded on 11/23/2020.

NB 1: This deliverable may be accompanied by several .ipnyb files containing all the code relating to the analysis and data processing, as far as OpenData is concerned .

NB 2: In the event that certain data has been analyzed through reserved access obtained in partnership with certain organizations, the .ipnyb files can also be shared if necessary (which in no way gives access to the raw data processed by the DGITM).

NB 3: As part of the work already carried out, .csv files have been created and allow faster access to certain categories of data, only those originating from OpenAccess data can be shared.

NB 4: The results can also be provided in the form of .csv for more readability and to facilitate dissemination.

1. General information on ONISR data

The ONISR data are described by the BAAC Guide (Ministère de l'Intérieur, Observatoire National Interministériel de la Sécurité Routière, 2017) on the one hand and by the document describing the variables used in the OpenData (Ministère de l'Intérieur, Observatoire National Interministériel de la Sécurité Routière, 2020). The BAAC terminology changed in 2006, which is why, in all the analysis proposed, only the years after 2006 will potentially be studied.

In addition, as the analysis are based on variables having evolved or having been inserted in the BAAC in 2012, we only study in this report the years after 2012. In the same way, due to the Covid-19 health crisis and significant consequences on road safety, the year 2020 has not been taken into account.

It is also useful to specify that the BAAC was updated in 2017, which results in a modification of certain variables from 2019, including for example the appearance of the maximum authorized speed. Thus, for more precise analysis, a study on aggregated years after 2019 will not be possible (in particular 2021).

a. The concept of accident[2]

The concept of a personal road traffic accident[3] is clarified by the French decree of 27 March 2017 relating to the conditions for compiling statistics relating to personal road traffic accident. A personal road traffic accident is defined as an accident which:

- causes at least one injury, that is to say a user who has required medical treatment;
- occurs on a road open to public traffic;
- involves at least one vehicle.

Roads open to public traffic are either roads that belong to the public domain and open to public traffic, or private roads open to public traffic.

The notion of vehicle is linked to the notion of movement permitted either by a motor, or by an animal, a pedal or human propulsion. As a result, personal transport devices (PTD) are assimilated to vehicles in the same way as a bicycle.

[2] The following definitions are from the BAAC guide and are taken from the wording explained in the guide.
[3]A material traffic accident can only be the subject of a BAAC file if it involves at least one injured person.

Among users involved, we can distinguish:

- uninjured: users involved in the accident who did not die and did not require any medical care;
- victims: users involved not unscathed.

Among the victims are:

- those killed: any person who dies either as a result of the accident or within thirty days of the accident;
- injured: any unkilled person who has received medical treatment for an injury, however minor.

Among the injured[4] are:

- hospitalized injured (previously called seriously injured): any victim admitted as a patient in a hospital for more than 24 hours;
- slightly injured: any victim who has received medical treatment but who has not been admitted as a patient to the hospital for more than 24 hours.

The BAAC forms are filled in for each accident known to the police by the territorially competent police or gendarmerie service. These sheets are then sent to ONISR and checked.

It should also be specified that the procedure for filling and supplying the BAAC is based on the assessment of the scene of the accident by the police and in particular by the officer filling out the form. It is therefore important to take this particularity into account during the analysis. The BAAC contains mandatory fields[5] and optional fields depending on the accident situation such as the presence or not of a special vehicle[6] involved in the accident. Most of the fields are only filled in if a particularity is to be reported, this particularity being itself dependent on the assessment of the agent on site. This concept is detailed in the next section.

b. Organization of accident reporting

The reporting of accidents is a multi-step process that it is not useful to describe in this report. All additional information is available in the ONISR BAAC guide. The reporting of information must follow the BAAC typology.

Accidentality for each of the years listed in the BAAC is made up of four different files. Each of these files provides access to particular accident characteristics (Ministère de l'Intérieur, Observatoire National Interministériel de la Sécurité Routière, 2017). The denominations are as follows:

- characteristics: which describes the general circumstances of the accident;

This file contains all the spatio-temporal characteristics of the accident including the date, the day, the atmospheric conditions, the environment (urban or rural), the type of collision, the general characteristics of the infrastructure (intersection or not) .

All characteristics-related columns are the same for each of the vehicles and users involved in the same accident.

- location: which describes the different places on which vehicles involved are associated; in the case of an intersection, the locations may be different;

This file contains all the information related to the place of the accident including the administrative category such as the typology of the road but also its geometry, the traffic regime (not available in OpenData), the number of lanes where the users involved were traveling in the accident or the development of the environment.

- vehicles: which describes all the vehicles involved;

[4] The harmonization of statistics at European level modified the definition of injury severity on 01/01/2005. We now speak of "hospitalized injured" and no longer of "seriously injured" and "killed at 30 days" and no longer just "killed".

[5] These non-compulsory fields are specified in appendix II of the BAAC guide.

[6] The BAAC describes the special vehicle as a motor vehicle having a particular use (winter service vehicle, lifting vehicle, fire-fighting vehicle, exceptional transport vehicle, household waste dumpster). Special vehicles include taxis, ambulances, firefighters, for example.

This file contains all the information relating to the vehicles involved, including the category of the vehicle, the direction of traffic, the maneuver performed and the number of occupants of the vehicle (in the case of public transport (PT)).

All vehicle-related columns are the same for all users of the same vehicle.

- users: which describes the users involved.

This file contains information specific to users such as the type of user (pedestrian, driver, and passenger), the action of the pedestrian and his location (in the case of pedestrians), the severity of the victims or the route taken by the users involved.

This last file makes it possible to distinguish all users involved in road accidents each year. Each identified user must be attached to a vehicle. In the case of pedestrians, they are attached to the vehicle that knocked them down.

Since the files are different, two variables make it possible to link each of the four files. These are essential keys for data analysis and the redesign of appropriate databases.

All of these fields make it possible to reconstruct each of the accidents. Any identification of the users and of the precise characteristics of the accident is made impossible because the data likely to transmit this type of information are hidden from the national database so as not to harm the people involved in these accidents[7].

c. Statistical terminology

The statistical study of accident data or accidentology is based on a certain number of statistical indicators to be defined and chosen according to the scope of the study. As part of this preliminary work on the accident rate of conventional vehicles in France, the aim is to establish initial quantitative data on the level of risk associated with the circulation of vehicles on motorways (divided carriageways) and on national roads (in rural areas), buses and coaches as part of urban public transport.

Therefore, the indicator chosen in this report is the accident rate. This is the risk of having an accident per kilometer travelled. The accident rate is defined as follows:

$$\text{Taux} = \frac{\text{Nb accidents}}{\text{L x nb années x TMJA x 365}} \times 10^8$$

Where nb of accidents: total number of accidents considered during the observed period
L: length of the observed section (or of the portion of the network)
Nb of years: number of years considered
TMJA: annual average daily traffic (in veh/d)

As part of this work, the traffic on the network concerned was considered in veh.km, i.e. by considering the number of kilometers travelled by vehicles of the vehicle category in question (with regard to the accidents considered) circulating per year.

The virtue of the accident rate indicator, compared to a density indicator for example, is that it enables the accident risk of sections which do not have the same traffic to be compared.

Establishing a level of safety for automated road transport systems relies on the allocation of a reference level to the components of the system. This safety objective will be based on an accident rate per hour, depending on the level of risk established from accident data. The establishment of a security objective per hour of operation and per vehicle is not the subject of this document. The switch to hourly rates was considered too sensitive at this stage and will be the subject of additional work. In fact, the change to an accident rate per hour of operation requires strong assumptions to be made on vehicle traffic speeds, which require further investigation.

[7]https://www.onisr.securite-routiere.gouv.fr/outils-statistiques/open-data

The accident rate is used in two ways in this deliverable:

- Fatal and non-fatal accident rates[8] which are calculated based on the number of fatal and non-fatal accidents respectively;
- Fatality and injury rates, which are calculated based on the number of fatalities and injuries (serious or minor) respectively.

In the remainder of this report, the terms “accident rate” always refer to a value per kilometer travelled.

2. Use of data and main limitations

The knowledge of the BAAC data is a first step in the appropriation of the database. As indicated previously, the progressive evolution of the BAAC guide implies that all the files are, on the one hand, not homogeneous between them, and that on the other hand, disparities between the years exist. This feature thus requires a formatting of the database according to the desired use and the purpose.

The processing of the files over all the years thus required choosing priority columns to be processed and kept, which constitutes an obstacle to the automation of the analysis of this data. Moreover, as the encoding is not the same, special treatment for certain years is sometimes necessary. The main limitation of this approach is that a modification of the initial database, even substantial, requires starting over. Thus, the manipulation of data in analysis requires a minimum knowledge of the structure of the database, which makes this formatting an essential step in analytical work.

Another feature of this database, already mentioned above but which makes it rich, is that it is the result of a long and multiple process of collecting road accidents. BAAC data is based on the completion of reports by the police (national gendarmerie and national police) at the time and place of the accident. These reports are therefore directly dependent on the agent present at the time of the events, which gives a fairly significant variability in the filling of the BAAC fields (Ministère de l'Intérieur, Observatoire National Interministériel de la Sécurité Routière, 2020).

Although the reports are harmonized and checked before designing the annual database several times, inconsistencies or questions may persist on the way certain fields are completed.

In addition, the great diversity of fields to be filled in by the agent can create biases for certain variables that can be filled in optionally. For example, the difference between the missing data, the non-referenced data, the data not applicable or the category other(s) can give rise to interpretation when the information is given by the police officer. The same may apply to an accident involving rail or guided transport. An accident involving a tram or a train in the event of a collision between the tram or the train with another road user, will have a specification provided in the BAAC sheet for the infrastructure variable (code "railway"). On the other hand, an accident occurring near railway tracks will be specified by the intersection variable (code "level crossing") but does not predict a possible collision or the involvement of rail or guided transport in the accident. Thus, the existence of two different variables to code two close situations, can generate differences during filling. Filling in the "level crossing" code is likely to reflect a cause and effect link in the accident or a simple coincidence on the location of the accident.

Similarly, the filling of weather conditions which can affect the driver's visibility or even factors linked to the place of the accident (see the previous example) can give rise to differences in assessment during filling.

All these examples therefore represent potential biases in the analysis of the data and therefore in the results obtained. These results should therefore be viewed with caution. In addition, the accident data discussed here are based on conventional vehicles, so caution should be exercised when using these results and potential interpretations.

In general, the establishment of a reference safety level for ARTSs in France is based on the study of empirical data. The dependence of the results on the data used and the assumptions made subsequently is irrefutable and must be taken into account when reusing these results. The quantified data cannot be dissociated from its hypotheses and from the method constructed.

[8] This notion linked to the accident event is taken up again in part 6.c.

3. Construction of the working database

ONISR provides pre-processed statistical analyzes in its annual road safety reports. Insofar as our analysis are specific to a particular use, they require their own treatment and therefore to format the base on which to rely. The construction of a new database in order to work more easily on the data is a first stage of the work and it constitutes an important part of the analysis because it is at this level that the selection of the variables is done, as explained previously.

This part describes the method of building the working database, with regard to the method elements independent of the use cases used. This method is broken down into the following steps:

1. *Building the database*
2. *Data processing*

This paragraph precisely describes the processes and stages of building the database[9].

NB: The script description of the .ipynb can be found in Appendix 1.

Database design follows a logical process of breaking down data and recomposing only the data that will be useful for analysis. The rest of this part is written as much as possible in the form of an enumeration so as to clearly identify all the stages:

- The data in OpenData being designed by year, it is necessary to select a period over which the analysis will be carried out - in our examples, this could have been from January 1, 2012 to December 31, 2019, which represents 8 years of study .
- It is then necessary to select the variables identified beforehand and useful in the context of the work carried out. Insofar as all the data is not necessary and in the same format, this step is a mandatory step.

  It is also, as indicated above, the main obstacle to the automation of the new database construction because it requires a fairly detailed knowledge of the database and its content on the one hand, but also of the encoding used.
- Then, once these variables have been identified in the correct files, a single database should be gradually rebuilt containing all the information desired per year or over the entire period depending on the desired use.

  The steps for this point are detailed in the code provided in Appendix 1, but are not particularly necessary for understanding since these steps are very largely dependent on the format of the ONISR database.
- At this stage the file contains all the filtered data (regardless of their original file among the four described in the first part) for all the years selected – if a global database is created – for the accidents listed in mainland France[10].

  The retention of accidents that occurred only in mainland France may be the subject of a choice for the concern of consistency with other data sources[11].
- This new database is used to conduct the analysis.

Before processing the data, the macro use cases for which accident data are used should be designed. Although the design of the database requires having already thought roughly about the types of use cases, it is at this stage that the choice of taking into account the victims and the way of counting them comes into play.

[9] The database was produced using Python software. Knowledge of the tool is necessary to reproduce the diagram but not essential to understand the reasoning followed.

[10] The concept of mainland France is easily selectable and is part of the selection on the variables. In the remainder of this document, it has been chosen to work only on continental territory + Corsica.

[11] In the following analysis, traffic flow data were listed on mainland France too, which requires having consistent and compatible accident data.

Indeed, road traffic accidents describe different types of victims including the killed, the physically injured[12] and the uninjured. Within the framework of these analysis, the choice of victim identification was different according to the goal. However, it is useful to remember that this choice must be consistent with the period selected, in particular because of the evolution of the fields of the BAAC.

Another important criterion to take into account before carrying out the analysis is the format of victim identification, which must correspond to the format of the associated traffic data in order to subsequently obtain kilometric rates and/or hourly rates[13]. Indeed, the determination of accidentality is carried out with the aim of obtaining accident rates – fatal or not – relating to certain use cases.

With regard to the transition to an accident rate, a choice must be made between a rate representing a number of accidents per kilometer travelled, a number of vehicles involved in accidents per kilometer travelled or even a number of victims per kilometer travelled. Each of these units can then be declined according to whether the aim is to establish a mortality rate or a rate of injuries.

On the other hand, the handling of damaged vehicles can introduce an additional bias by counting the same vehicle several times – if it is for example associated with both fatalities and bodily injuries. It is therefore important to reason strictly in this case:

Introduction to TRAXY

The TRAxy software is an information system (IS) relating to accidents in France. It is the property of ONISR and is presented as a database accessible under restricted access. In the rest of the report, the use of TRAxy is described and was necessary, in addition to the OpenData described briefly in this part.

Access to data complementary to OpenData was necessary in the context of the use cases studied. Some characteristics were not accessible via OpenData and in particular the selection of study areas in the form of public establishments for inter-municipal cooperation (EPCI) was not possible.

The use of TRAxy within the framework of this study was made possible by the creation of a specific access to the DGITM by the ONISR. The completeness of the analysis and the targeted work on certain specific variables made the work more exhaustive. The creation of an access for the DGITM was carried out in order to be able to process certain use cases not referenced in the OpenData because of the fields that are too specific and not representative of the overall accident rate.

More generally, the use of TRAxy for accidentality considerably facilitates the work of recording accidents since it is possible to extract specific features from the database by means of queries. This is why, in certain specific cases, the use of TRAxy has been preferred to that of a database designed on Python from OpenData .

Nevertheless and when TRAxy was needed, its use was complemented by downstream work on Python. In particular, the processing and use of databases extracted from TRAxy have been automated on Python, in particular the counting of victims and the comparison of accident data with traffic data.

Also, the use of TRAxy was favored for macro use cases associated with public transport (public or school) while the use of the database built on Python was favored for macro cases of use associated with the particular vehicle. The global methodologies associated with each of the use cases defined above are explained below. It is useful to specify that the methodologies described in the following part are intrinsically linked to the data used. In addition, insofar as Python and TRAxy are two SI describing the same database, the overall coherence of the method is ensured. Finally, the distinction between the two use cases VP and STPA was marked on the construction of the extractions while their analysis and their exploitation were analogous

[12] According to ONISR terminology, an injury means either a light injury or a hospitalized injury. The distinction is made according to the duration of the victim hospitalization, which should not exceed 24 hours for a minor injury. An injured person hospitalized or seriously injured is a victim hospitalized for more than 24 hours.

[13] This document does not propose an hourly rate. The difficulty of having access to reliable average speeds would force us to introduce other strong hypotheses on each of the use cases considered.

4. Accident data processing

Whichever analysis is chosen, the data processing process will always be the same[14]. The study of accidentality involves working with mixed data, each accident being the result of a collision between at least two entities. The collision may be between two or more vehicles, between one or more vehicles and a pedestrian, between one or more vehicles and/or an obstacle on or near the roadway[15].

As explained above, the study of accidentality and its analysis is based on the consideration of several macro use cases, for which characteristics at the location of the accident, the infrastructure, the category of vehicle and user are required.

In most analysis, it is a matter of having access not only to the number of accidents but also to the seriousness of the victims, i.e. the number of injuries, whether hospitalized or slightly injured[16], and to the number of killed people.

It is also common to make a selection on the type of vehicle, insofar as we seek to objectify the level of safety for the ARTS from the accident rate of conventional vehicles.

The analysis is then done in two stages, whatever the purpose due to the format of the data:

a. Selection of the identifiers of all the accidents in which at least one of the vehicles or users involved meets the chosen criterion[17].

b. Count of all the victims and/or all vehicles involved in an accident for the severity chosen – victims killed or injured – whose accident number is contained in the first selection. For this step, the query used in *Python* is *.isin()* [18].

The method described in this part is very generic, it has been automated in order to make the process adaptable to any new analysis based on these accident data. An update of the new databases available each year is necessary. The prerequisite for using this step-by-step method is to work with a formatted and pre-built database.

The code that appears in Appendix 1 is commented, it is the generic code. Editable fields according to the desired analysis are clearly identifiable and points of vigilance are also identifiable, in particular concerning the paths used and implementation precautions.

The rest of the deliverable describes more precisely the work established within the framework of the various working groups for which the data has been processed. The results were established for mainland France, for the chosen time sample.

[14] Modulo the fact that we choose to count the number of victims or vehicles involved, which changes the work of counting the data, the rest of the script works in a similar way.

[15] According to the definition of a road traffic injury accident specified above.

[16] ONISR experts recommend working by considering injuries as a single category and not dissociating hospitalized injuries from those with minor injuries insofar as the counting of injuries may be biased.

[17] It is also useful to ensure that the filters chosen are consistent with each other and do not give access to aberrant data whose meaning and significance could controversial. Multiplying filters usually results in destroying the desired precision on the data and the macro use case. Thus in certain situations, it is not possible with OpenData only, to go back to the desired use cases.

[18] It is also possible to reason with the inverse query so as to remove from the final result all the users who do not meet a criterion.

Levels of accidentality for the lighting of safety references

This section focuses more specifically on the methodology developed from the ONISR accident data presented above.

The objective of this work is to provide useful insights for subsequently defining reference safety levels, based on the analysis of the accident rates of conventional vehicles on relevant use cases.

The results obtained are not independent of the method used, which is why the results should be coordinated with the method used and its characteristics. In addition, the diversity of traffic and road accidents in European countries could require, on a European scale, to harmonize the quantitative data obtained, although the method for demonstrating the reference level could be free.

5. Reference level of safety based on accidentality

The following study was carried out in France on a national scale in order to establish benchmarks based on the accident rate of conventional vehicles, benchmarks that can be used to shed light on the level of safety of automated road transport systems.

The method is based on the study of different sufficiently precise and generic macro use cases, considered as particularly safe use cases. These macro use cases are not set in stone and can be modified according to identified needs.

a. Macro reference use case for accidentality

These particularly safe macro use cases, referred to as reference for ARTSs, refer to macro use cases with low accident potential related to the number of kilometers travelled on the roads in France. Among these macro reference scenarios are:

- Urban public transport for the automated public transport use case – which here designates an urban use case whose qualification in relation to the BAAC data is "in built-up areas"[19].
- Driving on the highway for the passenger car use case – which is characteristic of safe driving[20].

Furthermore, some macro use cases can be interesting to analyze from the point of view of accidentality because they represent a reference network that is conversely particularly accident-prone. Among these macro use cases is:

- Driving on national roads for the passenger car use case – which this time designates a rather interurban to rural use case designated by the term "outside urban areas" in BAAC terminology.

Urban public transport refers to all vehicles whose primary purpose is for public transport. As a result, all vehicles listed as "public transport" are considered, regardless of their platform or type. In TRAxy, the terminology refers to coaches and buses without distinction. The restriction to the urban area is permitted by a selection of accidents occurring "in built-up areas" only by considering that the behavior of coaches traveling in urban areas is similar to that of buses.

Other use cases could be designed and worked on later, depending on the needs and taking into account the available data, whether on accidents or on traffic. Needs for specification of use cases will appear as a result of the first analyses.

[19] Details on the nomenclature used are given in the following section.

[20] This analysis was supplemented by an analysis using ASFA data for the following two purposes:

- ensure the significance of the indicators presented in this report;
- present indicators according to the type of vehicles circulating on the infrastructure and in particular the distinction between passenger vehicle and heavy goods vehicle.

This part is presented in Appendix 2.

b. *First assumptions*

By carrying out these analyzes on accident rates in France, several hypotheses have been formulated:

i. The reference accident rate for conventional vehicles can be transposed to vehicles equipped with an automated driving system[21]. The data used and the results obtained are from and applicable to conventional vehicles, for which the requirements will certainly be less stringent than for automated vehicles. Being part of the innovative technology, the acceptable risk levels will definitely be higher.

   It is therefore appropriate to take these results with hindsight, considering that they are not absolute.

ii. The territory observed is metropolitan France. The results therefore do not take into account the overseas departments and territories, where the accident rate is higher and different from that observed in mainland France. In addition, the territories observed do not represent the entire French metropolitan territory insofar as macro use cases have been chosen.

   Therefore, for the case of the VP, the cases of use studied are those related to environments rather motorway or interurban; the urban component of accidentality is therefore totally absent from these results at first.

   In the case of the STPA, the use case studied is linked to urban public transport. A distinction has been made to specify urban public transport in the densest French agglomerations.

   Finally, the reasoning on the French national scale cannot be used without adaptation on a European or international scale. Similar work should be carried out on a national scale by the Member States so that guides can be drawn up by the European Commission.

iii. For the particular mobility use case, all types of vehicles circulating on the infrastructure were considered and no distinction was made between the circulation of light vehicles and heavy goods vehicles.

   Appendix 2 presents an analysis justifying this hypothesis based on data from motorway concession holders (ASFA).

iv. The results presented in the following are taken from averages taken over several years, which has the advantage of smoothing the data and reducing the impact of variability[22], but also has the disadvantage of smoothing the decreasing trend of accidentality in France in recent years[23].

   The literature recommends carrying out analysis over a minimum of 5 years; this is why it was chosen to work over a minimum of 5 years (7 years for the passenger car use case and 8 years for the urban public transport use case).

   For the use case of individual mobility on motorways, 7 years were considered sufficient (given the available data). Comparisons with analyzes carried out over 3 years were made; at first glance and without additional analysis, the accident rate results are comparable.

   For the case of urban public transport, the choice was made to take a significant number of years of 8 years (from 2012, year of transition of the BAAC) in order to compensate for the biases of remarkable accidents involving a transport vehicle in common that occurred in some years. The smoothing of these remarkable accidents was considered necessary within the framework of this work.

v. This work being based on empirical data, the uncertainties are numerous and accumulated on the final results although the overall methodology is verified and validated. In addition, the sources of data are

[21]Although the literature agrees that the arrival of automated vehicles in traffic will have a significant impact on accidentality by drastically reducing accidents – a study by the NHTSA carried out in 2016 concludes that more than 90% of accidents in United States would be caused by human error (National Highway Traffic Safety Administration's Center for Statistics and Analysis, 2017), it should be noted that it is currently difficult to predict the accident rate that could be directly attributed to the technical failures of the system.

[22]The year 2020 has been rejected due to the health crisis and the significant impact it has had on road safety, as announced in the first part.

[23]The impact of the study period on accident rates has not been quantified properly and may be subject to additional information to refine the results.

multiple, which implies being cautious about the exploitation and the standardization of the objectives and level of risks that emerge.

As mentioned in Appendix 2, the additional analysis based on ASFA data is not based on accident data comparable to BAAC data. Motorway concession companies do not use the same accident reporting process as ONISR: the proportions are comparable, the exact values of the number of accidents or victims are less so.

vi. Finally, the general hypotheses and the processing of the data explained below were an assumed bias on the part of the French experts. Once again, the quantitative results obtained should not be dissociated from the hypotheses and the method used.

6. Analysis methodology by macro use case

The analysis methodologies were applied to the use cases considered. These are not independent of the data available for the different use cases and in particular of their format. There are thus two methodologies associated with the two macro use cases studied:

- individual mobility, for which the accident data of the two associated macro use cases were built and then used in Python.

  Traffic data associated with individual mobility use cases was retrieved in OpenAccess from the website of the *Service de la donnée et des études statistiques (SDES).*

- public mobility, whose accident data for the associated use case was retrieved via TRAxy and then used in Python.

  The traffic data associated with the public mobility use cases was retrieved via a DGITM reserved access from the results of the annual urban public transport surveys.

a. Methodology for particular mobility

For particular mobility, two macro use cases were considered, representing reference scenarios for accidentality. The results obtained for these two use cases are based on OpenData, so it is quite possible to find them rigorously by following the method below.

Preliminary work for accident collection was to ensure the consistency of the different data sources to establish a reference level. As explained above, the accident data come from the national databases held and filled in by the ONISR, while the traffic data come from the annual SDES databases[24].

The annual reports of the SDES are published as national statistics and are not adjustable according to need. In this case, it is necessary to match the ONISR data to the data shared by the SDES[25].

Accidents on roads with motorway characteristics

The use case relating to accidents recorded on the motorway network must specify the scope of the data that is included in the analyses. More specifically, the SDES publishes data relating to kilometers travelled (vehicles.kilometres) on the entire motorway network defined as follows: the motorway category includes all motorways under concession and motorways not under concession and are considered in unconcessioned motorways urban expressways and national interurban roads with motorway characteristics (divided lanes).

As part of this study, the accident data corresponding to the category “roads with motorway characteristics” defined by the SDES were extracted from the ONISR database. It was necessary to build a new database, derived from OpenData as ONISR uses the administrative categories of roads. Thus, the “motorways” typology only designates in the BAAC concession and non-concession motorways within the meaning of the road typology. The extension to all lanes with motorway characteristics was carried out using the Marvell software[26].

The data from Marvell made it possible to obtain files containing all the accident numbers as well as the administrative categories of the roads on which accidents occurred. From these files, the selection of accidents that have occurred on roads with motorway characteristics is permitted and integrates data similar and consistent with those which were the subject of the traffic reports, published by the SDES.

[24] Traffic data was taken from Transport accounts and traffic reports for the year 2019, available in OpenAccess on the SDES website.

[25] This notion of correspondence between the databases is explained in the following section, relating to accident rates on roads with motorway characteristics.

[26] Marvell is a static simulation tool which, based on road traffic estimates updated every year, makes it possible to predict traffic demand on the national network. Marvell's accuracy on the main axes was deemed sufficient to use the traffic generated for the two use cases studied. Marvell was used to match the geolocation of accidents in the ONISR database to the administrative category of the road on which the vehicle was traveling before the accident. This manipulation was necessary insofar as the categories of roads defined by the ONISR and by the SDES are not the same.

Accidents on national roads

The national roads use case refers to the administrative category of national roads. For this category of road, no precision is necessary to understand the data. The data published by ONISR and the traffic data published by SDES refer to the same road traffic.

Analysis

Results presented in the last part of this report are averaged over the period from 2012 to 2018 in mainland France. It was also chosen to design an indicator based on the number of victims for each use case for each of the following two categories:

- the number of killed within the meaning of the ONISR terminology;
- the number of injuries, without distinction between minor injuries or hospitalized injuries.

In addition, the process also made it possible to identify the number of fatal accidents.

Traffic associated with each of the use cases for every year considered was retrieved directly via the transport accounts and the SDES traffic reports updated in 2020.

The accident rate was therefore calculated using four different indicators:

- the mortality rate which is presented as the number of people killed in relation to all the kilometers travelled by vehicles of the category considered in traffic;
- the rate of injuries which is presented as the number of injuries in relation to all the kilometers travelled by vehicles of the category considered in traffic.
- the rate of fatal accidents which is presented as the number of fatal accidents in relation to all the kilometers travelled by vehicles of the category considered in traffic.
- the rate of non-fatal accidents which is presented as the number of non-fatal accidents (having resulted in a maximum of serious injuries) in relation to all the kilometers travelled by vehicles of the category considered in traffic.

In the absence of traffic data by network and by type of vehicle, the indicators presented represent overall accident rates, all categories of vehicles considered[27].

All victims associated with accidents occurring on roads with motorway characteristics or national roads were counted, including pedestrians.

b. Methodology for collective mobility

For collective mobility, a macro use case has been studied, representing a reference use case for accidentality. For this use case, it is not possible to be able to go back precisely to the results since the data used is not accessible in OpenAccess.

All victims are considered, whether users inside urban public transport, or users outside and impacted by the accident (pedestrians, other vehicles). Collective mobility therefore implies in this study that all types of platforms and all categories of vehicles whose main function is attributed to the public transport studied are taken into consideration. Thus, the results do not distinguish between accidents and victims according to the type of vehicle and in particular according to the presence or absence of a seat belt or according to whether they are standing or sitting.

Urban public transport

For this use case, the study only integrated urban areas according to the BAAC nomenclature "in agglomeration" on French regions[28] on TRAxy software. The choice of regional territories makes it possible not to take into consideration Paris region, which on the one hand does not have an accident rate comparable to the rest of the metropolitan territory, and on the other hand because of the desired correspondence

[27] Appendix 2 presents an analysis based on ASFA data by dissociating the type of vehicles on the network.

[28] The 12 regions considered are as follows: Auvergne-Rhône-Alpes, Bourgogne-Franche-Comté, Bretagne, Centre-Val de Loire, Corse, Grand Est, Hauts-de-France, Normandie, Nouvelle-Aquitaine, Occitanie, Pays de la Loire, Provence-Alpes-Côte d'Azur; in order to match accident data as closely as possible to traffic data from TCU surveys.

between the data of accident and traffic data from traffic surveys. Indeed, traffic surveys are characteristic of urban public transport in mainland France outside Île-de-France, for which the *Observatoire de la mobilité en Île-de-France (OMNIL)* carries out data.

ONISR definition specifies that "an accident is declared in built-up areas if it occurs between panels EB10 (entrance to built-up area) and EB20 (end of built-up area)". It is also specified that the non-urbanized areas inside the agglomeration signs must be coded "in built-up areas".

The first results revealed a global reference level at the territorial scale, but it seems interesting to reflect on a finer scale at the local level. Moreover, the accident rate on the scale of the metropolitan territory places so-called urban areas in metropolises or large agglomerations and less urbanized or even predominantly rural areas, although in agglomeration within the meaning of the administrative classification. Indeed, the BAAC only specifies that "localities and hamlets not indicated by signs EB10 and EB20 are not considered as agglomerations and must be coded outside built-up area". Thus the villages inside road signs EB10 and EB20 are considered "in built-up areas".

The more local work (urban mobility focus in the report) was carried out at the level of the territories around the largest cities in France. The selection criterion for the central city was to have at least 50,000 inhabitants. The choice of municipalities and associated territories was very dependent on the data available in the databases studied and in particular induced by traffic data. In total, 27 territories were taken into account outside Paris and Île-de-France, all the territories studied being at least agglomerations and metropolises at best. This work was carried out thanks to the possible selection in TRAxy by territories, which considerably lightens the work to be carried out at the level of the administrative limits of the territories.

The hypothesis of a correspondence between territories and urban public transport network operator was made in most of the territories concerned (this is the case for 26 of them). The exception of Marseille, whose Aix-Marseille-Provence metropolis is managed by several operators, was overcome by verifying the location of the accidents and by deleting from the database those not taking part of the territory managed by the Régie des Transports de Marseille[29]. The case of the Île-de-France region has also been treated separately, according to the territories listed in the SDES traffic data: Paris, Petite couronne and Grande couronne. The Petite couronne includes the following Île-de-France departments: Hauts-de-Seine, Seine-Saint-Denis and Val de Marne; while the Grande Couronne is made up of the following departments: Essonne, Yvelines, Val d'Oise and Seine-et-Marne. For the case of the Paris region, it was not necessary to dissociate the network operators and the territories given that the traffic data is already aggregated according to the three aforementioned territories.

The accident data by territories, in total 30 territories, concerns all the victims of the 8 years considered from 2012 to 2019 for respectively each of the two categories of victims considered.

For this use case, it was possible to work with the killed and injured, given that the number of victims of each of the two categories is non-zero in most cases: 6 territories, i.e. 20%, did not record no fatal accident during the 8 years considered.

Traffic data from annual urban traffic surveys

With regard to traffic data, urban public transport surveys were used, based on reserved access[30] to the database shared between the organizations that generate this database, of which the DGITM is a part. More generally, these surveys are completed every year by operators of urban transport networks in France[31]. These surveys are distributed to all operators of public transport services in France.

[29] On the scale of public transport and accident rates in the RTM area of competence, this sorting action represents 1 fatality and 109 injuries, ie respectively 10% and 20% of accidents in the Aix-Marseille-Provence metropolitan area.

[30] Access to more complete data than the data downloadable via the SDES OpenData was necessary due in particular to a need to have access to the data by territory.

[31] This is also the reason why the hypothesis of a perfect or almost perfect correspondence between the EPCI territory and the territories operated by the public transport organizations was made.

Operators have no obligation to complete this survey, nor any obligation to complete it in full in the positive case. As a result, it was necessary to work on the data from this database prior to the extraction[32], in particular in order to remove any networks that had not completed all the surveys over the first five years considered (2012-2016)[33]. By carrying out this work and assuming that the data is only valid and taken into account if the survey has been carried out consecutively by the operators, barely 2% of the kilometer data travelled by the buses has been deleted (exactly 1.7 %).

This hypothesis was carried out insofar as we envisaged a more complete work on the scale of the whole of France and not only on a few targeted territories.

The database obtained contains all the data relating to the urban traffic surveys from 2012 to 2019 inclusive, a selection has been made on the kilometers travelled by bus[34].

Analysis

The results presented in the last part of this report are averaged over the whole of the years 2012 to 2019 for the whole of mainland France. It was also chosen to identify the number of victims for each of the use cases for each of the following two categories:

- the number of killed within the meaning of the ONISR terminology;
- the number of injuries, without distinction between minor injuries or hospitalized injuries.

In addition, the process also made it possible to identify the number of fatal and non-fatal accidents.

Contrary to the data for individual mobility, access to victims via TRAxy implies having worked via tables by travel mode. This distinction therefore implies that the codes associated with the use case of collective mobility had to be adapted. The traffic associated with the use case for the years considered was retrieved using the analysis described above.

The accident rate[35] was therefore calculated using four different indicators:

- the mortality rate which is presented as the number of people killed in relation to all the kilometers travelled by vehicles of the category considered in traffic;
- the rate of injuries which is presented as the number of injuries in relation to all the kilometers travelled by vehicles of the category considered in traffic.
- the rate of fatal accidents which is presented as the number of fatal accidents in relation to all the kilometers travelled by vehicles of the category considered in traffic.
- the rate of non-fatal accidents which is presented as the number of non-fatal accidents (having resulted in a maximum of serious injuries) in relation to all the kilometers travelled by vehicles of the category considered in traffic.

All users and in particular victims associated with accidents involving urban transport have been counted. All analysis were carried out in Python and can be shared.

[32] Again, the preliminary analysis were done via Python.

[33] The verification was not carried out over the whole period of 8 years due to the nature of the bases available. Indeed, for the years prior to 2016, a file directly processable on Python was collected, which was no longer the case for the years after 2017, where the nomenclature of certain fields and the format of the data changed. Therefore, a strong assumption was made and consisted in considering that if the network operators had completed the surveys for 5 consecutive years, then we could take them into account in the analysis.

[34] The urban traffic surveys do not only contain kilometric data travelled in an urban environment but also data outside built-up areas when operators operate bus lines deployed between two built-up areas, for example. Within the framework of this study, only the data relating to the operation of the urban transport networks of the central agglomerations were selected for the specification of the densest agglomerations.

[35] Reminder: in this report, the term “accident rate” is used specifically to designate the ratio of the number of victims considered by the number of kilometers travelled, which corresponds to an accident frequency. The objective here is once again to present initial work and not to give a definitive study of accidentality.

c. Generic method for taking accidents into account

In the previous paragraphs, the method consists in considering the number of victims of accidents associated with the use case studied. The association of accidentality with the number of victims by typology is often perceived as more explicit and intuitive. However, it can be useful to have an accidentality based on the notion of event and not on the notion of victims strictly speaking.

Indeed, it is possible to choose to consider accident events and not the number of victims caused by an accident. By considering the accident event, we then observe data relating to a precise situation which leads to a precise consequence: we speak of the uniqueness of the event. In terms of data processing, it is therefore useless to go back and have access to the vehicles involved and the precise victims involved. On the other hand, the notion of accident alone is not precise enough because it does not depend in any way on the seriousness of the latter. In the analysis of the safety demonstration, the notion of reference is made possible by access to the most restrictive notion which results in the indicator linked to the “number of accidents leading to fatalities”.

In the rest of the study, all the calculations were carried out a second time taking into account the "accident" event. For each of the use cases, the number of accidents was recovered in order to have a generic indicator on accidentality on the one hand and on the other hand, the number of fatal accidents was recovered.

These two indicators are comparable in magnitude to the following indicators respectively:

- The order of magnitude of the number of accidents is comparable to the order of magnitude of the number of injuries insofar as a bodily accident is referenced if and only if it contains at least one victim (a user not uninjured a casualty) ;
- The order of magnitude of the number of fatal accidents is comparable to the order of magnitude of the number of users who died at the time of the accident and therefore to the number of deaths.

In both cases, the results are lower when the hypothesis is taken on the accident event: an accident has at least one victim and a fatal accident has at least one fatality. In the rest of the report, the two analyzes were carried out and the results present the same orders of magnitude, whether we count the accident events or the number of actual victims.

Presentation of the results

The results presented below are therefore derived from the treatment processes presented above. It is useful to specify that the use cases of individual mobility were averaged over the period 2012-2018 while that relating to collective mobility was averaged over the period 2012-2019.

For the rest, the data is totally comparable because it comes from the same databases: that of ONISR for accident data and that of SDES for traffic data. Although the method of retrieving this data varies depending on whether the data is in OpenAccess or reserved access, the rates obtained are comparable to each other for the same use case as well as between the use cases studied.

It is useful to recall and specify that this part presents different results. First of all, the results by use case of the mortality and injury rates are presented, that is to say by counting the victims. Then, a second part presents the accident and fatal accident rates by use case, ie by counting the number of accidents that have occurred only. Finally, a focus is made on the use case of urban public transport.

7. Presentation by use case of fatality and injury rates

The presentation of the results of the mortality and injury rates by use case was a first entry for the appropriation of the results.

The results below present the death rates and injury rates for each of the use cases considered (for individual mobility on motorways and national roads and for urban public transport).

The results obtained are available in Table 3 below.

Table 3: Table presenting the mortality rates and the injury rates per km travelled [36] for each of the use cases concerned, averaged over each of the study periods

Average accident rate by macro use case	**Mortality rate**	Injury rate
Motorways and lanes with motorway characteristics	**$1.15.10^{-9}$**	$2.64.10^{-8}$
National roads	**$5.71.10^{-9}$**	$6.79.10^{-8}$
Urban public transport	**$3.96.10^{-8}$**	$1.49.10^{-6}$

In this table, the mortality rates for ARTSs are important insofar as they characterize a binding reference level for each of the use cases observed.

Nevertheless, and as mentioned previously in the methodological section, by working over a large number of years, the decreasing trend in accidentality is smoothed out. This hypothesis certainly has an advantage for the case of urban public transport in order to smooth out serious accidents that have occurred occasionally in certain areas, but prevents us from observing more particularly the decreasing trend of accidentality, which would contribute to developing a level of reference safety as restrictive as possible for automated road transport systems, for which we know that the acceptability of the level of risk will be lower than for conventional public transport.

The table below shows the rates observed for each of the three use cases over the last three years considered. It is quite easy to notice in this table the rather increasing trend in accident rates (whether in terms of mortality or the number of injuries) for particular mobility, while the use case of collective mobility tends to observe a slight decrease in accidentality, as expressed above.

In general, the rates obtained for urban public transport on a study only over the last three years ie. from 2017 to 2019, are lower than the overall rates from 2012 to 2019.

In the case of urban public transport, a fairly clear increase from 2017 to 2019 in the number of kilometers travelled is observed; multiplied by more than 2.5 for urban mobility. The number of accidents remains

[36] All the results presented in the body of the document are rates expressed in relation to the kilometers travelled.

generally stable annually, which suggests a rather decreasing trend in accidents per kilometer travelled[37].

Table 4: Table presenting the mortality rates and injury rates for each of the use cases concerned, averaged over the last three years of the periods studied (either 2016-2018 for special mobility, or 2017-2019 for shared mobility).

Average accident rate by macro use case	**Mortality rate**	Injury rate
Motorways and lanes with motorway characteristics	**$1.44.10^{-9}$**	$3.95.10^{-8}$
National roads	**$7.07.10^{-9}$**	$9.54.10^{-8}$
Urban public transport	**$2.59.10^{-8}$**	$9.19.10^{-7}$

In the case of special mobility, the results obtained seem to indicate a slight increase in accident rates. The SDES traffic reports, accessible in OpenAccess, nevertheless show a clear increase in vehicle kilometers travelled. Roughly and without going into more detail, these results tend to highlight a slight increase in the number of fatalities and injuries in accidents occurring on roads with motorway characteristics and national roads[38].

No statistical analysis was performed on the significance of these results compared to the previous ones. This work may be explored further in a future study.

8. Presentation by use case of accident and fatal accident rates

The following part focuses on accidents and more particularly on fatal accidents, which implies other reference levels, nevertheless of the same order of magnitude as the results counting the number of victims.

The results presented in the following table show similar orders of magnitude between the rates obtained respectively for mortality and accidents leading to at least one death and respectively the number of injuries caused and the number of accidents.

Table 5: Table presenting the rates of non-fatal accidents (injury rate) and fatal accidents for each of the use cases concerned, averaged over each of the study periods

Average accident rate by macro use case	Fatal accident rate	Non-fatal accident rate
Motorways and lanes with motorway characteristics	**$1.03.10^{-9}$**	$1.90.10^{-8}$
National roads	**$5.07.10^{-9}$**	$4.66.10^{-8}$
Urban public transport	**$3.64.10^{-8}$**	$8.31.10^{-7}$

At the end of these results, the following table presents the figures obtained for the reference linked to mortality respectively for the mortality rate and the rate of fatal accidents.

Table 6: Table presenting the mortality rates and the fatal accident rates for each use case, averaged over the period considered

Average accident rate by macro use case	Mortality rate	Fatal accident rate
Motorways and lanes with motorway characteristics	**$1.15.10^{-9}$**	**$1.03.10^{-9}$**

[37] It is useful to specify here that we must not lose sight of the fact that the correspondence in the number of accidents is inexact since we are working with the registered victims of accidents. In the case of public transport, the number of victims is always higher given the high occupancy rate, contrary to a study on private transport.

[38] Please note that it is important to distinguish between motorways and lanes with motorway characteristics. Based on the figures put forward by the Association of French Motorway Companies (ASFA), the trends are not quite the same. Indeed, the ASFA only assesses the concessioned motorways of the national motorway network and does not take into consideration the non-conceded motorways and more generally the roads with motorway characteristics.

National roads	**5.71.10 $^{-9}$**	**5.07.10 $^{-9}$**
Urban public transport	**3.96.10 $^{-8}$**	**3.64.10 $^{-8}$**

This comparison at the level of two different rate indicators based on accident mortality data shows that the considerations of the number of deceased victims and the number of accidents leading to fatalities lead to comparable reference levels for reasoning. per use case. On the other hand, no statistical study on the significance of the comparisons was carried out for this work.

9. Focus on public transport by agglomeration

For the use case of urban transport, a focus was carried out on 30 dense territories and particularly representative of the densest agglomerations in France. Among the 30 territories observed[39], there are 21 metropolises, 6 urban communities or communities of agglomerations and Ile de France (divided between Paris, the inner suburbs and the outer suburbs).

The presentation of the results is carried out in the report in two formats:

- Tables presenting the numerical values grouped into quartiles from the rate values obtained. These quartiles make it possible to categorize each of the agglomerations or metropolises according to whether it is not very accident -prone (which would induce a restrictive reference level) or whether it is rather accident -prone (which would induce a less restrictive reference level).
- Graphs presenting the distribution of the rates obtained according to the population density of each of the territories in the study. In this way, the heterogeneity of the representative rates of accidentality is transcribed in relation to the local population density[40], anonymously.

In the remainder of this paragraph, reference is made to metropolises or agglomerations by the term EPCI, which designates public establishments for inter-municipal cooperation.

a. Death rate and injury rate

The first indicator presented is that associated with the victims recorded: a distinction is thus made between the mortality rate and the rate of injuries, respectively by counting the number of people killed and the number of people injured (without distinguishing between minor injuries or serious injuries).

The established classification is thus based on reasoning based on the geometric quartiles of the values of the series of mortality rates and injury rates. The categorization of the results into 4 levels was chosen in order to show the heterogeneity of the accident rate of urban public transport in France, for the representation. The following table shows the levels obtained.

Table 7: Classification obtained from the distribution of mortality and injury rates of the networks studied – representation from the geometric quartiles of the rate values

Quartiles	1st quartile	2nd quartile	3rd quartile	4th quartile
Mortality rate	]0; 7.30.10 $^{-9}$]	]7.30.10 $^{-9}$; 2.59.10 $^{-8}$]	]2.59.10 $^{-8}$; 4.41.10 $^{-8}$]	]4.41.10 $^{-8}$; 1.06.10 $^{-7}$]
Injury rate	]0; 4.71.10 $^{-7}$]	]4.71.10 $^{-7}$; 7.38.10 $^{-7}$]	]7.38.10 $^{-7}$; 1.20.10 $^{-6}$]	]1.20.10 $^{-6}$; 4.03.10 $^{-6}$]

[39] Metropolitan areas: Angers, Brest, Caen, Dijon, Grenoble, Le Havre, Lille, Limoges, Lyon, Marseille, Montpellier, Nancy, Nantes, Perpignan, Rennes, Rouen, Saint-Etienne, Strasbourg, Toulon, Tours, Troyes;
Urban communities or agglomerations: Dunkirk, Lorient, Poitiers, Pau, La Rochelle, Saint-Nazaire.

[40] The density was taken for the year 2019, the last year of the study period.

The following graph presents the results of injury rates per territory around the central municipalities of more than 50,000[41] inhabitants for the period considered from 2012 to 2019. A color code was chosen in order to clearly show the belonging of each territory to one of the 4 categories (respectively from the least accident prone to the most accident prone, dark green, pale green, blue and orange).

Consequently, it appears that the accident rate is on the one hand very heterogeneous in France according to the territory considered. The representation according to the local population density also shows that there does not seem to be an obvious correlation between the rate of injuries and the population density.

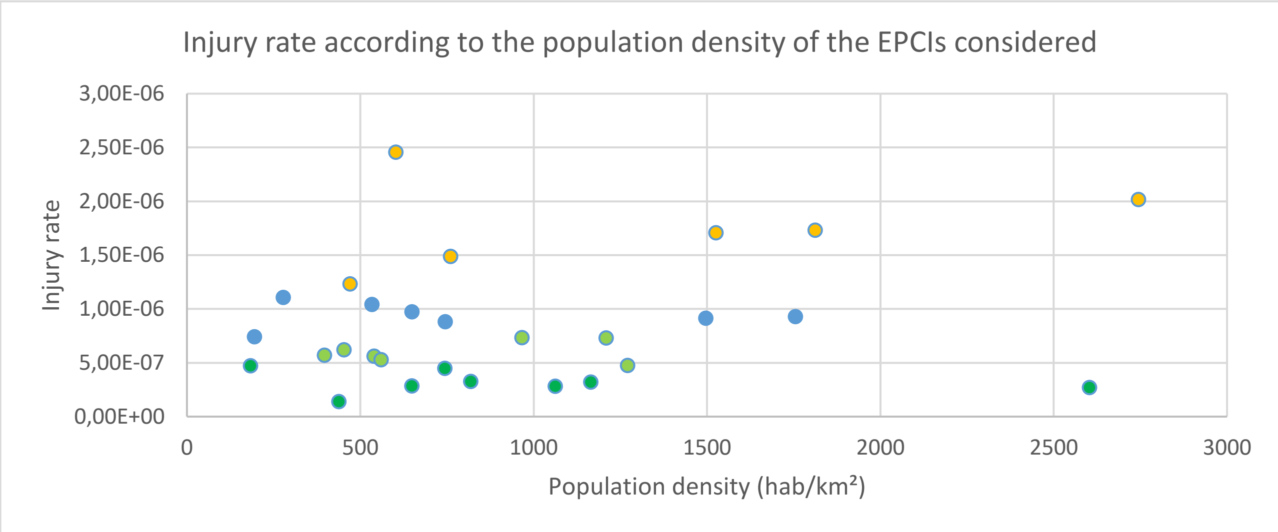


Figure 1: Graph showing injury rates by EPCI averaged over the period 2012-2019 according to the population density of each EPCI - average value at 1.00.10-6

The injury rates relating to each of the EPCIs during the period 2012-2019 are a first indicator of accidentality but are not sufficient to establish a reference level for automated public transport systems. The literature mostly uses a fatality rate to express the number of fatal accidents or even the number of deaths caused by traffic accidents involving public transport (here we are working with buses and coaches in undifferentiated way). This is why it seems important to define an overall level for safety, based on a binding reference emanating from fatal traffic accidents.

The following graph presents the results for the 27 + 3 EPCIs considered in this study for the period considered from 2012 to 2019 according to the population density of each territory. The representation of the results is made from the geometric quartiles of the series.

[41] It is useful to specify that not all EPCIs whose central municipality has a population of more than 50,000 inhabitants are studied, including Bordeaux and Toulouse in particular. However, and as introduced previously, the choice of EPCIs was conditioned on the presence or absence of data available on the kilometers travelled by public transport, reported in urban traffic surveys. The availability of data is conditional on the one hand on the completion of the survey by the operator(s) of the network in question and on the other hand on the specific completion of the kilometric data relating to high-level service buses (BHNS) and buses.

Once again, the graphical representation reflects a great heterogeneity in mortality rates, which still does not seem to be correlated with the local population density.

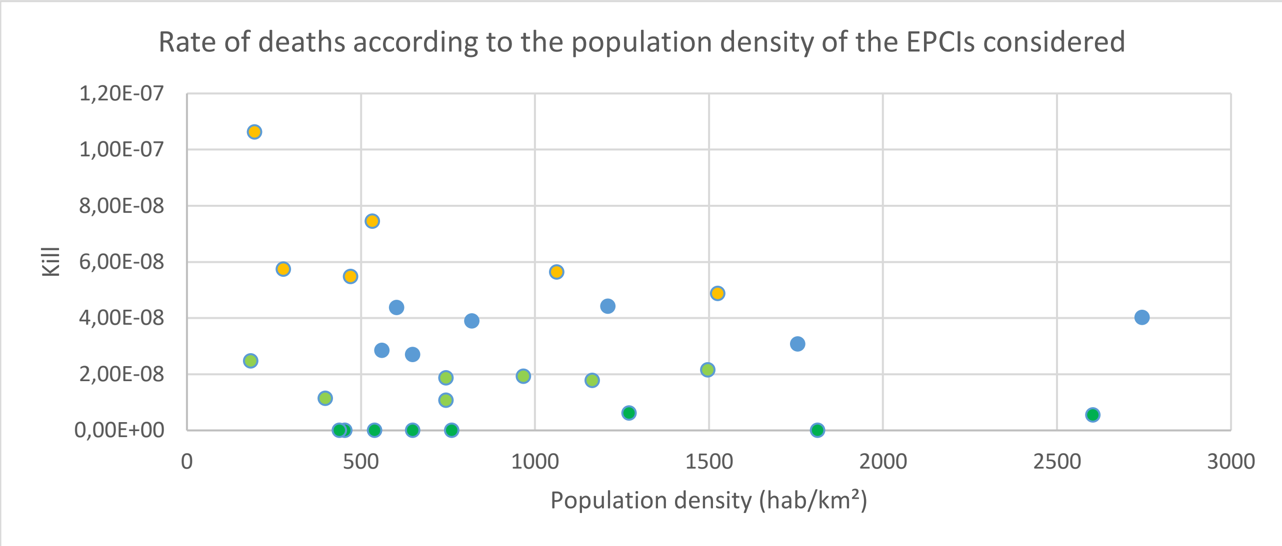


Figure 2: Graph showing mortality rates by EPCI according to the population density of each over the period 2012 - 2019 - average value at 2.91.10-8

This graph takes into account all the results obtained for the 27 + 3 EPCIs considered, including the potentially zero values. In fact, these values correspond to EPCIs in which no fatal accidents and therefore no deaths were recorded between 2012 and 2019. This is the case for 6 EPCIs, i.e. 20%, which partly distorts the classification obtained. More generally, the presence of 20% of null values can impact the establishment of an overall security objective. In addition, it was deemed relevant not to omit these EPCIs and to take them into account in an overall analysis (detailed below) insofar as these EPCIs are representative of France and its agglomerations.

Conversely, some serious accidents that occurred during the period studied contributed to an upward leveling of the average mortality rate. In some cases, the rate obtained is high and above the average due to the low number of kilometers travelled by public transport.

As part of the work on demonstrating the safety of ARTSs and in particular of automated public transport systems (APTSs), it is necessary to have a reference value that is common to the entire public transport use case. For the moment and without additional study, it has been chosen to work with an average of the mortality rates (possibly with a value linked to the rate of injuries in addition) on all the values obtained on each of the EPCIs over the period of 'study.

Table 8: Average mortality rates for the 27 + 3 EPCIs for the 2012-2019 study period

Victims considered	**Mortality**	Wounded
Average rates	**$2.92.10^{-8}$**	$1.00.10^{-6}$

This reference value resulting from the analysis presented in this report can be the subject of a national reference metric in France. It is useful to specify that this value is provisional and comes from a study in which strong assumptions had to be made. More generally, the study of accidentality is not set in stone and is intended to evolve according to new data available but also taking into account feedback. The method may also need to evolve according to new input data as well as the expected purposes.

b. Accident and Fatal Accident Rates

This second paragraph presents the results by EPCI by keeping the accident rate indicator according to whether it resulted in fatalities, independently of the number of fatalities caused by the accident. The indicator linked to the number of accident events, and not to the number of victims, has the advantage of not taking vehicle occupancy into account. For public transport, the occupancy rate is on average higher than for private mobility.

The classification established is based, as before, on reasoning based on the geometric quartiles of the values of the series of accident and fatal accident rates. The following table shows the levels obtained.

Table 9: Classification obtained from the breakdown of accident rates and fatal accident rates of the networks studied

Quartiles	1st quartile	2nd quartile	3rd quartile	4th quartile
Fatal accident rate	]0; $7.30.10^{-9}$]	]$7.30.10^{-9}$; $2.53.10^{-8}$]	]$2.53.10^{-8}$; $4.29.10^{-8}$]	]$4.29.10^{-8}$; $1.06.10^{-7}$]
Accident rate	]0; $3.65.10^{-7}$]	]$3.65.10^{-7}$; $5.56.10^{-7}$]	]$5.56.10^{-7}$; $9.82.10^{-7}$]	]$9.82.10^{-7}$; $3.13.10^{-6}$]

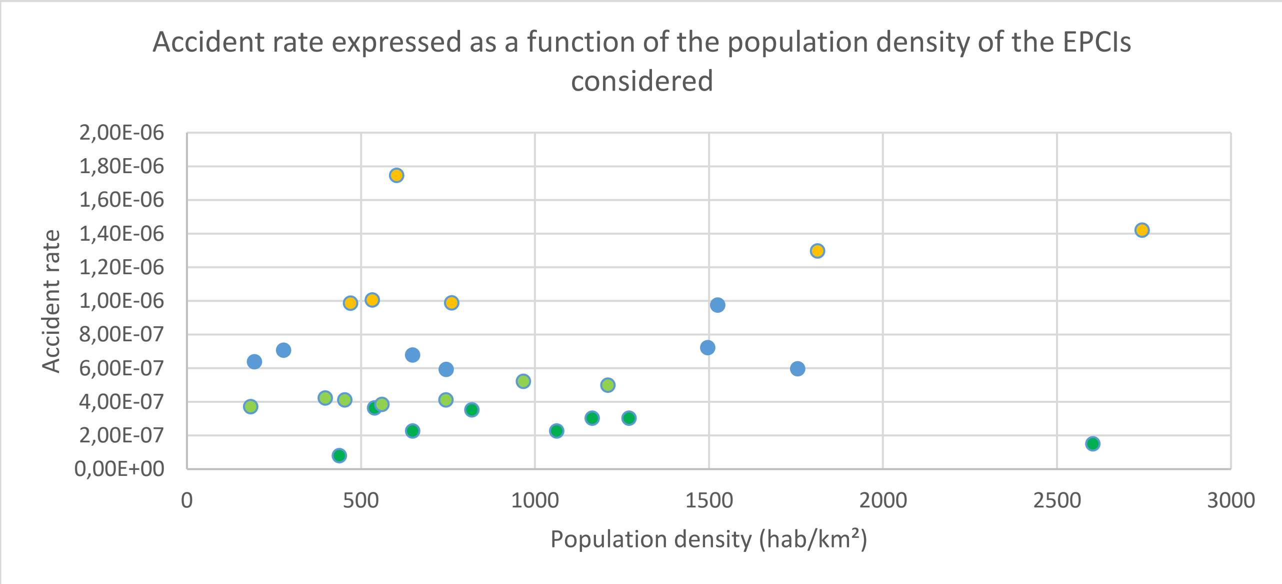
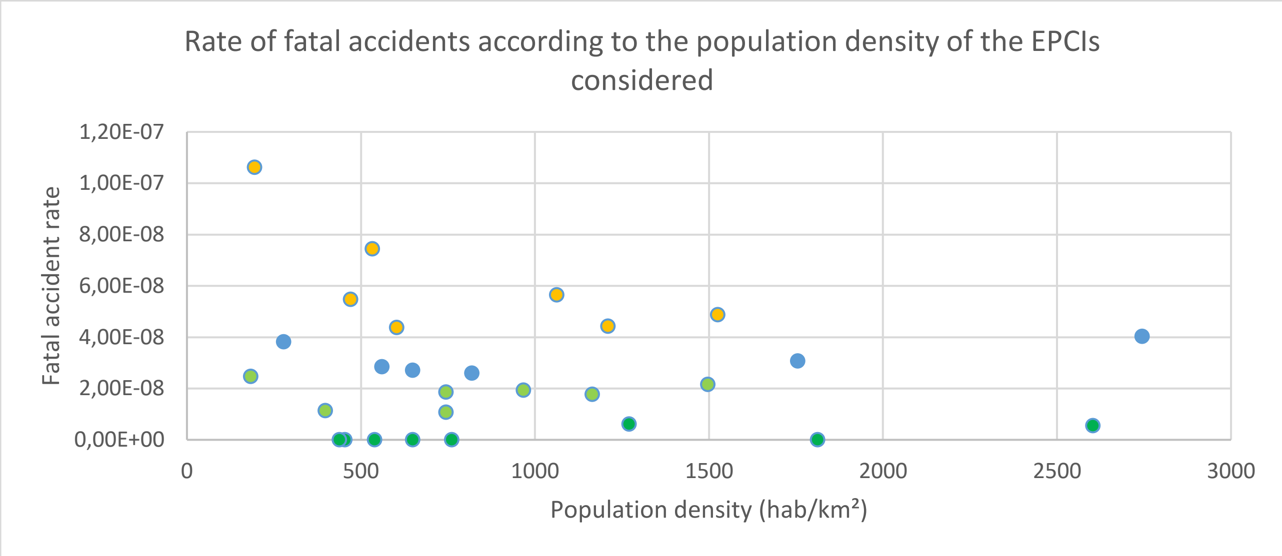


Figure 3: Graph showing accident rates by EPCI, averaged over the period 2012-2019 according to the population densities of each territory - average valuE at 7.31.10-7

The graph above shows the results of accident rates by EPCI around central municipalities with more than 50,000 inhabitants for the period considered from 2012 to 2019, according to the local population density. The representation of the results is made from the geometric quartiles of the series. By comparing this graph with that obtained for the injury rates, the classification obtained is globally similar to that obtained for the injured. As a result, we can hypothesize that the number of accidents leading to injuries follows the same trend as the number of injuries listed at the scale of French EPCIs, for the years 2012 to 2019.

Work on the number of accidents and not on the number of victims indicates in a less biased way the overall accident rate of an agglomeration by giving the rate of fatal and non-fatal accidents. On the other hand, on the question of seriousness, the indicator linked to the number of victims recorded is more explicit.

Figure 4: Graph showing fatal accident rates by EPCI, averaged over the period 2012-2019 according to the population density of each territory - average value at 2.81.10-8



The graph above shows the average fatal accident rates over 2012-2019 by EPCI, according to local population density. As with the assimilation of the indicator based on accidents to the indicator of injuries, it is quite easy to equate the indicator linked to fatal accidents to the indicator of the number of fatalities. In this case, the notion of fatal accident rates rather than the pure severity of accidents associated with the number of fatalities is much more poignant. Thus, some urban areas have a low rate of fatal accidents but a higher severity due to certain external factors not represented in the accident rate analysis by the BAAC.

Table 10: Average mortality rate and fatal accident rate for the 27 + 3 EPCIs for the 2012-2019 study period

	Mortality	**Fatal accidents**
Average rates	**2.92.10 $^{-8}$**	**2.81.10 $^{-8}$**

The preceding table presents the results obtained for the most representative accident rate indicators in order to establish a reference accident rate level. The notion of mortality is predominant in these analyses. Whether one chooses to work in relation to the frequency of occurrence of fatal accidents or in relation to the seriousness directly attributed to the accident by counting the number of fatalities, the indicator obtained remains significantly the same.

10. Limits

This last part is interested in placing this work in a more general framework on the safety demonstration and once again expressing some reservations on the use of this data. This paragraph aims to refocus the accidentology carried out within the framework of this study in a specific French context based on the national accident rate of conventional vehicles in 2022.

a. A national analysis

As indicated above, the French context of this analysis poses certain limits with a view to broadening the results obtained, particularly in the context of use by the European Commission. It is assumed that these results are taken as a national reference and are therefore the subject of support for European reflections on a reference safety level; however, it is not accepted that these results are used and decorrelated from all the hypotheses presented in this report and from any hindsight.

In this context, it is necessary and useful that other similar studies emerge within the Member States in order to supply the state of the art and to establish international guidance on the reference safety levels.

b. Feedback as a basis for development

As indicated previously, the analysis are based on the study of accidentality relating to conventional vehicles in France and do not represent the foreseeable accidentality for future automated road transport systems and in particular for automated public transport systems.

The French method deployed here does not dissociate accidents according to the presumed causes, whether it is a malfunction of a vehicle device, the inattention of a driver or an external event such as environmental conditions, the time of day by example. This type of analysis was considered in the context of preliminary work established in other working groups, but it was quickly decided that these analysis would not be representative of a reference safety level. Indeed, it is difficult and not easy to go back to the precise causes of an accident, even on the principle of a BAAC which would integrate declarative elements of the people involved. Similarly, the relatively low number of certain categories of accidents by possible cause would possibly hinder a substantial improvement in the results.

More generally, the analysis will be improved and updated. In addition, other analysis may accompany this report and its first data.

Bibliography

General Commission for Sustainable Development (CGDD). (nd). *https://www.statistiques.developpement-durable.gouv.fr/sites/default/files/2018-11/km-parcourus-en-bus-car-2016.xls.*

Legifrance. (2007). *Order of 27 March 2007 relating to the conditions for compiling statistics relating to traffic injuries.* NOR: EQUS0700567A.

Ministry of Ecology Transition, Directorate General for Infrastructure, Transport and the Sea. (2020). *National strategy for the development of automated road mobility 2020-2022.*

Ministry of the Interior, National Interministerial Road Safety Observatory. (2020). *Description of the annual databases of road traffic injury accidents - Years 2005 to 2019.*

Ministry of the Interior, National Interministerial Road Safety Observatory. (2020). *Road safety in France, Report on accident rates for the year 2019.*

Ministry of the Interior, National Interministerial Road Safety Observatory. (2017). *Guide to drafting the Traffic Accident Analysis Report.*

National Highway Traffic Safety Administration's Center for Statistics and Analysis. (2017). *2016 Fatal Motor Vehicle Crashes: Overview.* Traffic Safety Facts - Research Note.

Table of tables

Shapes table

Appendix 1: Construction of a new database

The part below specifies the part entitled 3. Construction of a new database of the Familiarization and exploitation of the data part of the report. It is a question here of briefly presenting the reasoning that was deployed for the raw processing of ONISR data from OpenData.

Objective:
Retrieve all the accidents occurring on the metropolitan territory involving public transport by distinguishing the urban and rural environment. It is obvious that the selection presented here has no vocation to remain closed, it is easily possible to adapt the code to the desired data (here public transport) or even other categories of vehicles or users.

Importing working libraries
Only *pandas* and *numpy* are required. These libraries are already available with *Anaconda,* which normally does not require any additional downloads.

Data formatting
This step only needs to be done once; when the work file is created, the direct call to the database is sufficient (in this case, go directly to the Import paragraph).
For this step, all .csv files must be in a single folder.
Files can be saved in a folder other than the .ipynb if they are used as part of several GTs or several different analyzes (which is my case). Instead, the Python file should be saved where the job and any .csv files created should be saved.
i. Reading data
It is necessary to pay attention to the types of files in particular to the separators used: they are not the same between all the files nor between each year. In the code, this is marked by the use of several different structures. For the *caracteristiques.csv* files it is also necessary to check the encoding and therefore leave a specific command for the particular encoding. For the rest, the coding is done in the American way, the decimals are marked with dots.
ii. Database formatting
As initially raised, the problem of not having a unique key by accident is problematic for doing joins between files. This is why it was chosen to create a new column using the following two columns: *Num_Acc* and *veh_num.* Thus, for each vehicle involved in an accident, a unique key makes it possible to make the joins correctly. The concatenation can then be performed on the files by joining on the two following columns.
Before starting the concatenation of all the files to create a single and homogeneous base, we must rework each of the columns because there is no homogeneity between the years (variables appear or disappear, variables change names): we do not can therefore not easily resume the basics simply without modifying the columns. It was chosen in this case to select the useful columns for the work; a fairly rough initial selection provides a new database that is exportable and usable enough for a variety of analyses.
iii. Joins
In order to follow the evolution of the base formed, the formation of the bases was done in three stages:

- Concatenation of files related to the geographical location, the general characteristics and the place of the accident (*characteristics.csv* and *places.csv*). In this case, the work is simple because each line corresponds to a different accident, we have uniqueness of the *Num_Acc key.*
- Concatenation of files related to the types of vehicles involved and to the users (*vehicles.csv* and *users.csv*). In this case, we use the previously created column (*Id_unique*). There is indeed uniqueness of the accident considered although each of the lines is not distinguishable by this column (several users per vehicle for example). The goal is to finally be able to concatenate these two DataFrame (df) without duplicating certain rows or overwriting data.
- Concatenation of the two previous df according to *Num_Acc*.

THE concatenation of all the new df obtained by year then makes it possible to obtain a single database which contains all the files for all the years selected: here from 2012 to 2019.
iv. Removal of DOM TOM

We want to work only with accidents in mainland France. The use of column *M* as indicated is not possible because for the year 2019, this column is non-existent (like many others, which prevents this data from being easily processed due to the inconsistency of the 2019 files compared to others). Dealing with 2019 separately from the other years would have been laborious and would have led to successive manipulations of the data, which on the one hand complicates the work but also statistically multiplies the potential errors on the final database. I therefore chose to work with the *dep column* which gives the department in which the accident took place (again, the coding is done on the left so the 1 is found as 10 and the 10 is found as 100 which does not is not easy to take only departments below 100 for example). I thus had to create a list in which appear all the DOM TOM (although all are not inevitably present in the bases) in order not to omit to delete accident. The command used is then an inverse request because we only want to keep those that are not DOM TOM.

USING THE .ISIN() QUERY

This query is one of the most important in this script because it is used very often with this kind of data. It makes it possible to work very quickly by avoiding going through for loops *and* would greatly increase the calculation time and would generally cause the machine to crash for dfs of more than 10,000 lines (as is the case for metropolitan accidentality over a period as big).

To work on the accident data, the reasoning is done in two stages: it is first a question of finding all the accidents which have taken place with the criterion sought (for example according to the type of vehicle or the type of user) then it is a question of recovering all the lines which correspond to the victims of these accidents. The requests are thus made in two stages:

- Find all the accidents that meet the chosen criterion. I propose for simplicity to create a variable containing all the lines in question.

Note: this step is sufficient when it comes to finding all the accidents and therefore the victims with a criterion on the infrastructure such as the type of road or the location because all the victims of the same accident have this same criterion. It is no longer sufficient when it comes to having details on the types of vehicles involved, for example where all the victims of an accident are not necessarily in a vehicle or even drivers. In this case the second step is inevitable.

- Retrieve from the initial database all the rows (therefore all the human victims) for which the *Num_Acc* is in the variable described in the previous step.

When the reverse query to *. isin()* is desired, just add a sign – or a ~ in front of the request so as to mean that it is in fact NOT *. isin()*.

v. Saving the .csv

The next step is saving the new df obtained so that you don't have to rework each time on the base files. In addition, working directly on a saved and stored file avoids inadvertently modifying the basic files and using unnecessary RAM space.

Importing the working .csv

If the *.csv file* is already created, the initial import is sufficient.

Appendix 2: Analysis based on ASFA data

This appendix presents an analysis of the data shared annually by ASFA in its “Key figures”. The particularity of this analysis is that it is based on the reports published by the ASFA every year, the accident data as well as the kilometers travelled emanate from these data and no longer from the ONISR reports. The consequence is that the results obtained with the ASFA data are comparable to those of the ONISR but are not the result of the same accident reporting processes:

- ASFA records the accidents reported by the agents who travel during an accident;
- ONISR uses the BAAC, from the reports filled out by the police who travel to the scene of the accident.

For fatal accidents, the data is normally consistent, which is less the case for injury accidents. The comparison between the results obtained must therefore be handled with caution: the orders of magnitude are comparable while the real values are less so.

The use of ASFA data also made it possible to obtain accident rates by type of vehicle circulating on the infrastructure (divided between light vehicles including motorized two-wheelers and heavy goods vehicles).

The following table presents for the period from 2012 to 2018, the number of accidents recorded on the motorway concession network as well as the proportion of bodily accidents according to the type of vehicles involved.

Year	Number of bodily accidents	Number of fatal accidents	Share of bodily accidents by type of vehicle	
			LV (including 2WD)	PL
2012	1486	127	0.84	0.16
2013	1516	138	0.81	0.19
2014	1523	123	0.93	0.07
2015	1699	148	0.84	0.16
2016	1834	147	0.86	0.14
2017	2073	150	0.86	0.14
2018	1957	140	0.87	0.13

The following table shows the traffic (kilometres travelled) on the entire network for the period from 2012 to 2018 as well as the share of traffic by type of vehicle.

Year	Total traffic (in billions of km travelled)	Traffic travelled by light vehicles (including PTW) (in billions of km travelled)	Traffic travelled by heavy vehicles (in billions of km travelled)
2012	83.7	71.9	11.8
2013	85.2	73.2	12.0
2014	87.2	74.9	12.3
2015	89.7	77.0	12.7
2016	92.6	79.4	13.2
2017	94.4	80.6	13.8
2018	95.0	80.7	14.3

The following table presents the rates calculated as part of this analysis (they may differ slightly from those presented in the ASFA balance sheets).

Year	Fatal accident rate	Injury accident rate	Light vehicles fatal accident rate	Heavy vehicles fatal accident rate	Light vehicle accident rate	Heavy vehicle accident rate
	The rates presented represent a number of deaths per billion km travelled.					
2012	1,517	17,754	1,484	1,722	17,361	20,149
2013	1,620	17,793	1,527	2,185	16,775	24.003
2014	1,411	17,466	1,527	0.700	18,910	8,667
2015	1,650	18,941	1,615	1,865	18,535	21,405
2016	1,587	19,806	1,592	1,559	19,864	19,452
2017	1,589	21,960	1,600	1,522	22,119	21,030
2018	1,474	20,600	1,509	1,273	21,098	17,791

Finally, the table below compares the fatal accident rate indicators obtained in this report and by the ASFA analysis. The rates are averaged over all the years studied.

Average accident rate by macro use case	Fatal accident rate	
Motorways and roads with motorway characteristics (ONISR)	**1.03.10** $^{-9}$	
Concessioned motorways (ASFA)	**1.55.10** $^{-9}$	
	Light vehicles (including PTW)	*Heavy vehicles*
	1,550.10 $^{-9}$	**1,546.10** $^{-9}$

In conclusion, this analysis based on ASFA data makes it possible to consolidate and validate the results presented in this report, or at least to validate the orders of magnitude of the accident rates determined by the methodology explained in this report. In addition, this analysis, by providing additional information on the accident rate of the types of vehicles on the motorways under concession, makes it possible to pose an additional hypothesis for the rest of the work: the accident rate is not dependent on the type of vehicle used, between LV and HGV for the use case of special mobility on motorways.