HARMONIZED PRE-CRASH SCENARIOS FOR REACHING GLOBAL VISION ZERO

Antonio Lara
Mexico
Jeffrey Skvarce
United States of America
Harald Feifel
Michael Wagner
Germany
Toshihisa Tengeiji
Japan

Continental AG – Division Chassis & Safety

Paper Number 19-0110

ABSTRACT

Crash data are essential for the development and introduction of new of active safety systems. At first, the target population of a new system is evaluated to understand the situations in which the system shall become active. The respective crashes are then analyzed and requirements towards the system development are derived. Finally, an effectiveness evaluation validates the potential benefit of the system in real-world crashes.

Multiple in-depth databases are available for different regions of the world. They are generally based on different crash collection and data coding methods. Thus, comparable data analysis is hard to achieve. This is however necessary for a systematic worldwide approach towards reaching "Vision Zero".

Crash scenarios describe the scene of the crash including the participants and their respective actions and intentions. They are the basis for developing sensor-based active safety systems.

The paper discusses possibilities of analyzing in-depth crash data and deriving harmonized crash scenarios. Different databases and their limitations are considered, and a scenario catalogue is proposed.

This catalogue will enable various stakeholders to compare and analyze crash scenarios of different regions and countries. The catalogue serves as a new and efficient tool to enhance the policy making for vehicles and the development of safety technology to drive "Vision Zero" worldwide.

INTRODUCTION

Crash data are needed to evaluate the benefit of safety systems in real-world crashes. The field-of-action is analyzed in which the system can become active to avoid or mitigate crashes, and essential requirements for the system development are derived. Additionally, the potential effectiveness of the safety system within its defined field-of-action is evaluated. For both development phases, a classification of crashes based on common characteristics, before and during the collision, is needed. Such common characteristics can be the trajectories of crash participants or the actual collision geometrics.

To classify traffic crashes, a set of pre-defined crash scenarios can be used. The commonly used terms scene, situation and scenario differentiate by the added level of detail. A scene describes all players and their local and dynamic properties within the surrounding environment. A situation additionally includes goals and values of the players. Besides the properties of the scene and the situation, a scenario also contains actions of the players and other decisive events [1]. Generally, a crash scenario describes the course-of-events that lead to a traffic crash, based on intentions and movements of the participants and other events and circumstances, at the scene and within the environment of the crash, and including the collision outcome. Thus, crash scenarios are well-suited for the description and definition of a safety system.

Vehicle safety systems are divided into primary, secondary and tertiary systems, with active safety systems (ADAS) addressing the primary safety by performing driver warnings and active interventions in the vehicle dynamics. This is based on a critical assessment due to ego kinematics data and object information provided by environment sensors. Therefore, a classification of crashes into crash scenarios, that are to be used for ADAS development, should be done using common sensor-relevant properties in the pre-crash phase. These are mainly the positions and movement directions of the crash participants.

Crash types describe the conflicts that lead to traffic crashes. They are generally represented by pictograms which show the first conflict between two traffic participants, regardless whether other participants are involved. Crash types are used to systematically classify and group traffic crashes. Each crash is classified by the respective crash causer and non-causer. The crash types are partly characterized by a very high level of detail [2]. Due to this segmentation they are generally unfavorable to represent the overall crash occurrence in a compact way.

This paper shows a method to cluster crash types into crash scenarios, considering characteristics and limitations of active safety systems. The focus shall be on the usability of the defined crash scenarios during the development of active safety systems. A scenario catalogues is proposed based on the Cyclist-AEB Testing System (CATS) [3]. The method is demonstrated using in-depth databases from four of the biggest markets worldwide (USA, Germany, China, Japan). As an example, traffic crashes between passenger cars and motorcycles are analyzed and visualized.

METHOD

Active safety systems prevent crashes by direct or indirect intervention in the longitudinal or lateral vehicle dynamics based on sensor information in the pre-crash phase. A classification of traffic crashes that is based on the crash type definition is therefore particularly suitable for defining the field-of-action for an ADAS.

For this paper, the authors used data from USA, China and Japan. Depending on the database used, the authors were able to identify variables which classify the crash configuration and are suitable for clustering. Table 1 gives an overview of the databases and the respective variables used for the scenario generation.

Table 1.
Databases used for the analysis and the respective variables for scenario generation

Database	Country	Variables	Reference
GIDAS	Germany	UTYP, UTYPA, UTYPB	[4]
FARS	USA	ACC_TYPE, PEDCTYPE, BIKECTYPE	[6]
ITARDA	Japan	SIP-code	[7]
CIDAS	China	UTYP, UTYPA, UTYPB	[5]

The GIDAS database describes the three-digit crash type UTYP for each recorded crash and classifies the two participants in the causal conflict as UTYPA and UTYPB. In general, the causing crash participant is coded as UTYPA. The exception are crashes with pedestrians, who are always coded as UTYPB regardless of the question of guilt. Based on the parameters UTYP, UTYPA and UTYPB, the crashes are clustered into crash scenarios.

In [8] the methodology to derive a scenario catalogue based on the GIDAS database has been extensively documented. Each step of the methodology is almost exactly applicable to the data found in CIDAS and ITARDA database. The authors propose a scenario mapping for the ITARDA data in Appendix 1. An example for the ITARDA data is given in Figure 1. Note that scenario C1 describes crossing scenarios from nearside, thus left-hand driving in Japan must be considered.

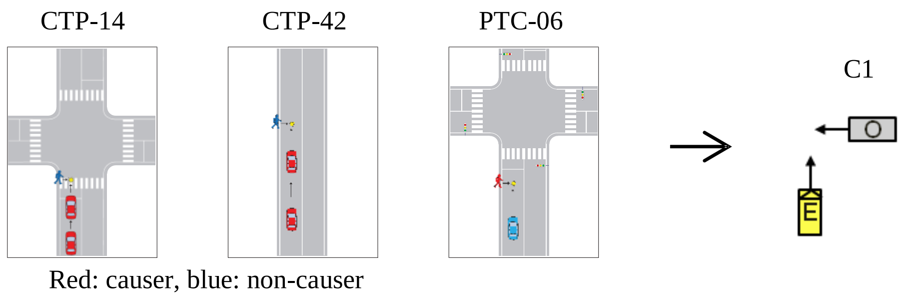


Figure 1. Example of scenario C1 using SIP code an ITARDA data

The FARS database differs in many ways from the previously mentioned databases (GIDAS, ITARADA, CIDAS). Therefore, in this paper the authors propose a methodology to derive a scenario catalogue based on the FARS database. Figure 2 describes the necessary steps of the method when using FARS. In the following text, the authors are using the terminology of FARS variable as described in Appendix 2.

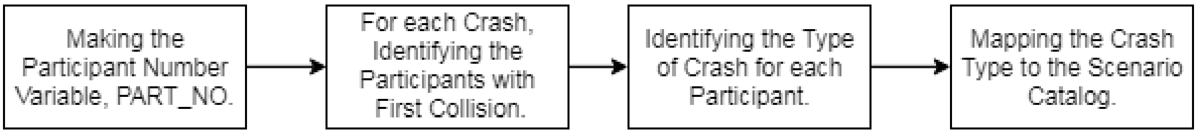


Figure 2. Generation of scenarios based on FARS data

Introduction of a new variable in FARS

The FARS database describes the type of crash at the level of the vehicle for each of the motorized vehicles. It does not contain a causal conflict and it does not contain a participant variable equivalent to the GIDAS data. A comparison of the database hierarchies of both GIDAS and FARS is visualized in Figure 3.

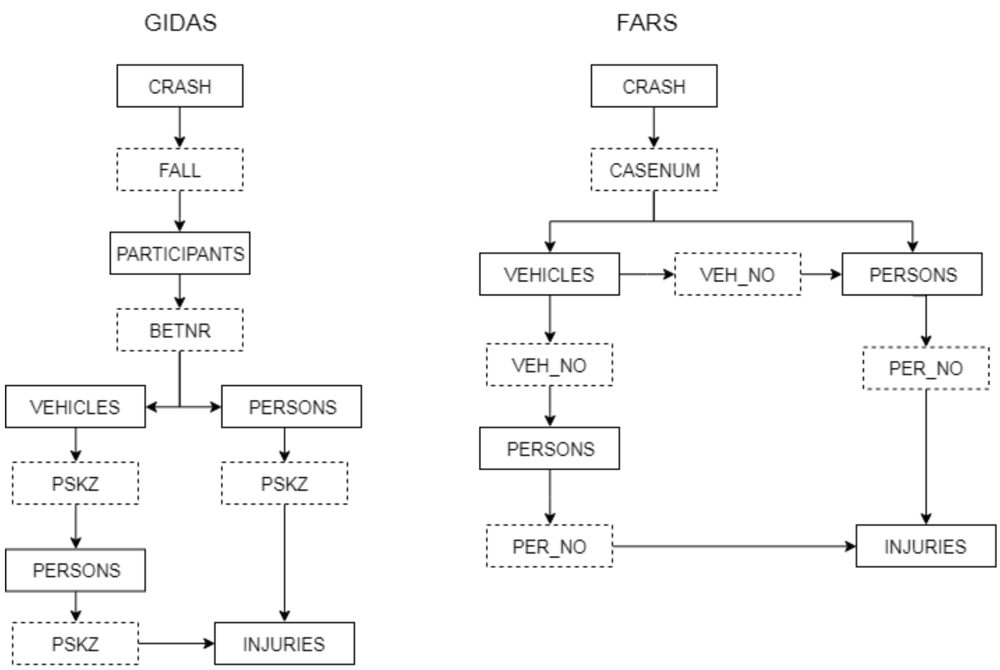


Figure 3. Comparison of GIDAS and FARS database hierarchies

For instances, in GIDAS, a conflict between a vehicle and a pedestrian is classified as a crash between two participants. In contrast, for FARS this would be considered as a single crash given that only one motorized vehicle was involved. To harmonize the data and to be able to apply the proposed scenarios found in [8], one of the primary goals was the introduction of a "participant" layer in the FARS data, which takes the number of the vehicle and the number of the pedestrian/bicyclist in a crash as an input, and then maps them into a participant number.
To this end we merge the person and vehicle data sets. Then for each state cases, the following equation is proposed, and it gives a solution to this problem.

$$Participant\ Number = \begin{cases} \text{VEH\_NO (num. of vehicle)}, & If\ \text{VEH\_NO} \neq 0 \\ max(\text{VEH\_NO}) + PER\_NO, & If\ \text{VEH\_NO} = 0 \end{cases} \quad \text{(Equation 1)}$$

For the Equation 1, suppose a State Case with n Vehicles, $\{V_1, V_2, \ldots, V_n\}$, and Persons, $\{P_j^k\}$ for k in $\{0,1,\ldots,n\}$, where P_j^k represents the Person j in the Vehicle k; if $\text{k} = 0$, the Person is a Pedestrian or Bicyclist.

It follows from the above equation that the participant number is the same for the vehicles. We do not consider the case where $\text{k}, \text{j} = 0$, since this case reduces to a conflict among motorized vehicles. For $\text{k} = 0$, the pedestrians/bicyclist case, that is $\{P_j^0\}_{1 \leq j \leq m}$, the participant number is $\text{j} + \text{n}$ for all the j, as n corresponds to the maximum number of vehicles. Hence, the total number of participants is given as follows.

$$\#\{P_j^0\}_{1 \leq j \leq m} + \#\{V_1, V_2, \ldots, V_n\} = m + n \quad \text{(if j} \neq 0\text{) (Equation 2)}$$

A visualization of the above described process can be found in Figure 4.

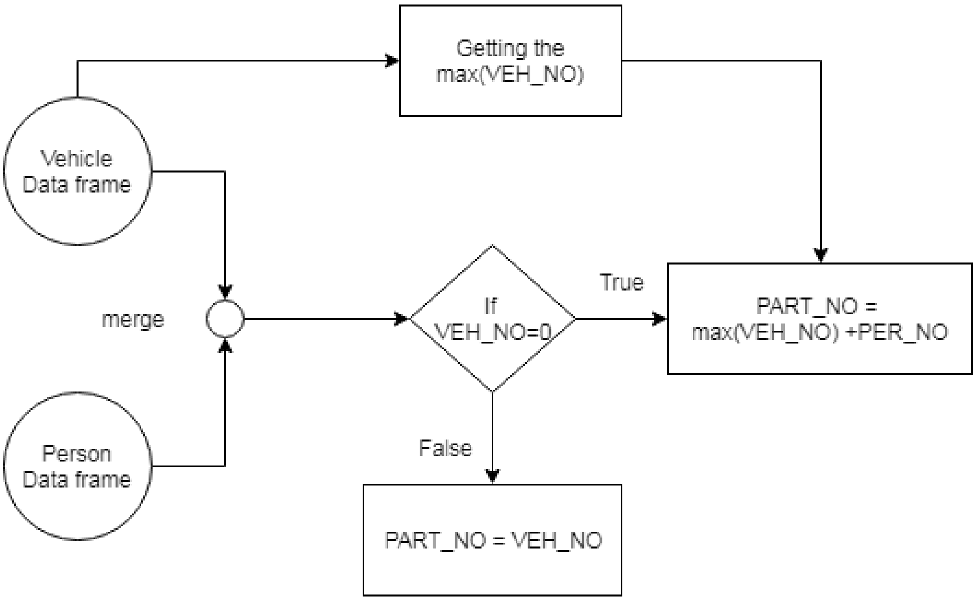


Figure 4. Generation of the participant variable from the number of vehicles and pedestrian.

Table 2 shows the above visualized process for a specific example taken from FARS 2015.

Table 2.
Example for the introduction of a new variable in FARS

ST_CASE	VEH_NO	PER_NO	PART_NO
10712	1	1	1
10712	2	1	2
10712	0	1	3 (*2+1 = max (VEH_NO)+PER_NO*)
10712	0	2	4 (2+2)
10712	0	3	5 (2+3)

Identifying the participants with the first collision

The next step is the identification of the participants with the first collision in a crash to apply the proposed scenarios. Since the authors could not identify a variable indicating the participants directly, this information was obtained using different steps. We select the first two participants in the crash using the lowest possible number in the event variable, EVENTNUM, from the VEVENT data set.

If the first participants are vehicles, we use the VNUMBER1 and VNUMBER2 variables to infer the causal conflict. The limitation with this approach is that this field is only applicable when the event is a collision between two motor vehicles.

If the first participants are a vehicle and pedestrian/bicyclist, we use, depending of the case, PEDCTYPE and BIKECTYPE for the analysis. One of the limitations with this approach is the lack of information to get a complete classification, particularly on the direction of travel of the pedestrian and cyclist. Thus, if there is a doubt about who the causer of the crash could be, the authors assume that the motorized vehicle is the causer of the crash. However, this is regardless of the question of who the (legally) guilty participant of a crash is, like it can be found in GIDAS. The overall process is visualized in Figure 5. To see our assumptions for the cause conflict in the vehicle/pedestrian and vehicle/bicyclist conflict, see Table 2 and Table 3.

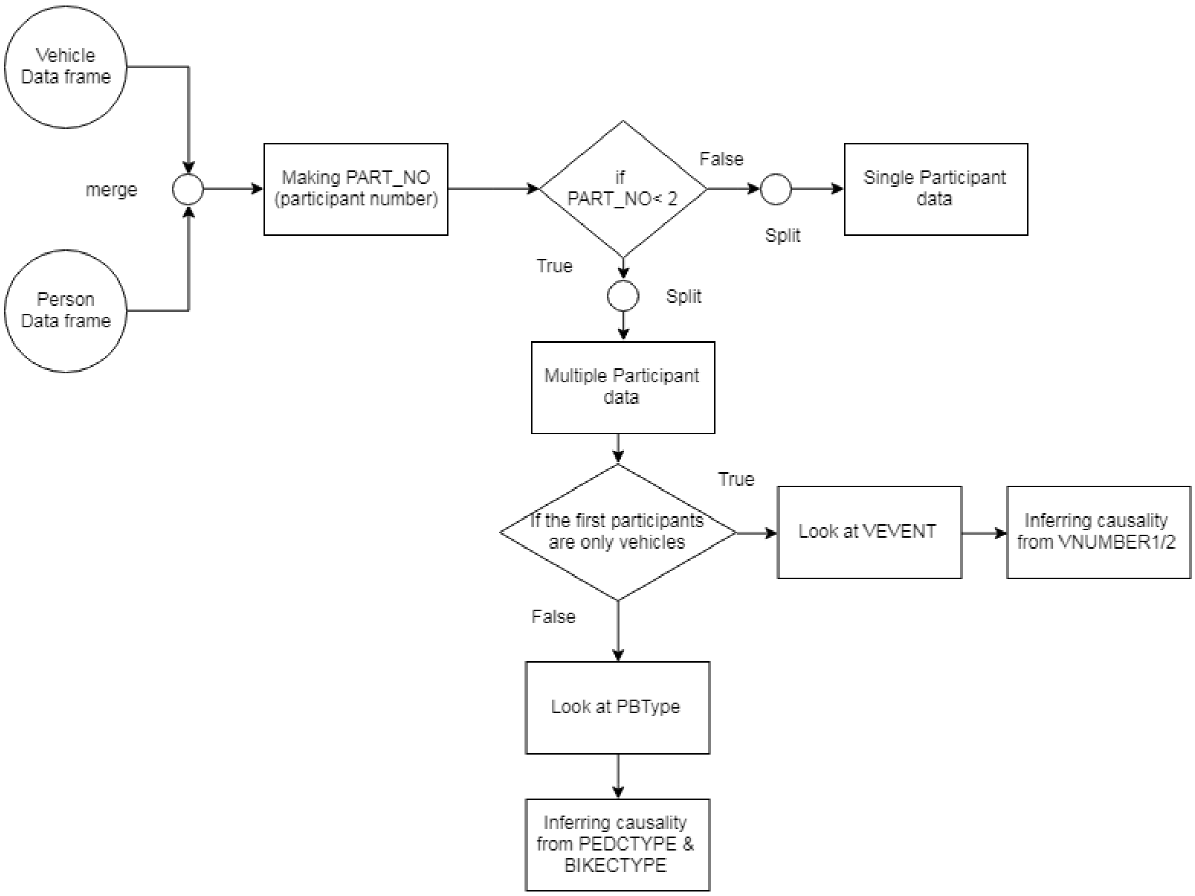


Figure 5. Identifying participants with the first collision in the crash.

Table 3.
PEDCTYPE indicates pedestrian causation in vehicle/pedestrian conflict

PEDCTYPE	**Title**	**Description**
160	Pedestrian Loss of Control	is used when the pedestrian stumbled, fell or rolled into path of a vehicle due to surface conditions, medical issue, blackout or unconsciousness, alcohol or drug impairment, falling asleep, or other mishap.
313	Lying in Roadway	is used when the pedestrian is lying in the roadway when involved with a collision with a motor vehicle. This includes someone sitting, getting up, asleep/unconscious, kneeling, etc.
742	Dart-out	is used when the pedestrian walked or ran into the roadway and was involved in a collision with a vehicle where the driver's view of the pedestrian was blocked until an instant before impact. A dart-out can only occur if there is some documented visual obstruction (e.g., parked vehicle, building or vegetation).

Table 4.
BIKECTYPE indicates bicycle causation in vehicle/bicycle conflict

BIKECTYPE	**Title**	**Description**
114, 115, 116	Bicyclist Turning Error	is used when the bicyclist made a left turn/right at an intersection or a commercial driveway, cut the corner and entered the opposing traffic lane (travel lane, bike lane, paved shoulder, parking lane) occupied by the motorist.
122, 123, 124	Bicyclist Lost Control	is used when the bicyclist lost control due to mechanical problems, alcohol, drug impairment, surface condition, improper breaking, etc.
142,153, 311, 312, 313, 318, 319	Bicyclist Ride-out	is used when the bicyclist rode from a driveway access into the path of a motor vehicle
155	Bicyclist Ride-Through	is used when the case materials indicate that the motorist had the right-of-way and the bicyclist did not stop at a sign (stop or yield) or flashing light-controlled intersection.
156, 157, 159	Bicyclist Failed to Clear	is used when the bicyclist entered the intersection on green, did not clear the intersection before the signal changed for the cross-street traffic giving those operators the right-of-way, and was involved in a collision with a vehicle whose view was not obstructed by standing or stopped traffic
250	Wrong way/Wrong side	is used when the bicyclist was traveling the wrong way on a one-way roadway or on the wrong side of a two-way roadway and collided with a motor vehicle.

Applying the scenario catalogue to FARS

After these steps, we have a dataset which satisfies all the premises to apply the proposed crash scenarios. An example of the catalogue mapping is shown in Figure 6. For each of the crashes, the solid red point represents the participant to which the scenario "C1" is attributed. Notice that only one of the involved vehicles belongs to the scenario.

Passing the variables in Table 1 of the processed data through the mapping in Appendix 3 results in a list of scenarios for a comparison with other regions.

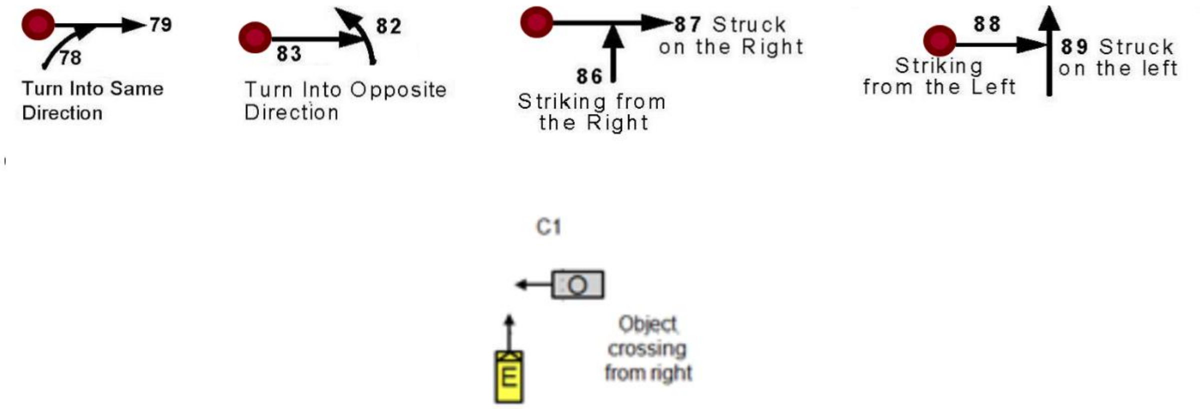


Figure 6. Mapping crash types for scenario "C1" as defined in [8]

RESULTS

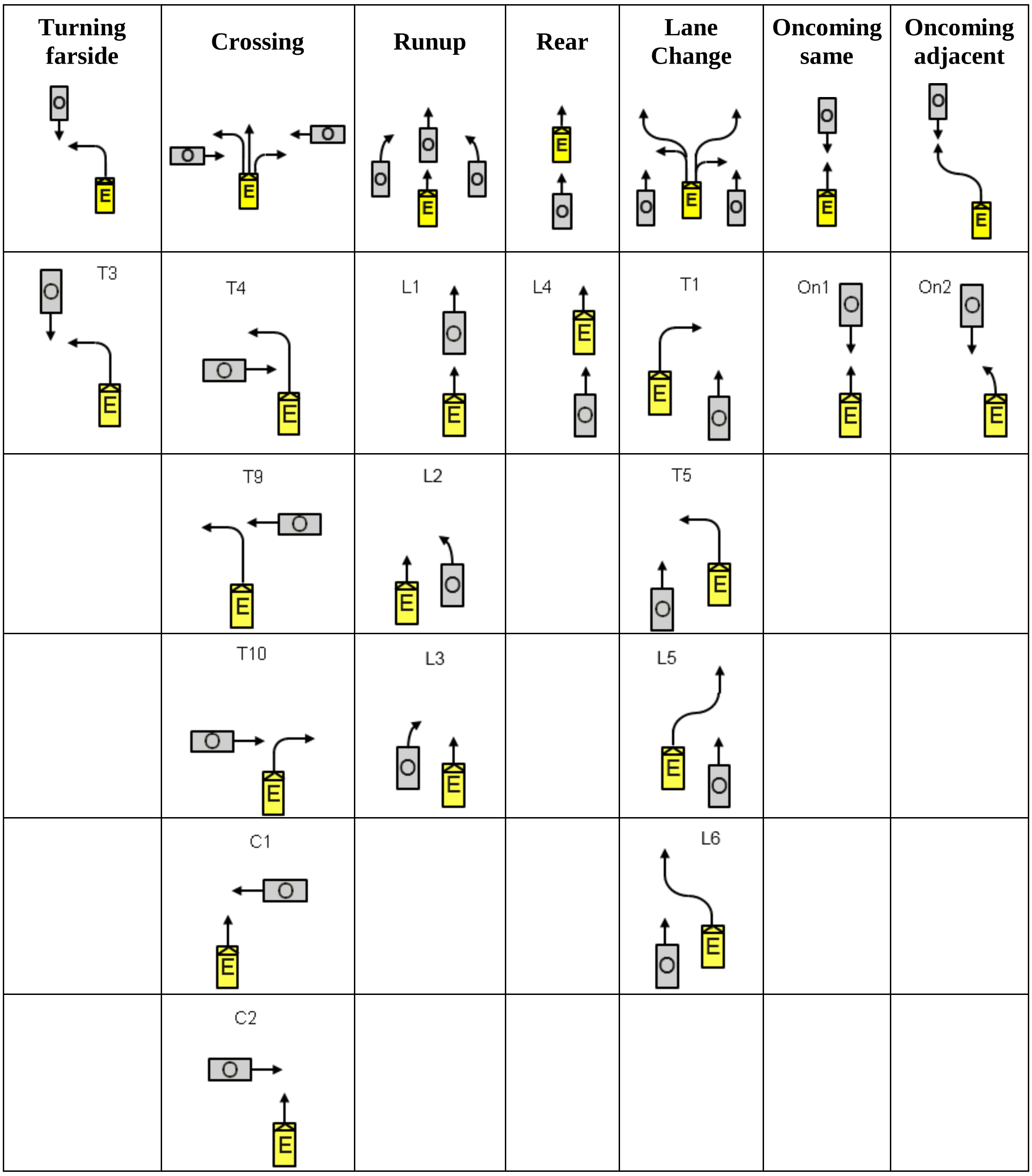
FARS describes 92 different crash types for motorized vehicles (ACC_TYPE), 56 for pedestrians (PEDCTYPE) and 78 for bicyclists (BIKECTYPE). This totals more than 220 crash types to describe the overall traffic crashes in the USA. Using a proven methodology [8] this paper shows that with slight modifications for the FARS database a mapping of nearly all FARS crash types to an existing harmonized pre-crash catalogue is possible.

This mapping allows for a reduction of the overall number of crash type categories from over 220 to 22 (-90%). For Japan this reduction is even greater since there are currently 255 SIP-Codes defined in ITARDA (-91%). As for Germany and China there are almost 300 crash types (UTYP) defined in the database, which means a reduction of about 92%. The complete overview of the mapping for the ITARDA variable SIP-Code and the FARS variable ACC_TYPE can be found in Appendix 1 and Appendix 3, respectively. The mapping for GIDAS (UTYP) has already been published in [8].

Besides reducing the number of categories, another benefit of the harmonized pre-crash scenarios can be found in the comparability of traffic crashes between countries, regions, databases, etc. The authors are not aware of any studies that have shown a practical mapping between more than two databases up to this point in time.

To demonstrate the practicability of the harmonized pre-crash scenarios, crashes between passenger cars (including light trucks) and motorcycles were analyzed across four different databases. For this example, the scenarios are clustered further into bundles to simplify the visualization and to allow a better comparability. In Table 5, the scenario bundles are displayed.

Table 5.
Scenario Bundles



From the point of view of the cars and the motorcycles, the complete accident occurrence can be presented using the proposed scenario bundles. This shows how the scenario catalogue can be applied to various participants types. In Figures 7 and 8 both, the cars and the motorcycles are depicted in the role of the ego vehicle.

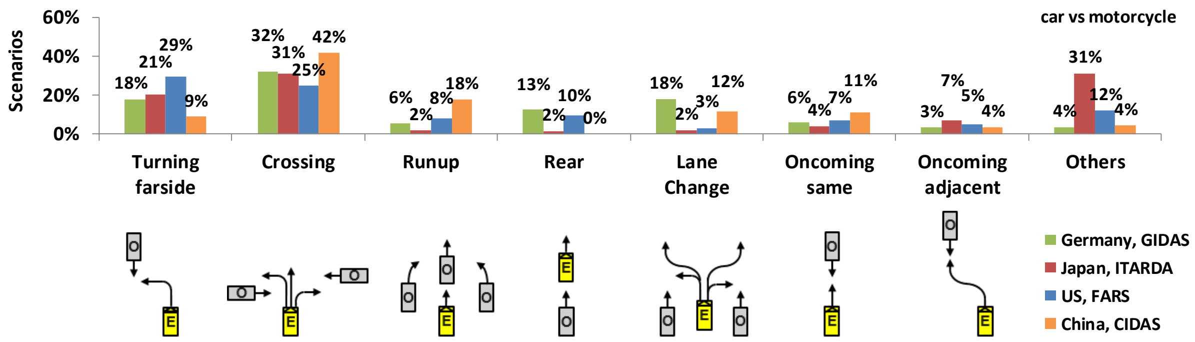


Figure 7. Car perspective: Ego is car, object is motorcycle

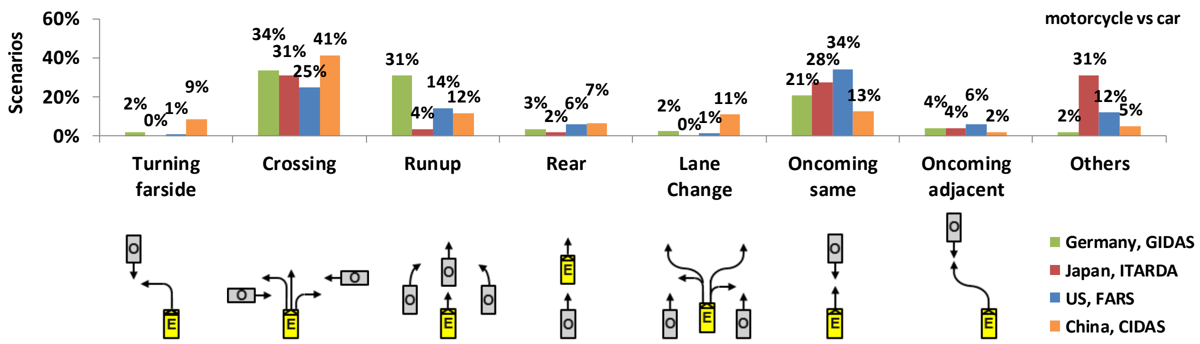


Figure 8. Motorcycle perspective: Ego is motorcycle, object is car

Appendix 4 gives an overview of all scenarios for crashes with car vs. motorcycle participation in four countries (Germany, USA, China, Japan).

At this level of the analysis of traffic scenarios it is already possible to derive the most relevant scenarios for a specific crash constellation. In the example above, “Crossing” scenarios are the most relevant in all four crash databases with “Turning farside” scenarios as second most relevant (Figure 7). Looking from the motorcycle perspective at the same crashes, “Crossing” followed by “Oncoming same” scenarios are the most relevant.

Following this high-level analysis, which can be used for identifying for example consumer test scenarios, data analysts can also take a deeper dive into the existing data to provide input for the requirements of advanced driver assistance systems (ADAS) and of advanced rider assistance systems (ARAS®).

CONCLUSIONS

The crash scenarios are consolidating for regions by combining them into a common form PCAS scenarios. Since the GIDAS and FARS coding hierarchy differs, it requires an extra effort to map the existing data upon the proposed scenarios. The limitations found within the FARS data are as follows:

i) Non-motorized participants (Pedestrian and Bicyclist) do not have a crash type in the ACC_TYPE variable.
ii) Crash type for non-motorized participants needs to be obtained through PEDCTYPE and BIKECTYPE and adjust to each of the cases.
iii) Changes in the data throughout the years (ACC_TYPE is introduced after 2010).

Nevertheless, the methodology presented in this paper shows a systematic way to deal with these limitations.

Since the FARS coding does not explicitly state if the pedestrian or cyclist was at fault (causer), we had to make the choice based on the parameter descriptions (See Table 3 and 4). Mapping results shown in Appendix 3 is incomplete because we are unable to map directly a FARS value into a PCAS scenario since some values are coded unknown (i.e. specifics unknown, specifics other). However, the authors understand that improvement in the classification as well as the processing of the data still can be made. We may improve the method by using the pre-crash variables to get further details of the critical events which lead to the collision.

Considering the challenge of properly mapping the N/A's presented in Appendix 3, more research is needed to analyze and understand these parameters. We found that our algorithm works for the years 2015 to present; however, the database has significant changes for 2010 and prior which would need further analysis.

The crash scenario catalogue presented in this paper is the result of an analytical approach of available in-depth crash data worldwide. Existing crash classifiers of different databases are used to create a harmonized dynamic scenario description. This enables a comparability of crash research results regardless of regional differences in data collection and coding formats. Therefore, harmonized safety system development and simulation methods and tools can be utilized.

The scenario generation has been demonstrated on four different crash databases. The focus of this paper lays on the US fatality database FARS. The detailed mapping of the German GIDAS crash data is explained in [8], which can also be applied to the Chinese CIDAS data. Due to its simplicity, the Japanese ITARDA SIP crash codes can directly be mapped to the proposed crash scenarios. SIP codes are defined by 255 typical accident types with more than three fatalities; these cover around 80% of fatal accidents in Japan, however, accident types with less than 3 fatalities are not considered.

OUTLOOK

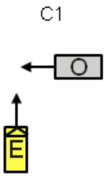
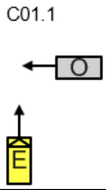
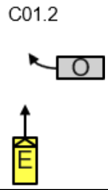
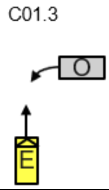
This paper describes a method to cluster crash types to a harmonized set of scenarios, by looking at each crash from the perspectives of the causer and of the non-causer and by considering additional pre-crash information. This inductive approach will naturally leave a quantity of crashes that cannot automatedly be mapped to crash scenarios, since the available classification is not sufficient, see Table 6. To further increase the respective percentages, additional research should be performed to add further available crash parameters. Ultimately, manual re-coding might be needed to reach 100% coverage, which however will be difficult to justify for existing data. It is therefore suggested that the proposed scenario catalogue is introduced as a standard crash parameter to all relevant worldwide databases and is consequently populated for all new cases.

Table 6.
Percentage of crash participants classified by automatic mapping method

Database	Region	Percentage covered
GIDAS	Germany	75%
FARS	US	70%
ITARDA	Japan	82%
CIDAS	China	75%

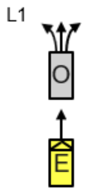
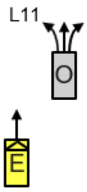
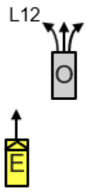
The crash scenarios allow for a dynamic crash description from the perspective of the ego vehicle. They include ego movement and object direction, however do not differentiate between possible object intentions. In V2V communication systems, the object intention is communicated over the air, therefore the crash dynamic scenarios should be extended to reflect this extra information. An extension to the scenario definition with additional object intentions is proposed. Table 7 gives an example for scenario C1 "Crossing from right" with added object intentions.

__Table 7.__
__Scenario C1 "Crossing from right" with different object intentions__

No object intention	**Object going straight**	**Object turning right**	**Object turning left**
			

The proposed scenario catalogue has been developed by aggregating crash types from different crash databases. The catalogue does not cover normal driving scenarios that are not crash relevant. To allow the classification of all real-world driving data, such as normal driving, near miss incidents and crashes, the scenarios catalogue is further extended. Therefore, non-crash relevant scenarios are added. Table 8 shows following-scenarios with traffic objects in same or adjacent lanes.

__Table 8.__
__Scenarios L1, L11, L12 "Run-up in same lane", "Following in adjacent lane"__

Run-up in same lane	**Following in adjacent lane**	
		

The proposed scenario catalogue should be applied to a maximum number of worldwide crash and naturalistic driving databases (NDD). Table 9 lists several databases that are suggested for further research.

__Table 9.__
__Possible variables for scenario generation in other databases__

Database	**Region**	**Variables**	**Reference**
RASSI	India	PRECREV, PRECRA, PRECRB	[9]
iGLAD	Worldwide	ACCTYPE, ACCTYPEA, ACCTYPEB	[10]
SHRP2 NDS	USA	Crash Type	[11]
TUAT NDS	Japan	Incident / Collision Type	[12]

REFERENCES

[1] Ulbrich, S. et.al. 2015. “Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving." In Proceedings of 18th International Conference on Intelligent Transportation Systems (Las Palmas, 2015). IEEE, pp. 982-988.

[2] German Insurance Association (GDV). Accident Types Catalogue 2016. Retrieved from URL: https://udv.de/sites/default/files/tx_udvpublications/unfalltypen-katalog_udv_web_2.pdf

[3] Op den Camp, O., et al. 2017. “Cyclist Target and Test Setup for Evaluation of Cyclist-Autonomous Emergency Braking.” International Journal of Automotive Technology, Vol. 18, No. 6, pp. 1085−1097.

[4] VUFO GmbH. 2017. “Codebook German In-Depth Accident Study.”

[5] CATARC. 2018. “Codebook China In-Depth Accident Study.”

[6] National Highway Traffic Safety Administration. 2017. “Fatality Analysis Reporting Systems (FARS) Analytical User’s Manual 1975 – 2016.” US DOT HS 812 447.

[7] Cabinet Office, Government of Japan. 2018. “Survey on Analysis Method to Estimate Traffic Fatalities Reduction Effect by Automated Driving System.”

[8] Feifel, H. and Wagner, M. 2018. “Harmonized Scenarios for the Evaluation of Active Safety Systems based on In-Depth-Accident Data.” Expert Symposium on Accident Research ESAR 2018.

[9] JP Research India PVT Ltd. 2016. Road Accident Sampling System - India (RASSI) Database Coding Manual.”

[10] IGLAD Technical Working Group. 2018. “Codebook Member Year 2018 Phase III.”

[11] Virginia Tech Transportation Institute. 2015. “SHRP2 Researcher Dictionary for Video Reduction Data.”

[12] Akagi, Y. 2017. "Ontology based collection and analysis of traffic event data for developing intelligent vehicles," IEEE 6th Global Conference on Consumer Electronics, pp.219-223.

APPENDICES

Appendix 1: Scenario mapping Japan for crashes with at least 3 fatalities (SIP code)

Code	Causer	Non-causer	Code	Causer	Non-causer	Code	Causer	Non-causer
CTC-01	C2	C1	CTC-11	On2	N/A	CTC-20	On1	On1
CTC-02	C1	C2	CTC-12	On2	On1	CTC-21	On2	On1
CTC-03	On1	T3	CTC-13	L1	L4	CTC-22	On2	On1
CTC-04	T3	On1	CTC-14	L1	L4	CTC-23	On2	On1
CTC-05	On2	On1	CTC-15	On2	On1	CTC-24	On2	On1
CTC-06	C2	C1	CTC-16	L1	L4	CTC-25	L1	L4
CTC-07	C1	C2	CTC-17	On2	On1	CTC-26	L1	L4
CTC-08	T4	C1	CTC-18	On2	On1	CTC-27	On1	On1
CTC-09	T9	C2	CTC-19	On2	On1	CTC-28	On1	On1
CTC-10	T3	On						

Code	Causer	Non-causer	Code	Causer	Non-causer	Code	Causer	Non-causer
CTM-01	C2	C1	CTM-09	T4	C1	CTM-16	On2	On1
CTM-02	C1	C2	CTM-10	T3	On1	CTM-17	L1	N/A
CTM-03	On1	On2	CTM-11	On2	On1	CTM-18	L1	L4
CTM-04	T5	L2	CTM-12	L5	L3	CTM-19	T4	N/A
CTM-05	T3	On1	CTM-13	T3	On1	CTM-20	L5	L3
CTM-06	C2	C1	CTM-14	On2	On1	CTM-21	T3	On1
CTM-07	C1	C2	CTM-15	L5	L3	CTM-22	L5	L3
CTM-08	T10	C2						

Code	Causer	Non-causer	Code	Causer	Non-causer	Code	Causer	Non-causer
CTB-01	C2	C1	CTB-11	T1	L2	CTB-20	On1	On1
CTB-02	C1	C2	CTB-12	T3	On1	CTB-21	L1	L4
CTB-03	T1	L3	CTB-13	On1	On1	CTB-22	C2	C1
CTB-04	T2	On1	CTB-14	L1	L4	CTB-23	C1	C2
CTB-05	T5	L2	CTB-15	C2	C1	CTB-24	L5	L3
CTB-06	T3	On1	CTB-16	C1	C2	CTB-25	L5	L3
CTB-07	L1	L4	CTB-17	C2	C1	CTB-26	On1	On1
CTB-08	C2	C1	CTB-18	L1	L4	CTB-27	C2	C1
CTB-09	C1	C2	CTB-19	L1	L4	CTB-28	C1	C2
CTB-10	T14	C2						

Code	Causer	Non-causer	Code	Causer	Non-causer	Code	Causer	Non-causer
CTP-01	C2	C1	CTP-18	C1	C2	CTP-35	C1	C2
CTP-02	C1	C2	CTP-19	T1	C2	CTP-36	L1	N/A
CTP-03	T2	C1	CTP-20	T5	C1	CTP-37	L5	L3
CTP-04	T1	C2	CTP-21	T3	C2	CTP-38	L6	L2
CTP-05	T5	C1	CTP-22	L1	N/A	CTP-39	C2	C1
CTP-06	T3	C2	CTP-23	T3	N/A	CTP-40	C1	C2
CTP-07	T3	C2	CTP-24	L5	L3	CTP-41	C2	C1
CTP-08	C2	C1	CTP-25	C1	C2	CTP-42	C1	C2
CTP-09	C1	C2	CTP-26	C2	C1	CTP-43	C1	C2
CTP-10	T3	C2	CTP-27	C1	C2	CTP-44	L1	N/A
CTP-11	L1	N/1	CTP-28	T3	C2	CTP-45	L5	L3
CTP-12	L5	L3	CTP-29	C1	C2	CTP-46	C1	C2
CTP-13	C2	C1	CTP-30	L1	N/A	CTP-47	L1	N/A
CTP-14	C1	C2	CTP-31	T3	N/A	CTP-48	B1	N/A
CTP-15	T5	C1	CTP-32	L1	N/A	CTP-49	B1	N/A
CTP-16	T3	C2	CTP-33	L5	L3	CTP-50	B3	C2
CTP-17	C2	C1	CTP-34	C2	C1			

Code	Causer	Non-causer	Code	Causer	Non-causer	Code	Causer	Non-causer
MTC-01	C2	C1	MTC-06	C1	C2	MTC-10	On2	On1
MTC-02	C1	C2	MTC-07	T4	C1	MTC-11	On2	On1
MTC-03	On1	T3	MTC-08	T3	On1	MTC-12	L1	L4
MTC-04	T3	On1	MTC-09	On2	On1	MTC-13	L5	L3
MTC-05	C2	C1						

Code	Causer	Non-causer	Code	Causer	Non-causer	Code	Causer	Non-causer
BTC-01	C2	C1	BTC-04	C1	C2	BTC-06	L6	L3
BTC-02	C1	C2	BTC-05	C1	C2	BTC-07	L6	L3
BTC-03	C2	C1						

Code	Causer	Non-causer	Code	Causer	Non-causer	Code	Causer	Non-causer
PTC-01	C1	C2	PTC-05	C1	C2	PTC-08	C1	C2
PTC-02	C2	C1	PTC-06	C2	C1	PTC-09	C2	C1
PTC-03	C1	C2	PTC-07	L4	L5	PTC-10	N/A	L1
PTC-04	C2	C1						

Code	Causer	Non-causer	Code	Causer	Non-causer	Code	Causer	Non-causer
HCTC-01	L1	L4	HCTC-04	L1	L4	HCTC-07	L1	L4
HCTC-02	L1	L4	HCTC-05	On2	On1	HCTC-08	On2	On1
HCTC-03	L1	L4	HCTC-06	L1	L4			

Code	Causer	Non-causer	Code	Causer	Non-causer	Code	Causer	Non-causer
HCTM-01	L1	L4						

Code	Causer	Non-causer	Code	Causer	Non-causer	Code	Causer	Non-causer
HCTP-01	C1	C2						

Appendix 2: FARS Terminology

For this paper the following data files and variables are used from FARS, year 2015 (FARS Analytical User's Manual 1975 – 2016).

Accident: This data contains information about crash characteristics and environmental conditions at the time of the crash. There is one record per crash.

> **ST_CASE**: This data element is the unique case number assigned to each crash. It appears on each data file and is used to merge information from the data files together.

Vehicle: This data file contains information describing the in-transport motor vehicles and the drivers of in-transport motor vehicle who are involved in the crash. There is one record per in-transport motor vehicle.

> **VEH_NO**: This data element is the consecutive number assigned to each vehicle in the case. It is used in conjunction with the ST_CASE data element to merge information from vehicle level data files.

> **ACC_TYPE**: Identifies the attribute that best describes the type of crash this vehicle was involved in based on the "First Harmful Event" and the pre-crash circumstances.

Person: This data file contains information describing all persons involved in the crash including motorists (i.e., drivers and passengers of in-transport motor vehicles) and non-motorists (e.g., pedestrians and pedal cyclists). There is one record per person.

> **PER_NO**: This data element is the consecutive number assigned to each person in the case (i.e., each occupant, pedestrian, or non-motorists involved in the crash). It is used in conjunction with the ST_CASE data element (and sometimes the VEH_NO data element) to merge information from person level data files.

Vevent: This data file contains the sequence of events for each in-transport motor vehicle involve in the crash.

> **VNUMBER1 & 2**: This data element identifies the "Vehicle Number" (VEH_NO) of this in-transport motor vehicle described in this event. This is the vehicle described in "Sequence of Events" for this event. If Vehicle #1 (V1) impacts Vehicle #2 (V2), then we have at least 2 Vevent records.

VEH_NO	EVENTNUM	VNUMBER1	SOE	VNUMBER2
1	1	1	12	2
2	1	1	12	2

> The explanation of these 2 records is as follows:
> V1 was involved in event 1 where V1 impacts V2
> V2 was involved in event 1 where V1 impacts V2

> **EVENTNUM**: This data element is the consecutive number assigned to each harmful and nonharmful event in a crash, in chronological order.

PBType: This data file contains information about crashes between motor vehicles and pedestrians, people on personal conveyances and bicyclists. There is one record for each pedestrian, bicyclist or person on a personal conveyance.

> **PEDCTYPE**: This data element summarizes the circumstances of the crash for this pedestrian.

> **BIKECTYPE:** This data element summarizes the circumstances of the crash for this bicyclist.

Using the above variables, the data sets, and the formula 1, we can infer the variable PART_NO (Participant number). Thus, a participant number is a number assigned to each of involved parties in a given crash (pedestrian, bicycle, car type). This shall not be mistaken by the PER_NO or the VEH_NO, however is obtained from these two.

Appendix 3: Scenario mapping proposal for USA (FARS)

Crash Type	Scenario
1	D1
2	D1
3	D1
4	N/A
5	N/A
6	D2
7	D2
8	D2
9	N/A
10	N/A
11	L1
12	O2
13	L1
14	O2*
15	N/A
16	N/A
20	L1
21	L4
22	L4
23	L4
24	L1
25	L4
26	L4
27	L4
28	L1
29	L4
30	L4
31	L4
32	N/A
33	N/A
34	D3*
35	D3*
36	D3*
37	D3*
38	D3*
39	D3*
40	D3*
41	D3*
42	N/A
43	N/A
44	L2*
45	L3*
46	L5
47	L6
48	N/A
49	N/A
50	On2
51	On1
52	N/A
53	N/A
54	D3*
55	D3*
56	D3*
57	D3*
58	D3*
59	D3*
60	D3*
61	D3*
62	N/A
63	N/A
64	On2
65	On1
66	N/A
67	N/A
68	T2/T3*
69	On1
70	T1
71	L3
72	T5
73	L2
74	N/A
75	N/A
76	T9
77	C2
78	T10
79	C1
80	T14
81	C2
82	T4
83	C1
84	N/A
85	N/A
86	C2
87	C1
88	C1
89	C2
90	N/A
91	N/A
92	B
93	L4*
98	N/A
99	N/A

Note: Scenarios with * indicate that the mapping needs to be optimized.

Appendix 4: Overview of all scenarios for crashes "Car vs. Motorcycle" in four major countries

Car vs. motorcycle crashes from the perspective of the car

Scenario	Share in %			
	Germany (KSI)	USA (K)	China (KSI)	Japan (K)
T1	0.1	0.5	6.0	0.0
T2	0.0	0.0	0.7	0.0
T3	18.1	29.5	9.0	22.8
T4	14.6	9.7	9.0	2.0
T5	9.7	1.8	3.0	0.0
T9	2.2	0.9	0.0	0.0
T10	2.2	0.9	0.7	0.5
T14	0.2	0.2	0.0	0.0
C1	6.4	4.0	13.4	10.8
C2	6.8	9.5	18.7	12.2
L1	3.4	5.8	6.7	2.7
L2	1.5	0.8	9.7	0.0
L3	1.0	1.8	1.5	0.0
L4	12.6	9.6	0.0	1.1
L5	2.6	0.6	2.2	1.6
L6	5.6	0.4	0.7	0.5
On1	5.9	7.2	11.2	9.3
On2	3.4	4.9	3.7	2.5
S1	0.0	0.0	0.0	0.0
S2	0.0	0.0	0.0	0.0
B	1.2	0.2	0.7	0.0
N/A	2.3	11.9	3.0	34.1
Total	**100**	**100**	**100**	**100**

Germany: GIDAS 2005-2018, weighted to national crash statistics

USA: FARS 2015, not weighted to national crash statistics

China: CIDAS 2014 - 2018, not weighted to national crash statistics

Japan: ITARDA 2013, not weighted to national crash statistics

Car vs. motorcycle crashes from the perspective of the motorcycle

Scenario	Share in %			
	Germany (KSI)	USA (K)	China (KSI)	Japan (K)
T1	0.1	0.0	0.7	0.0
T2	0.0	0.0	0.0	0.0
T3	2.3	1.2	8.9	2.3
T4	1.2	0.6	3.0	0.0
T5	0.6	0.1	6.7	0.0
T9	0.0	0.1	3.0	0.0
T10	0.2	0.1	0.0	0.0
T14	0.0	0.2	0.7	0.0
C1	24.8	19.8	25.2	14.0
C2	7.4	4.3	10.4	11.5
L1	12.9	9.6	0.0	0.0
L2	15.4	2.4	4.4	0.0
L3	2.8	2.1	7.4	2.0
L4	3.4	5.9	6.7	2.7
L5	0.9	0.6	0.7	0.0
L6	0.9	0.6	3.0	0.0
On1	20.9	34.4	12.6	31.6
On2	4.0	6.0	2.2	0.7
S1	0.0	0.0	0.0	0.0
S2	0.0	0.0	0.0	0.0
B	0.0	0.0	0.0	0.0
N/A	2.2	11.9	4.4	35.2
Total	**100**	**100**	**100**	**100**

Germany: GIDAS 2005-2018, weighted to national crash statistics

USA: FARS 2015, not weighted to national crash statistics

China: CIDAS 2014 - 2018, not weighted to national crash statistics

Japan: ITARDA 2013, not weighted to national crash statistics