Draft Prospective Safety Assessment Framework

Milestone MS 10 – WP2 – PU

This project has received funding from the European Union under grant agreement No. 101075068

Draft Prospective Safety Assessment Framework – Method

Work package 2, Milestone 10

Please refer to this report as follows:
Fahrenkrog, F., Fries, A., Das, A., Kalisvaart, S., Charoniti, E., Op den Camp, O. et al. (2024). "Draft Prospective Safety Assessment Framework - Method", Milestone Report MS10 of the Horizon Europe project V4SAFETY.

Project details:	
Project start date: Duration: Project name:	01/10/2022 36 months V4SAFETY | Vehicles and VRU Virtual eValuation of Road Safety
Coordinator:	Ir. S.H. (Sytze) Kalisvaart | TNO Integrated Vehicle Safety Automotive Campus 30, 5708 JZ Helmond
	This project has received funding from the European Union under grant agreement No. 101075068

Milestone document details:	
Version: Dissemination level: Due date: Submission date:	1.2 PU (public) 30/09/2023 16/10/2023

Lead contractor for this milestone document:
Felix Fahrenkrog– BMW Group

Report Author(s):	**Fahrenkrog, F., Fries, A., Das, A., BMW Group, Germany** **Op den Camp, O., Charoniti, E., Kalisvaart, S., TNO, the Netherlands** **Sander, D., Bourauel, B., Skudlarek, H., BASt, Germany** **Bärgman, J., Chalmers, Sweden** **Urban, M., Fraunhofer IVI, Germany** **Menzel, T., IDIADA Germany, Germany** **Glasmacher, C., ika RWTH Aachen University, Germany** **Chajmowicz, H., LAB, France** **Davidse, R. SWOV, the Netherlands** **Denk, F., TH Ingolstadt, Germany** **Lorente Mallada, J. Toyota Motor Europe, Belgium**

La Torre, F., Paliotto, A., Meocci, M., University of Florence, Italy
Wimmer, P., Virtual Vehicle, Austria
Sander, U., Ek, P., Volvo Cars, Sweden
Metzger, J., ZF, Germany
Wijnen, W., W2Economics, the Netherlands

Revision history

Date	Version	Reviewer	Description
01/08/2023	Preliminary draft 1	Felix Fahrenkrog – BMW	Structure of the document available.
16/10/2023	V1.0	Felix Fahrenkrog – BMW	Input by all partners included
26/11/2023	V1.1	Felix Fahrenkrog – BMW	Version for external consultation
22/04/2024	V1.2	Sytze Kalisvaart – TNO	Changed to a public document

Legal Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [name of the granting authority]. Neither the European Union nor the granting authority can be held responsible for them.

Table of contents

About V4SAFETY vii
Terms and definitions viii
Executive summary 1
1. Establishing the V4SAFETY Framework 3
1.1. Process of developing the V4SAFETY Framework 3
1.2. Background for the V4SAFETY Framework 4
2. Overview V4SAFETY Framework 9
2.1. Requirements for the V4SAFETY Framework 9
2.2. V4SAFETY Framework 12
3. General aspects of V4SAFETY Framework 14
3.1. Definitions 14
3.1.1. Visualization 14
3.1.2. Detailed Description 15
3.1.3. Input Output 16
3.1.4. Consequences 16
3.1.5. Examples 16
3.1.6. Q&A 16
3.2. User 16
3.2.1. Visualization 17
3.2.2. Detailed Description 19
3.2.3. Input Output 19
3.2.4. Consequences 20
3.2.5. Examples 20
3.2.6. Q&A 20
3.3. Example Assessment 20
3.3.1. Visualization 20
3.3.2. Detailed Description 21
3.3.3. Input Output 23
3.3.4. Consequences 23
3.3.5. Examples 23
3.3.6. Q&A 24
3.4. Formulate Conclusions 24
3.4.1. Visualization 24

3.4.2. Detailed Description 24
3.4.3. Input Output 24
3.4.4. Consequences 24
3.4.5. Examples 24
3.4.6. Q&A 24
4. Assessment Preparation 25
4.1. Define Evalution Scope 25
4.1.1. Visualization 25
4.1.2. Detailed Description 26
4.1.3. Input Output 31
4.1.4. Consequences 33
4.1.5. Examples 33
4.1.6. Q&A 35
4.2. Select Baseline Approach 35
4.2.1. Visualization 36
4.2.2. Detailed Description 37
4.2.3. Input Output 42
4.2.4. Consequences 42
4.2.5. Examples 42
4.2.6. Q&A 44
4.3. Prepare Data 44
4.3.1. Visualization 45
4.3.2. Detailed Description 46
4.3.3. Input Output 51
4.3.4. Consequences 52
4.3.5. Examples 52
4.3.6. Q&A 53
4.4. Select Models 53
4.4.1. Visualization 54
4.4.2. Detailed Description 54
4.4.3. Input Output 54
4.4.4. Consequences 55
4.4.5. Examples 55
4.4.6. Q&A 56
5. Simulation Execution 57
5.1. Configure Simulation 57

5.1.1. Visualization 57
5.1.2. Detailed Description 57
5.1.3. Input Output 58
5.1.4. Consequences 59
5.1.5. Examples 59
5.1.6. Q&A 59
5.2. Manage Simulation 59
5.2.1. Visualization 59
5.2.2. Detailed Description 60
5.2.3. Input Output 61
5.2.4. Consequences 61
5.2.5. Examples 61
5.2.6. Q&A 61
5.3. Simulate Baseline 61
5.3.1. Visualization 62
5.3.2. Detailed Description 62
5.3.3. Input Output 62
5.3.4. Consequences 62
5.3.5. Examples 62
5.3.6. Q&A 62
5.4. Simulate Treatment 63
5.4.1. Visualization 63
5.4.2. Detailed Description 63
5.4.3. Input Output 63
5.4.4. Consequences 63
5.4.5. Examples 63
5.4.6. Q&A 64
6. Assessment Analysis 65
6.1. Evaluate Safety Performance 65
6.1.1. Visualization 65
6.1.2. Detailed Description 66
6.1.3. Input Output 67
6.1.4. Consequences 68
6.1.5. Examples 68
6.1.6. Q&A 69
6.2. Project the Results 69

6.2.1. Visualization 69
6.2.2. Detailed Description 71
6.2.3. Input Output....73
6.2.4. Consequences 74
6.2.5. Examples....74
6.2.6. Q&A 76
6.3. Analyse Cost / Benefit.... 76
6.3.1. Visualization 76
6.3.2. Detailed Description 80
6.3.3. Input Output....80
6.3.4. Consequences 80
6.3.5. Examples....80
6.3.6. Q&A 80
7. Cross Assessment Topics 81
7.1. Conduct Validation & Verification 81
7.1.1. Visualization 81
7.1.2. Detailed Description 81
7.1.3. Input Output....85
7.1.4. Consequences 86
7.1.5. Examples....86
7.1.6. Q&A 87
7.2. Document Assessment.... 87
7.2.1. Visualization 88
7.2.2. Detailed Description 90
7.2.3. Input Output....95
7.2.4. Consequences 95
7.2.5. Examples....95
7.2.6. Q&A 95
8. Summary 96
References 97
Appendix A User Roles 1

About V4SAFETY

To set policies for road safety in the coming decades and push for Vision Zero, an accepted and reliable method for the comparison of safety measures for any type of safety measure that involves at least one motorized vehicle including CCAM measures is needed. The V4SAFETY method will deal with the safety of all road users, from vulnerable road users to vehicle occupants.

V4SAFETY will provide a prospective safety assessment framework that can handle a large variety of safety measures, ranging from in-vehicle safety measure, new vehicle types, infrastructure measures to regulations that influence road user behaviour. It includes methods to project the results in future scenarios and over EU regions for use by policy makers, authorities, and consumer organizations.

To understand differences between studies and to understand the influence of underlying data, assumptions and models, the method provides tools to characterise the influence of the contributing factors and their uncertainties. The resulting transparency and consistency in simulation-based safety assessment leads to much improved comparability and reliability of assessment conclusions.

The V4SAFETY consortium, led by TNO, consists of following international partners: BMW, BASt, Chalmers, Frauenhofer IVI, IDIADA, IKA, LAB France, SWOV, THI, TME, UNIFI, VIF, VCC, ZF, ERTICO, W2Economics. The project will run for 3 years.

Terms and definitions

Autonomous Emergency Braking (AEB)
A system that starts braking automatically if a collision is imminent and the driver is not taking any action (or is not doing so fast enough). AEB is able to detect a potential collision and activate the braking system to decelerate the vehicle with the purpose of avoiding a collision, or at least mitigating its impact.

Automated Driving System
The hardware and software that are collectively capable of performing the entire DDT on a sustained basis, regardless of whether it is limited to a specific operational design domain (ODD); this term is used specifically to describe a Level 3, 4, or 5 driving automation system.

Baseline
A set of data without the safety measure under study, to be compared when performing prospective assessments of a safety measure's performance.

Pre-crash phase
The time phase prior to the crash which ends with the contact between participants and/or objects involved in the crash.

Predictive Safety Assessment Framework
(or simply: framework) is a structured process for the use of simulation models and tools to predict the safety performance of a road safety measure.

Projection of data
An estimation of (future) changes for a population or target area based on the results of a smaller sample of input data that does not represent the population/area.

Prospective assessment
A predictive assessment of the (future) performance of given safety measures usually before their deployment.

Road Safety Measure
A measure intended to increase road safety, whether implemented as an in-vehicle safety measure, infrastructure measure, policy, or regulation.

Safety performance
The quantified capability of a road safety measure to achieve an improvement in road traffic safety in terms of fewer crashes and fewer/less serious injuries.

Safety measure
The collection of techniques, processes, and systems capable of temporarily or permanently directing, restricting, or controlling traffic participants; the expected safety benefit will be predicted in the prospective assessment.

Note: Safety solution and safety measure are used in the document synonymous. The word solution implies in this document that a measure has the potential to improve the traffic safety. Whether a solution does and does not improve the traffic solution is the matter of the assessment and its outcome.

Simulation Model
A computational model which allows the virtual evaluation of the safety solution, process, or behaviour it represents. A simulation model can also contain other simulation models.

Treatment
The specific safety solution applied during a prospective assessment. Treatment simulations provide data on the performance of the safety solution under assessment for comparison with baseline data.

Test
The use of quantitative measures to evaluate a safety solution under a set of specified conditions (test case), with reference to values that represent an acceptable outcome.

Vulnerable Road User (VRU)
Pedestrians, cyclists, and powered two wheelers (including e-scooters and mobility scooters) as well as people with disabilities, the elderly, and children.

[Further information on definitions is provided in section 3.1]

Executive summary

The draft document presents the initial version of the V4SAFETY framework for the virtual evaluation of road safety. The draft is intended to forester discussion in the consortium as well as with external stakeholders. The work on the framework will continue for the entire duration of V4SAFETY. The final deliverable will present the final framework.

After describing the process and background of the V4SAFETY framework, the framework is presented. The process of the framework definition started with the definition of requirements that should be fulfilled. This has been the basis to develop the framework (see Figure 1.1) which consist of four main topics that are divided in different topics and two cross-topics (conduct V&V and documentation), which are relevant for all process related topics. The four main topics which have been structure along the logical process (not necessarily in chronological order) are:

- V4SAFETY Framework (topics of this main topic: Definition, User & Stakeholder, Examples and Formulate Conclusions) – addresses the general aspects of the framework and is the only not directly process related main topic.
- Prepare Assessment (topics: Define Evaluation Scope, Select Baseline Approach, Prepare Data, Select Models)
- Execute Simulation (topics: Configure Simulation, Manage Simulation, Simulate Baseline, Simulate Treatment) describes the actual simulation. Configure and manage simulation describe the general process, while simulate baseline and treatment address the specific aspect of these simulations.
- Analyse Assessment (topics: Evaluate Safety Performance, Analyse, Cost / Benefit, Project the Results) covers the handling of the simulation output including following assessment steps, such as project the data to a different region or time and cost-benefit analysis).

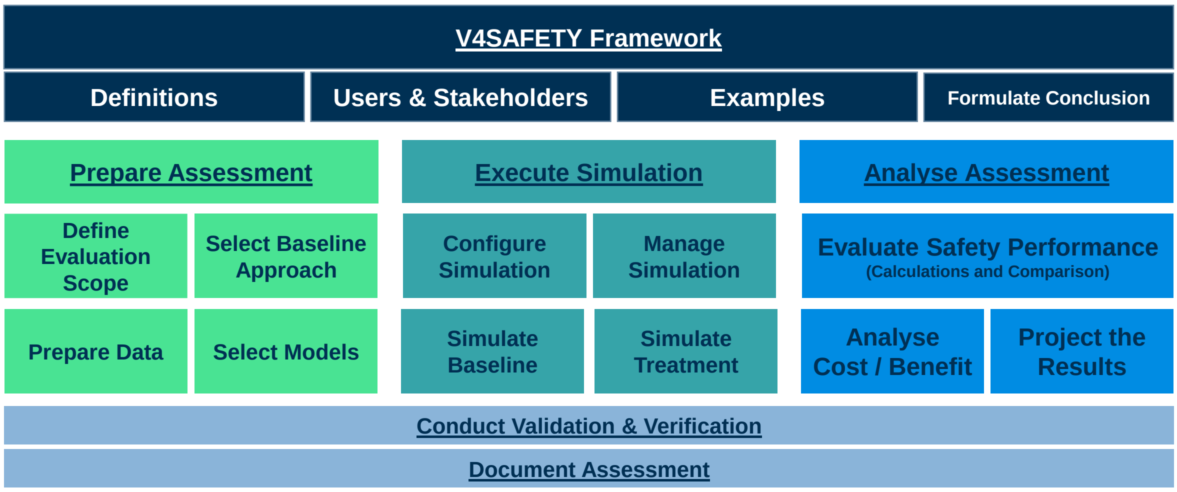


Figure 1.1: V4SAFETY Framework

All topics are described along a common structure. First, a high-level introduction of the topic is given. Then a visualization of the topic's process, a detailed description of the topic and the input to as well as the outputs to the topic is provided. In addition, to this core information in the draft a few examples are provide – as far as they are already available. For the final deliverable additional examples, a Q&A section as well as the discussion of potential consequences of certain decisions in the process will be added.

1. Establishing the V4SAFETY Framework

This chapter describes the process and the background to establish the V4SAFETY framework. The process section covers several steps – starting from the literature overview to the preparation of the detailed content of the different framework's topics. In the background section an overview about the most relevant literature for the framework is given.

The V4SAFETY framework is the cornerstone of the project. It has interactions with all content related other work packages, namely WP3 to WP7, see Figure 1.1. it requires the content input from the work packages WP3 "The human in simulation", WP4 "Modelling baseline" and WP5 "Modelling safety measures". From WP6 "Demonstration on use cases and benefits estimation" and WP7 "Communication, exploitation and stakeholder involvement" the input are rather requirements – from WP6 rather internal ones regarding the ability to execute assessments; from WP7 in terms of external expectations.

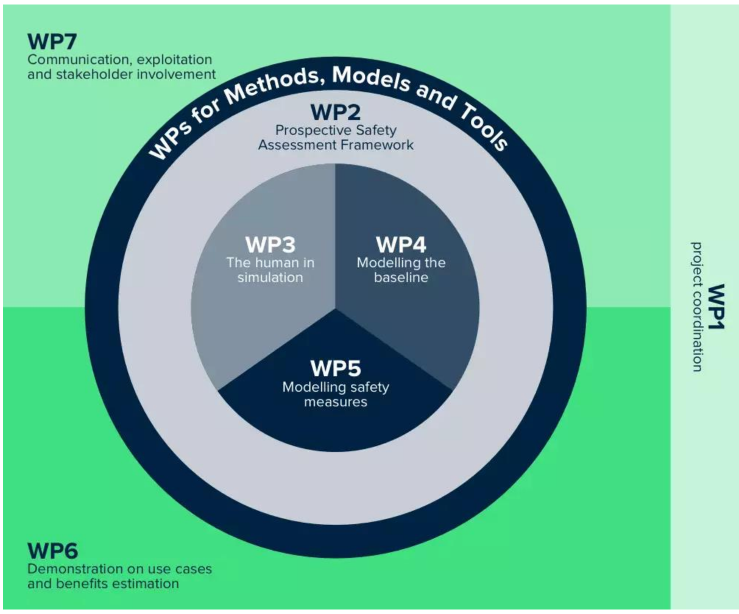


Figure 1.1: V4SAFETY workpackages and interaction of workpackages.

In terms of output, WP3 to WP5 need to consider the framework of WP2 and the requirements to it in their practical work. WP6 demonstrates the applicability of the framework in the use cases that are assessed. And WP7 covers the external communication of the framework and distribution.

1.1. Process of developing the V4SAFETY Framework

The kick-off for the V4SAFETY Framework was literature review by the work package partners. Within the review the target was to identify the most relevant source for the framework development. Therefore, each partner should nominate the five most relevant publications. These were collected and merged and build the basis for the development of the V4SAFETY Framework. The data sources are presented in the following section 1.2.

In a second initiative the requirements on the V4SAFETY framework were defined. Here all V4SAFETY partners discussed the objectives of the V4SAFETY framework. The requirements and the associated success criteria enabled to track the progress and ensure that the targets are met in the end. The definition process had been an iterative process consisting of several review and discussion rounds. The resulting requirements are presented in section 2.1.

After the preparation work was completed, the focus turned onto the actual definition of the framework. Here, starting from a review of different frameworks of other projects the topics of the framework were defined first. Once the topics, which are presented in chapters 3 to 7 were defined, the interaction between them was analysed and a logical structure of the topics were defined. The objective was to keep the framework simple and clear. In several iterations updates regarding the topics and their names were implemented. Afterwards the level of information per topic was discussed. Within the process a driving requirement was to provide information at different granularity to address several different stakeholders. Different information layers were defined, namely a short introduction, a visualization, a detailed description, input & output, consequences & limitation, examples, and a Q&A section. The common structure per topic should allow to move the framework also to a digital format.

1.2. Background for the V4SAFETY Framework

The V4SAFETY framework builds up on the knowledge of previous and ongoing projects. To most relevant documents for the work in the work package are listed in Table 1-1.In addition, the V4SAFETY Partners in work package contributed to the development of the framework based on their knowledge and experience.

Table 1-1: Background for the VSAFETY Framework

ID	Title	Project (if applicable)
1	ISO 21934 – 1 "Road vehicles — Prospective safety performance assessment of pre-crash technology by virtual simulation — Part 1: State-of-the-art and general method overview"	P.E.A.R.S. | ISO
2	ISO 21934 – 2 "Road vehicles — Prospective safety performance assessment of pre-crash technology by virtual simulation — Part 2: Guidelines for application"	P.E.A.R.S. | ISO
3	Road Safety Guidelines	N/A
4	openPASS documentation	openPASS
5	Development of a Methodology for the Evaluation of Active Safety using the Example of Preventive Pedestrian Protection	N/A
6	L3Pilot Deliverable D7.4 - Impact Evaluation Results	L3Pilot
7	New Assessment/Test Method for Automated Driving (NATM) Guidelines for Validating Automated Driving Systems (ADS)	UNECE
8	HEADSTART D.2.3 Assessment method for each of the use cases defined	HEADSTART

9	Headstart D2.1 Common methodology for test, validation and certification & D3.1 Guideline of a comprehensive validation and certification procedure to ensure safe CAD systems	HEADSTART
10	ASAM SIM:Guide Standardization for Highly Automated Driving	ASAM
11	Prospective Effectiveness Assessment of ADAS and Active Safety Systems via Virtual Simulation: A Review of the Current Practices	P.E.A.R.S.
12	Toward harmonizing prospective effectiveness assessment for road safety: Comparing tools in standard test case simulations	P.E.A.R.S.
13	The addressed VRU scenarios within PROSPECT and associated test catalogue (D3.1)	PROSPECT
14	SafetyCube - the European Road Safety Decision Support System	SafetyCube
15	SafetyCube Decision Support System (DSS)	SafetyCube
16	Crash cost estimates for European countries. Deliverable 3.2 of the H2020 project SafetyCube.	SafetyCube
17	Guidelines for priority setting between measures with practical examples, Deliverable 3.5 of the H2020 project SafetyCube.	SafetyCube
18	REGULATION (EU) 2019/2144 (General Safety regulation GSR)	EU
19	A systematic cost-benefit analysis of 29 road safety measures. Accident Analysis and Prevention, 133, 105292.	Safety Cube
20	Marktdurchdringung von Fahrzeugsicherheitssystemen 2019 (Market penetration of vehicle safety systems 2019)	BASt
21	An extrapolation method on European accident data based on weighting and data harmonization	TASC
22	A METHODOLOGY FOR BUILDING SIMULATION FILES FROM POLICE RECORDED ACCIDENT DATA (FOR ADAS EFFECTIVENESS ASSESSMENT)	PASTAS
23	PARAMETERIZATION OF STANDARD TEST SCENARIOS OF AUTOMATED VEHICLES USING ACCIDENT SIMULATION DATA	VISA
24	ANIMAL STREET CROSSING BEHAVIOR - AN IN-DEPTH FIELD STUDY FOR THE IDENTIFICATION OF ANIMAL STREET CROSSING BEHAVIOUR USING THE AIMATS-METHODOLOGY	AIMATS
25	Ein neuartiger Ansatz zur Energy Equivalent Speed (EES)-Berechnung sowie zur Stoßberechnung von Pkws mittels EES-Modellen	impacEES
26	Cost-benefit analysis of innovative automotive safety systems	VIRTUAL
27	Ten years of sustainable safety in The Netherlands: An assessment. (Transportation Research Record vol. 2213, p. 1-8.	n/a
28	Defining Reasonably Foreseeable Parameter Ranges Using Real-World Traffic Data for Scenario-Based Safety Assessment of Automated Vehicles	SAKURA (JPN)
29	Harmonized Approaches for Baseline Creation in Prospective Safety Performance Assessment of Driving Automation Systems (ESV2023; to be published)	P.E.A.R.S
30	PRYSTINE SC7 Demonstrators Use Case Definition	PRYSTINE

31	SAFE-UP D5.2 SAFETY IMPACT ASSESSMENT	SAFE-UP
32	D1.2: SHOW Use Cases	SHOW
33	AUTOPILOT D-3.7 Test data management platform architecture	AUTOPILOT
34	iRAP Methodology Fact Sheets	iRAP
35	LEVITATE Policy Support Tool (PST)	Levitate
36	Current and future accident and impact scenarios for pedestrians and cyclists	VIRTUAL
37	Sustainable Safety 3rd edition – The advanced vision for 2018-2030; Principles for design and organization of a casualty-free road traffic system	SWOV-project
38	Potenzieller gesellschaftlicher Nutzen durch zunehmende Fahrzeugautomatisierung	BASt
39	A Traffic-based Method for Safety Impact Assessment of Road Vehicle Automation	(uni-das)
40	The exiD Dataset: A Real-World Trajectory Dataset of Highly Interactive Highway Scenarios in Germany	exiD
41	Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain	Wyamo whitepaper
42	Cyclist target and test setup for evaluation of cyclist-autonomous emergency braking	CATS
43	Generation of tests for safety assessment of V2V platooning trucks	HEADSTART
44	Risk Quantification for Automated Driving Systems in Real-World Driving Scenarios	TNO StreetWise
45	Vulnerable Road User Protection	White paper
46	SAFETY OF VULNERABLE ROAD USERS	n/a
47	Improving the safety and mobility of vulnerable road users through ITS Applications [VRUITS] D2.2 assessment methodology	VRUITS
48	Microsimulation of cyclists' behavior. Evaluating the impacts of traffic demand and infrastructure design on cyclists' behavior.	MSc thesis
49	The Impact of Infrastructure Design on Cycling Safety	MSc thesis
50	Can non-crash naturalistic driving data be an alternative to crash data for use in virtual assessment of the safety performance of automated emergency braking systems?	QUADRIS (FFI - Swedish funding)
51	Counterfactual simulations applied to SHRP2 crashes: The effect of driver behavior models on safety benefit estimations of intelligent safety systems	EFRAME (FFI - Swedish funding)
52	Holistic assessment of driver assistance systems: how can systems be assessed with respect to how they impact glance behaviour and collision avoidance?	QUADRAE (FFI - Swedish funding)
53	Development and validation of a generic finite element vehicle buck model for the analysis of driver rib fractures in real life nearside oblique frontal crashes	VINNOVA funded project (FFI - Swedish)

54	Validated and Computationally Robust Active HBMs	OSCCAR
55	Virtual testing of speed reduction schemes on urban collector roads	N/A
56	Drivers' performance in response to engineering treatments at pedestrian crossings	N/A
57	Evaluation of the vehicle/safety barrier/sign support interaction by means of FEM simulations	N/A
58	Drivers' speed behaviour in real and simulated urban roads. A validation study	N/A
59	The role of infrastructure for a safe transition to automated driving	N/A
60	GIDAS-AIDED QUANTIFICATION OF THE EFFECTIVENESS OF TRAFFIC SAFETY MEASURES IN EU 27	GIDAS
61	OSCCAR D1.1: "Accident data analysis - remaining accidents and crash configurations of automated vehicles in mixed traffic"	OSCCAR
62	Preliminary Guidelines for Priority Setting Between Measures - Deliverable D3.4 of H2020 Project SafetyCube	SafetyCube
63	What travel modes do shared e-scooters displace? A review of recent research findings	N/A
64	A comprehensive analysis of the relationships between the built environment and traffic safety in the Dutch urban areas	N/A
65	A Surrogate Model-enhanced Simulation Framework for Safety Performance Assessment of Integrated Vehicle Safety Systems	-
66	A Survey on Modelling of Automotive Radar Sensors for Virtual Test and Validation of Automated Driving	InVADE (ffg.at)
67	Refining Object-Based Lidar Sensor Modeling — Challenging Ray Tracing as the Magic Bullet	Set Level
68	SAFE-UP: D2.6 USE CASE DEFINITIONS AND INITIAL SAFETY-CRITICAL SCENARIOS	SAFE-UP
69	Advanced Crash Avoidance Technologies (ACAT) Program – Final Report of the Volvo-Ford-UMTRI Project: Safety Impact Methodology for Lane Departure Warning – Method Development and Estimation of Benefit	ACAT
70	Market penetration of intersection AEB: Characterizing avoided and residual straight crossing path accidents	QUADRAE
71	An omni-directional model of injury risk in planar crashes with application for autonomous vehicles	N/A
72	Comparing motor-vehicle crash risk of EU and US vehicles	N/A
73	Human Body Model Muscle Activation Influence on Crash Response	N/A
74	Prospective safety assessment of highly automated driving functions using stochastic traffic simulation	
75	Virtual safety performance assessment for automated driving in complex urban traffic scenarios	SAVe

76	Assessment of traffic safety interventions using virtual randomized controlled trials: potential of connected and automated driving including V2X for collision reduction at urban intersections	SAVe
77	Safety Performance Assessment of Assisted and Automated Driving in Traffic: Simulation as Knowledge Synthesis.	
78	A Combined Simulation Approach to Evaluate Overtaking Behaviour on Two-Lane Two-Way Rural Roads	
79	Assessing the Impact on Road Safety of Automated Vehicles: An Infrastructure Inspection-Based Approach	CoEXist
80	Methodology for determining maximum injury potential for automated driving system evaluation	

2. Overview V4SAFETY Framework

This chapter presents the overview on V4SAFETY framework. The individual topics are presented in more detail in the following sections of this milestone document. First, the requirements on the framework are described and subsequently, the framework is presented.

2.1. Requirements for the V4SAFETY Framework

The requirements that have jointly been defined by the V4SAFETY partners are shown in Table 2-1. Overall, 28 requirements were identified that the V4SAFETY framework should address.

Table 2-1: V4SAFETY framework requirements.

No	Category	Requirement	Criteria
1	Application	The framework should be simulation-tool-agnostic.	• Description should only mention tools as examples. • Deliver overview/ example list of tools when utilization of standard(s) is recommended. • Demonstration of framework with at least two tools.
2	Application	The framework should be applicable for different road environments from e.g., urban, rural, motorway.	• Demonstration of framework for at least two different road environments (Urban and Motorway) is performed
3	Application	The framework should be able to assess different traffic participants (for instance passenger cars, HGV, PTW, cyclist, Pedestrian, E-scooter, Shuttle)	• Demonstration of framework in scenario with different traffic participants - at least covering Car to PTW, Car to pedestrian and car to cyclist conflict - is performed
4	Application	The framework should allow to consider changes in the driving scenario exposure (i.e., frequency of how often a vehicle encountered a certain driving scenarios) in the assessment.	• Demonstrate assessment of driving scenario exposure for at least two measures (one per expert assessment, one based on simulation) is performed.
5	Application	The framework should allow to consider changes in social-economic assessment exposure factors (e.g., driven milage, use form of transport).	• Demonstrate assessment of social-economic exposure for at least one measure is performed.
6	Application	The framework should be applicable on early technology descriptions/abstractions (concept phase, upper left side of V-model), but also on developed technologies (validation phase, upper right side of V-model)	• Describe of at least two different application purposes of the framework in deliverables available.

7	Application	The framework should enable projection of results to the EU and to the coming years incl. consider future changes in the traffic system, e.g., new modes of mobility, and modal shifts.	• Projection to EU for at least for one WP6 use case is performed. • Demonstration for at least one use case with potential future traffic is performed. • Projection to future years for at least for one use case in WP6 is performed.
8	Application	The framework should provide a clear guideline, template/checklist and example on how to - select, configure and connect relevant models	• Describe guidelines in relation to evaluation scope in D2.2 available. • Checklist to check the selection of models (D2.1) is available. • Description of examples in D2.1 is available.
9	Application	The framework shall be able to integrate the penetration of vehicle and/or infrastructure measures.	• Consideration of penetration rates in assessment for in least one use cases demonstrated. • Statements on penetration rates in D2.1 included.
10	Application	The framework shall provide recommendations on which baseline approach to use depending on the evaluation objective (research question)	• Description of selection process in D2.1 is available. • At least 6 examples for baseline selection provided. • Decision support (maybe as a part of D2.1) tool provided.
11	Application	The framework should allow to perform a socio-economic impact assessment.	- Social Economic assessment impact performed at least for one safety measure in WP6. • Definition of interfaces to social economic assessment available. • The documentation provided with information that are needed to perform the analysis (e.g., alignment to the SafetyCube DSS calculator for Economic Efficiency Evaluation).
12	Application	The framework shall allow to consider different types of driver models in the pre-crash phase for the relevant traffic participants and surrounding traffic.	• Demonstration with the usage of at least two different driver models is performed. • List of types of road-users models that can be demonstrated is available.
13	Baseline	The framework shall provide a clear description of difference(s) of the baseline approaches.	• Discussion and explanation of all baseline approaches differences and characteristics is available.
14	Commu-nication	The description of the framework allows non-experts (anyone who is not working with the framework on a regular basis, such as politicians, policymakers or managers) to understand it.	• V4SAFETY Deliverable D2.2 is incl. glossary is available. • Management summary-like framework description is available.
15	Documen-tation	The framework should provide a clear guideline, template/checklist and example on how to document	• Template/ checklist creation for results and correct application are available.

		- the results of safety assessments & the process, in which it is checked whether the framework is correctly used.	• Exemplary filling of templates (for at least one example) has been performed. • Clear guidelines to fill the templates are described.
16	Documen-tation	The framework should provide clear guidelines on how to document (template) -the assumptions, limitations as well as the generalizability and area of applicability of the study.	• Templates to document assumptions, limitations and generalizability are defined. • Clear guidelines to fill the templates are described.
17	Documen-tation	The framework should provide clear guidelines, template/checklist and example on how to - document all applied models incl. limitations, generalizability, data sources used in the development of models, and previous V&V activities.	• Templates to document model logic, assumptions, limitations and generalizability are defined. • Setup of clear guidelines incl. examples to fill the templates are described.
18	Documen-tation	The framework should provide a clear guideline, template/checklist and example on how to document - both implicitly (e.g., used for model development) and explicitly (e.g., the crashes used as baseline for A and B approaches) used data sources and processing of data	• Template/ checklist creation for data source reliability evaluation is available: a) Example of filled templates (for at least one example) b) Report data bias and sampling method (WP4, WP3?)
19	Model	The framework should enable the consideration of diversity of road users in pre-crash and post-crash phase, e.g., cover gender and age differences in terms of response times, modal split, helmet use, safety effect.	• Report available that describes how gender and age should be considered in the model and/or input data is available. • Demonstration the effect of gender in one use case is performed. • Demonstration the effect of age in one use case is performed.
20	Model	The framework should consider the usage of standardised interfaces for simulation models (e.g., for infrastructure) as specified in WP5	• Definition for all interfaces in the framework (WP5 input) is available. • Documentation of relevant standards for interfaces in the framework is available.
21	Output	The framework should provide a list of KPIs for measuring safety that covers the state of art and includes advantages and disadvantages.	• List of KPIs (next to primary safety related KPIs also secondary safety related KPIs) related to measuring safety fused through knowledge of partners is available.
22	Safety Measure	The framework should be able to assess all types of safety measures (including in-vehicle - incl. integrated safety -, infrastructure, behaviour and regulatory measures)	• Demonstration of framework for different types of safety measures - at least one per area of safety measures - is conducted.
23	Safety Measure	The framework should be able to consider V2X technologies.	• Demonstration of framework in (at least) one use case with V2X technology is conducted.

24	Application	The framework shall provide hints /reflections for the consideration of automated driving.	• List of consideration for the framework (mainly scenario related) regarding the assessment of automated vehicles is available.
25	Taxonomy	The framework should provide/be based on a taxonomy that is clear and concise, taking previous work (incl. standards) into account.	• Document that provides taxonomy for all utilized terms including references is available
26	V&V	The framework should be able to cover validation and verification on model level, (sub-) system level, and complete process level. Describe the concrete steps in the process of V&V.	• Short description how the entire V& V process is available. • A list of KPIs (definition) which must be looked at/ fulfilled is available. • Example argumentation on why the results are fine is available.
27	V&V	The framework should enable the inclusion of sensitivity analysis and analysis of uncertainty.	• Demonstrate at least one time the framework with a sensitivity analysis (covering different components) is conducted.
28	V&V	The framework should consider different kind of validation and verification approaches, techniques, and tools.	• Demonstration of framework for at least two different ways regarding V&V approaches, techniques and tools is conducted.

2.2. V4SAFETY Framework

The V4SAFETY Framework that is shown in Figure 2.1 consist of four main topics and two overarching assessment topics.

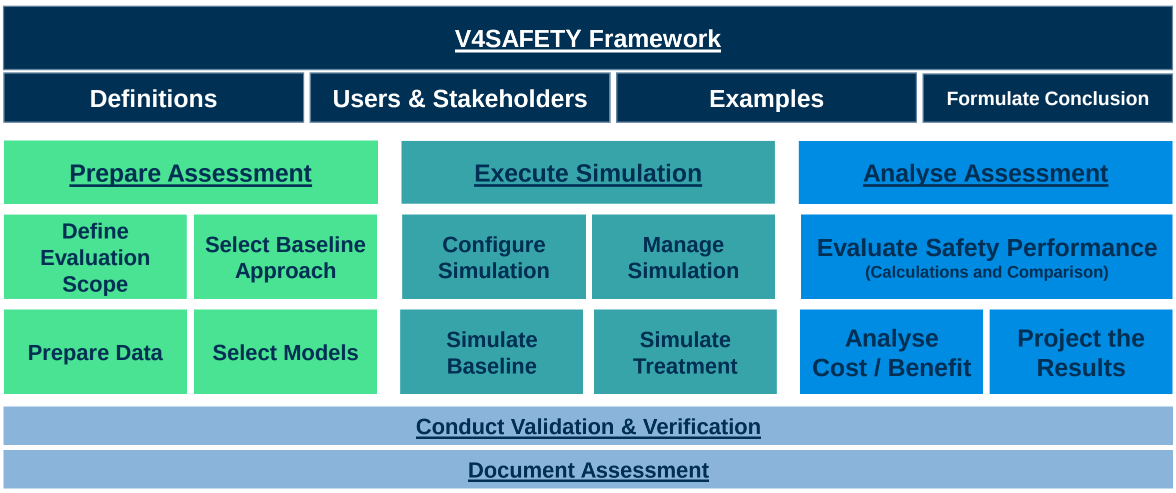


Figure 2.1: V4SAFETY Framework

The V4SAFETY Framework shown in Figure 2.1 consist of four main topics and two overarching topics. The first main topic addresses the general aspects of the framework and provides the definitions for the framework, describes the users of the framework and provides examples used within the framework. It further includes the formulation of conclusion from the application of the framework in V4SAFETY. The next three topic are following the sequence of a virtual assessment starting with the assessment preparation. This main topic covers four sub-topics: defining the evaluation scope, selecting the right baseline approach, data preparation and the model selection. Once the preparation is completed, the execution of the simulation is conducted. Here, the simulation is configured and managed. The topic baseline and treatment simulation address special aspect to be adhered when run these simulations. Once the simulations are conducted the output is analysed, which is the last main topic, the assessment analysis. Besides the evaluation of the safety performance the main topic also covers the cost-benefit analysis and projection of results. It should be noted that the relevance of these two topics depends on the evaluation scope i.e., the execution of a projection and cost benefit analysis is not necessary in every assessment. The two overarching topics which cover all other main topics deal with the documentation and the validation and verification (V&V) of the virtual assessment incl. the simulation tool and the models.

A driving aspect in the definition of the framework was the organisation of the topics in a logical order and to rather distribute them equally to prevent that assessment preparation, which probably takes most efforts among all main task, is overloaded with topics. Thus, the framework should not be seen as a strict chronological process. All topics include methodological aspects that could be relevant for the assessment preparation. The interactions between the topics are highlighted by the input and outputs that are reported in the related chapters.

3. General aspects of V4SAFETY Framework

The V4SAFETY Framework covers different general topics associated with the framework. The topics of the general aspects of the framework are Definitions, Users & Stakeholders, Examples, and Formulated Conclusion. As the title implies, the topic Definitions provides the definition of relevant terms in context of virtual evaluation of safety measures. User & Stakeholders" gives an overview about the roles and interest in the V4SAFETY framework. The topic example lists the examples that are used in the context of the framework. Emphasis is put on the use cases that are later evaluated in WP6. The last topic Formulate Conclusion discusses how to derive clear and concise conclusions from the assessment. This topic will not be addressed in detail in the draft report, but later in the final document once the V4SAFETY use case have been analysed.

3.1. Definitions

The topic "Definitions" aims at providing a clear and concise explanation of the terms used within the V4SAFETY framework. V4SAFETY does not aim to provide new definitions for terms already used in the safety assessment community but rather to provide the right level of information to the different stakeholders that will be using the V4SAFETY framework.

As starting point, since there is long experience in dealing with several terms through past research and standardization activities, appropriate sources are taken into consideration so that an original definition related to a referenced project or initiative is provided, if the term exist in previous literature/standards.

Additionally, an adapted definition is provided for terms used within V4SAFETY framework, when the original definition does not provide enough clarity to the used terms, or when no definition is available. Furthermore, to enable the relation of each term to the V4SAFETY Framework, labels are provided, so that users can understand the main topics which are relevant to the used terms. Finally, detailed descriptions as well as examples are provided when that helps to provide the right background or context to each term. When there is no adapted definition, these examples and descriptions correspond to the original definition. When an adapted definition is provided, they correspond to it.

3.1.1. Visualization

Figure 4.9 provides a summary of the process followed to list and document the definitions used in the V4SAFETY framework.

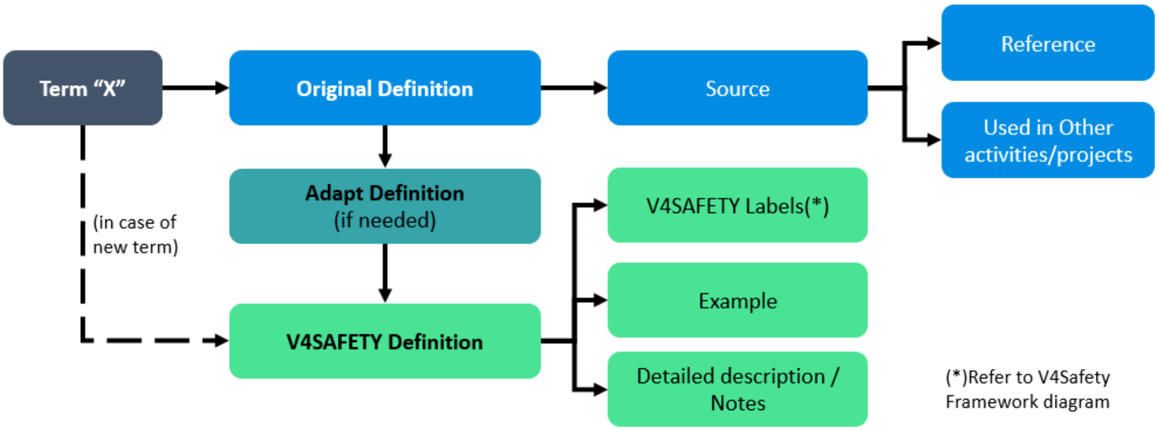


Figure 3.1: Process to list and document Definitions.

3.1.2. Detailed Description

The complete list of relevant terms identified within V4SAFETY framework, can be found under Annex X. Below, some examples are provided.

Term: Target population

- Original Definition: All situations or accidents that are addressed by the function under assessment.
- Source: ISO21934-1
- Reference: ISO/TR 21934-1:2021. Road vehicles – Prospective safety performance assessment of pre-crash technology by virtual simulation – Part 1: State of the art and general method overview
- Used in Other projects: N/A
- V4SAFETY Definition: all driving situations or accidents that are addressed by the technology under assessment.
- V4SAFETY Labels: Select Baseline Approach, Prepare Data, Evaluate Safety Performance, Project Results.
- Example: N/A
- Detailed description / Notes: terminology of function is replaced by technology.

Term: Mixed traffic environment

- Original Definition: Traffic in which participate VRUs, conventional vehicles with SAE Levels below 3 and SAE level 3 and upwards
- Source: L3Pilot
- Reference: L3Pilot Glossary
- Used in Other projects: Hi-Drive
- V4SAFETY Definition: Traffic in which participates with different SAE automation levels are operating.
- V4SAFETY Labels: Define Evaluation Scope, Select Baseline Approach, Prepare Data
- Example: N/A
- Detailed description / Notes: N/A

Term: Real-world data

- Original Definition: data collected in a non-virtual situation and environment
- Source: ISO21934-2

- Reference: ISO/AWI TS 21934-2. Road vehicles – Prospective safety performance assessment of pre-crash technology by virtual simulation – Part 2: Guidelines for application.
- Used in Other projects: N/A
- V4SAFETY Definition: data collected in a non-controlled and non-virtual situation and environment
- V4SAFETY Labels: Select Baseline approach
- Example: N/A
- Detailed description / Notes: N/A

In addition to the provided definitions, to help the reader understand how well different terms have been understood, a survey is provided in the Q&A section.

3.1.3. Input Output

Not relevant for the draft.

3.1.4. Consequences

Considering that the number of definitions is generally large when dealing with assessment of safety performance, when defining a term, there is the tendency to refer to an existing reference or source where the term has been used. This is generally a common approach, to avoid ending with duplicate definitions of terms. For example, the term baseline is defined by several sources, such as P.E.A.R.S., ISO21934-1, ISO21934-2 and within V4SAFETY proposal text.

However, even when dealing with the same topic, it is possible that a different definition is provided, something which can be the case not only on research project or activities but also at standardisation activities. To avoid a similar situation while providing the flexibility to adapt a term to a specific content, V4SAFETY aims to consider the “original” definition provided by previous activities, while providing also with an adapted version of the term to the V4SAFETY context.

Such adaptation shall make clear deviations from the original definition as well as specific needs that need to be documented, while providing a clear and concise definition of the terms. Examples may also be included when it helps understanding the user on a practical way, how such term may be applied. This may be especially the case for terms which are very specific, as for example the term, “Distribution-based pre-simulation models”, which deviates from the more generic term “pre-simulation models” and therefore requires an adaptation on what is meant by “distribution-based”, as well as a description on what is a pre-simulation model which is not distribution-based. Examples for this term, explain how such distribution-based pre-simulation models may be developed.

3.1.5. Examples

See detailed description in section 3.1.2.

3.1.6. Q&A

Not relevant for the draft.

3.2. User

The development of a comprehensive safety assessment framework involves engaging multiple stakeholders to ensure its effectiveness and successful implementation and use. In the context of our safety assessment framework, stakeholders can be categorized into direct and indirect users. The distinction between direct and indirect users lies in their level of involvement and usage. Direct users are actively engaged in its implementation and utilisation, while indirect users benefit from the framework's results and use them to make informed decisions related to traffic safety. Both groups play crucial roles in promoting safer roads and have a collective impact on the overall acceptance and usage of the safety assessment framework.

These types of users represent diverse entities, such as governmental bodies, transportation and road authorities, original equipment manufacturers (OEMs), academia, insurance companies and non-governmental organizations (NGOs). Each user role has unique expertise, resources, interests and perspectives.

Consideration of various stakeholders enhances the usability of the framework, ensures a holistic approach to traffic safety, reinforces compliance with regulations, and, ultimately, promotes a culture of safety among road users.

3.2.1. Visualization

Figure 3.2 visualizes the various users, direct and indirect ones, of the V4SAFETY framework. Many roles in different organizations have been identified and a clustering of them appeared necessary to present them in a more distinct way. In the visualization, the clustering is shown, based on some formulated larger organization groups, while under each one of these groups, the various roles are presented. It must be noted that the assignment of the roles to the groups is merely an introductory overview here, and not an exhaustive representation, as many roles can belong to more groups. However, for the sake of a concise visualization and to avoid too many repetitions, we have made a selective representation.

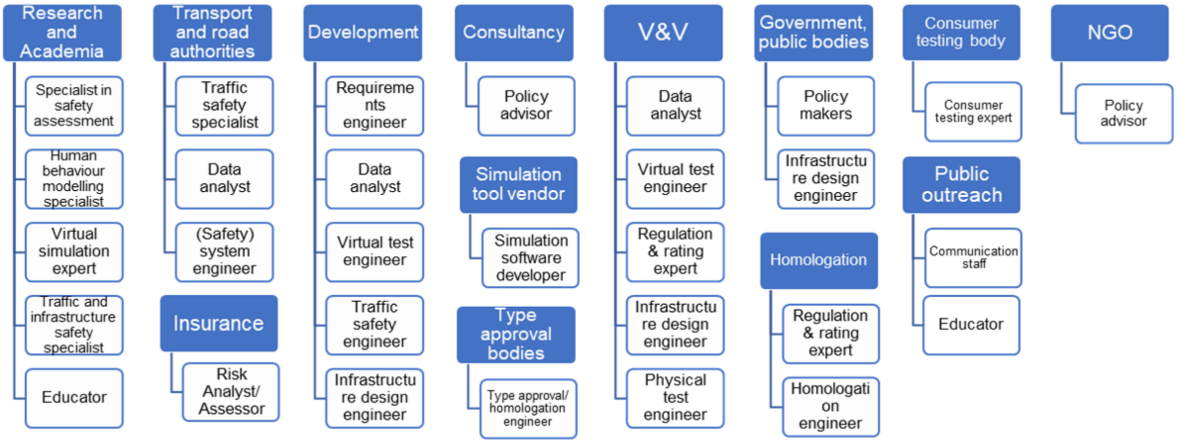


Figure 3.2: Overview of user roles.

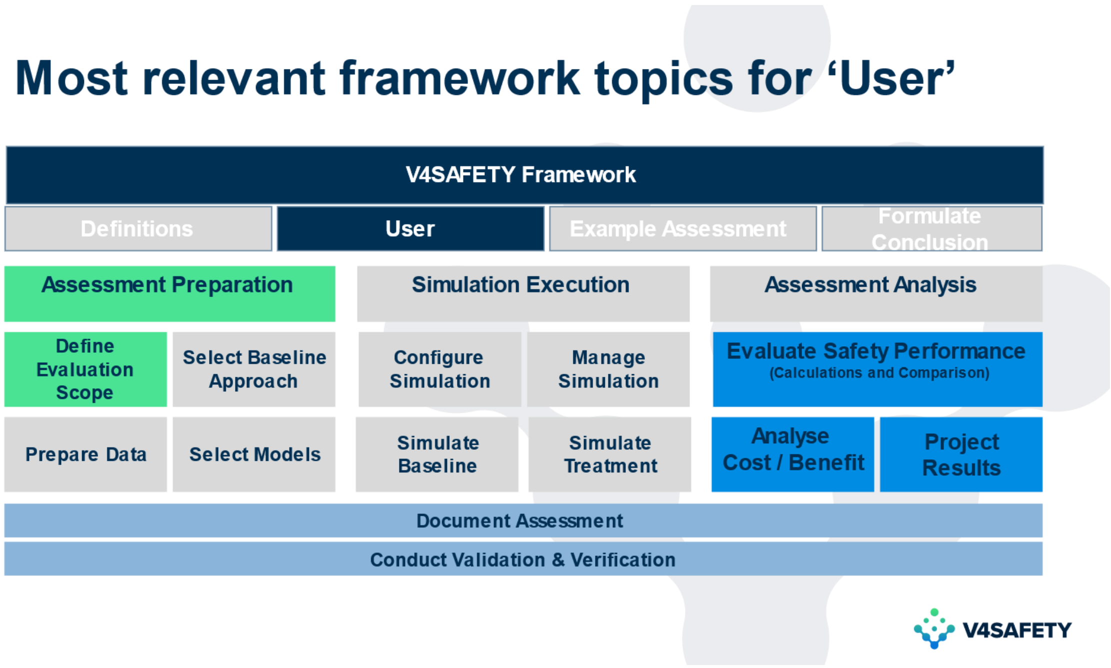


Figure 3.3: Framework aspects most influenced by user specification.

Figure 3.3 shows the parts of the V4SAFETY that will probably be most used by the various user groups. The middle three columns show the stages of the framework (Assessment preparation; Simulation execution; Assessment analysis) and the right column shows the indirect user groups that take an interest in the output of the Safety Assessment Framework. The diagram in Figure 3.4 may also be used as a reading guide for users.

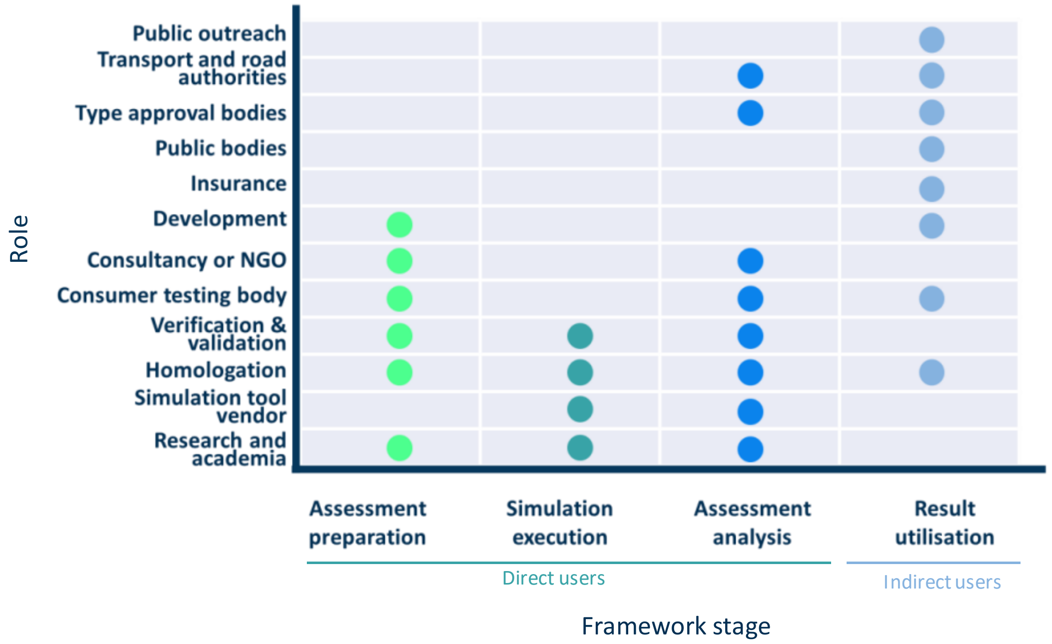


Figure 3.4: User types by framework stage.

3.2.2. Detailed Description

The following sections provide an overview of all user roles and their motivations to use the V4SAFETY framework and how they will apply the results. Additionally, success factors, bottlenecks, and key requirements are formulated for each role to provide background to the various expectations and needs. Some extra clarifications on these aspects are provided below:

- As success factors are considered the enablers that are related to the environment of the stakeholder and the application of the results.
- Bottlenecks are potential impediments that limit the use of the framework.
- Key requirements are requisites that must be addressed within the development of the V4SAFETY framework.

While specific bottlenecks can vary depending on the user group and their unique requirements, there are some generic, common bottlenecks that many users of a safety assessment framework may encounter. These are presented below:

- Local procedures might block full adoption of the V4SAFETY framework,
- Availability of models, in particular human road user models,
- Data access, ownership, quality, and availability issues,
- Lack of interdisciplinary collaboration,
- Resources and time constraints,
- Domain knowledge gaps.

Before proceeding to the detailed description per user role, some generic steps in adopting the framework have been identified. These apply to all direct users and include:

- Pilot the methodology on one project,
- Tune the methodology to internal way of working,
- Create local variants of the templates,
- Follow the V4SAFETY inspired local methodology.

An overview about all V4SAFETY relevant user roles (Specialist in safety assessment, Human behaviour modelling specialist, Virtual simulation expert, Virtual test engineer, Traffic and infrastructure safety specialist, (Safety) System engineer, Physical test engineer, Requirements Engineer, Regulation & rating expert, Type approval engineer, Consumer testing specialist, Developer using simulation, and Policy maker) for the categories “motivation”, “How will tey apply the V4SAFETY results”, “Succes factors”, “Bottlenecks” and “Key requirements” is given in Annex A.

3.2.3. Input Output

Different users interact with the framework in distinct ways and have unique requirements regarding the data they input into the system and the insights or information they expect to derive from it.

Input:
Users such as Safety Specialists contribute domain knowledge, policies, and incident data. Data Analysts provide data sources and analysis parameters. Policy Makers input high-level directives and regulatory requirements. Academics specify research questions and contribute to the development of models and method. Simulation Software Developers share simulation models. Transport Authorities input regional data. Engineers offer technical specifications. Regulators input compliance criteria. Researchers outline research objectives.

Output:

Safety Specialists seek detailed safety assessments. Data Analysts expect structured datasets and visualizations. Policy Makers receive reports for policy shaping. Academics access safety data for research. Simulation Software Developers use simulation results. Transport Authorities get jurisdiction-specific assessments. Engineers receive engineering recommendations. Regulators use compliance reports. Researchers access research-ready data and insights.

3.2.4. Consequences

Not relevant for the draft.

3.2.5. Examples

See detailed description.

3.2.6. Q&A

Not relevant for the draft.

3.3. Example Assessment

The chapter "Examples" describes the individual examples that are used to describe and to demonstrate the V4SAFETY framework. The scope of the examples is to assess their effect on the safety in terms of crashes avoidance and mitigation.
The examples for the assessment of the V4SAFETY framework are clustered into three groups. The first cluster contains in-vehicle safety measures (e.g., AEB w/o V2X, turning assist), the second cluster is about infrastructure measures (e.g., streetlights, speed bumps, etc.) and the last cluster includes regulatory and behavioural changes (e.g., traffic rules especially speed limits).

3.3.1. Visualization

Figure 4.9 provides the clusters of safety measures applied in or for traffic and a few use cases per cluster as examples for the application of the V4SAFETY framework for the safety assessment.

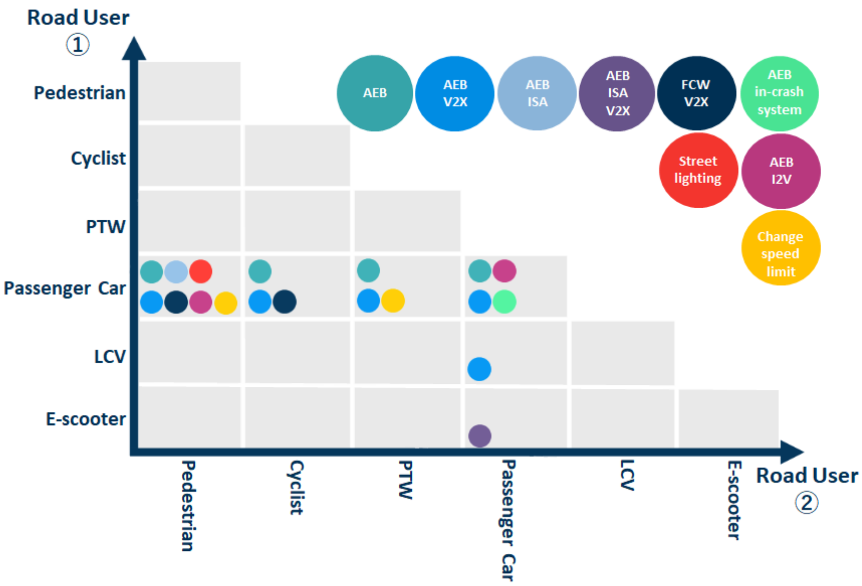


Figure 3.5: Clusters of traffic safety measures with some V4SAFETY examples per cluster.

3.3.2. Detailed Description

The V4SAFETY framework is used for the safety assessment of different safety measures. This section links the theoretical work of the framework that is described in this document with practical work of the work package 6 “Demonstration of Use Cases & Benefit Estimation”. The examples that are further explained in this section are used to describe and demonstrate the application of the framework. In addition to the WP6 examples, further examples that use to explain the framework are also described in this section.

In general, the V4SAFETY safety measures can be clustered in three categories, like in-vehicle safety measure, infrastructure, and legislation and regulation.

The first cluster named in-vehicle safety measure and is mainly driven by automotive industry (OEMs and suppliers) and / or consumer testing authorities like Euro NCAP and homologation requirements from legislation. The latter ones are less involved in the development of the technologies, but mainly responsible for a wider use of them.

This category covers all safety measures that are based on safety measure that is installed in the vehicle. The first sub-category of the safety measure considered in V4SAFETY are active safety measures. Their first objective is typically the avoidance of an imminent collision. Once this is not any longer possible the reduction of crash severity is the second objective. V4SAFETY will consider next to this active safety measures also technologies that act when the crash is a fact (passive and post-crash safety). The starting point for all simulation in the framework – independent whether an active safety and passive safety measure is considered – will be pre-crash phase (ranging from seconds to minutes prior to the crash). This poses a difference towards traditional in-crash simulations.

Some examples for such in-vehicle technologies are AEB, ACC, ISA, LDW or LKA. Also, higher automated driving – i.e., meaning system with SAE Level 3 (SAE J3016:2021) or higher – are part of the category. Some of such systems can be coupled with additional technologies like V2X communication, like a cooperative AEB (C-AEB). The passive safety measures will build up on the SAFE-UP project.

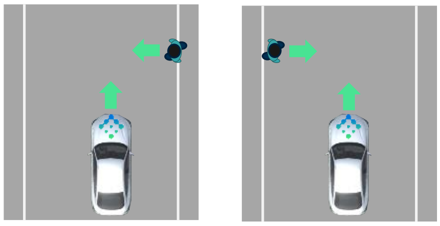
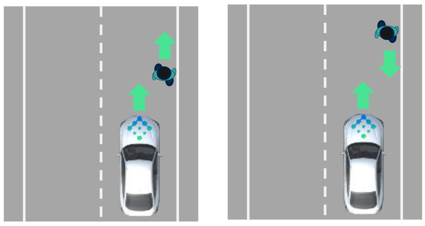
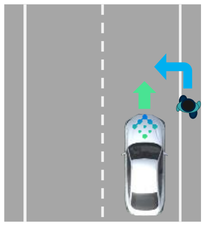
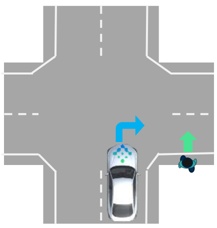
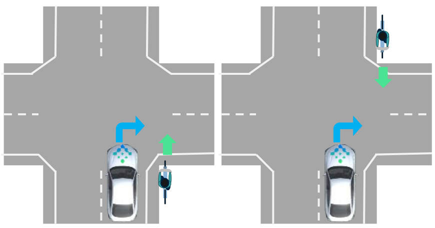
The second cluster consists of infrastructure safety measures. Usually, such safety measures cost a lot of money and effort for installation. Additionally, usually it takes a long time until such a safety measure is established in the whole target region. At the same time such safety measures have influence on a lot of traffic participants and are usually independent of any penetration rates or fleet survival curves which is the case for the upper mentioned in-vehicle safety measures.

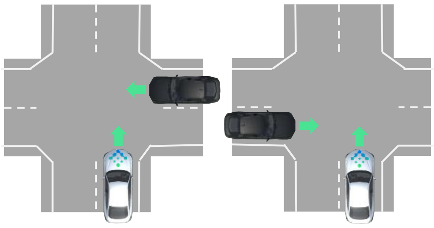
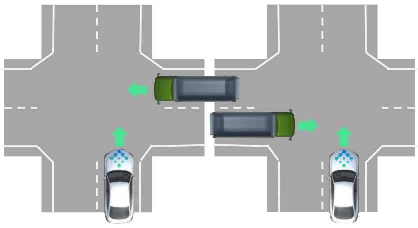
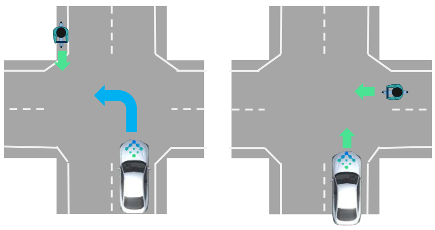
Some examples for infrastructure safety measures are speed bumps, traffic lights, digital speed limit signs, and different restraint systems like crash barriers or guardrails (reference https://safety.fhwa.dot.gov/hsip/hrrr/manual/sec47.cfm).

The third cluster is describing behavioural safety measures which are usually described by regulatory and legislation. But also other factors may have influence on the behaviour of road users like modal shifts or new technologies. Behavioural changes due to legislation depend on law enforcement and compliance.

The examples of the in-vehicle category to demonstrate the V4SAFETY framework are given in the Table 3-1.

Table 3-1: V4SAFETY Examples.

Safety Measure	Example description	Visualization of Example
In-car (AEB) Infra. (streetlight) Infra. (I2V + AEB) Behav. (speed limit)	Driving along the road with a Pedestrian crossing the road from right (EUNCAP CPNa / c) or left side (CPFa / c): The driver is driving along the road and does not recognise the intention of the pedestrian to cross the road (orthogonal to the road but also other angles are considered).	
In-car (AEB) Infra. (streetlight) Infra. (I2V + AEB) Behav. (speed limit)	Driving along the road with a Pedestrian in front going along the road in the same (EUNCAP CPLa / c) or opposite direction: The driver oversees the pedestrian on his lane going along the road.	
In-car (AEB) Infra. (streetlight) Behav. (speed limit)	Driving along the road with a pedestrian in front crossing the road, which shall be a mix of the two upper cases (jay crossing): The driver oversees the pedestrian on the road whose trajectory not only consisting of a straight line.	
In-car (AEB) Infra. (I2V + AEB)	Left turn with Pedestrian crossing the road from right to left, before left turn the pedestrian came from opposite direction: The host vehicle performs a left turn and the pedestrian is crossing the left arm of the intersection from right to the left side with respect to the host vehicle.	
In-car (AEB) Infra. (I2V + AEB)	Right turn with Pedestrian crossing the road from right to left thus the pedestrian is coming from same side as host vehicle: The host vehicle performs a right turn and the pedestrian is crossing the right arm of the intersection from the right to left side with respect to host vehicle.	
In-car (AEB)	Right turn with Bicyclist crossing the right arm of intersection: The host vehicle performs a right turn and the bicyclist is crossing the right arm of the intersection coming from same or opposite direction as host vehicle.	

In-car (AEB) Infra. (I2V + AEB)	Intersection with two passenger cars crossing the intersection straight: The host vehicle crosses the intersection straight while another passenger car from left or right side passes also the intersection straight. Both drivers don’t see or recognise the upcoming collision.	
In-car (AEB)	Intersection with host vehicle and light commercial vehicle, both crossing the intersection straight: The host vehicle crosses the intersection straight while a light commercial vehicle from left or right side passes also the intersection straight. Both drivers don’t see or recognise the upcoming collision.	
In-car (AEB) Behav. (speed limit)	Left turn with an oncoming Powered Two Wheeler (PTW): The host vehicle wants to perform a left turn. The driver overlooks the oncoming PTW. The AEB should intervene to stop the vehicle before it enters the driving path of the PTW.	
In-car (AEB) Infra. (I2V + AEB)	Left turn with passenger car crossing the intersection straight: The host vehicle performs a left turn while another passenger car passes the intersection straight. Both drivers don’t see or recognise the upcoming collision.	
In-car (AEB)	Intersection conflict with straight going or left turning host vehicle and orthogonally crossing e-scooter: The driver of host vehicle crosses the intersection straight or performs a left turn while an e-scooter crosses orthogonally its driving path straight. The upcoming collision is not recognised by both riders.	

Note: Section will be extended by adding further examples that are mentioned in the document.

3.3.3. Input Output

Not relevant for the topic.

3.3.4. Consequences

Not relevant for the draft.

3.3.5. Examples

See detailed description.

3.3.6. Q&A

Not relevant for the draft.

3.4. Formulate Conclusions

Conclusions should be formulated for all assessments. They are typically drawn and written up after the full assessment has been completed. That is, after all simulations, analysis, safety assessments, and projections and cost/benefit analysis have been completed and documented. The conclusions reflect on the results from the entire assessment, but also on the selected evaluation scope, methodological choices and decisions, assumptions and limitations that may impact results, and the subsequent conclusions. This section provides guidelines on how (typically) conclusions are written.

Note: The guidelines on formulating conclusions will be defined based on the experience with the V4SAFETY assessments. Therefore, this section will be written at the end of the project.

3.4.1. Visualization

Will be added at the end of V4SAFETY.

3.4.2. Detailed Description

Will be added at the end of V4SAFETY.

3.4.3. Input Output

Will be added at the end of V4SAFETY.

3.4.4. Consequences

Will be added at the end of V4SAFETY.

3.4.5. Examples

Will be added at the end of V4SAFETY.

3.4.6. Q&A

Will be added at the end of V4SAFETY.

4. Assessment Preparation

This main topic describes all necessary steps to prepare for the virtual evaluation of safety measures. Typically, the four steps including in this main topic are conducted at the beginning of the assessment. However, input from other main topics might be required at this stage, to enable meaningful preparation (e.g., metrics from the evaluation safety performance topic). The four steps of "Prepare Assessment" are Define evaluation scope, Select baseline approach, Prepare data and Select models. The first two categories are methodology oriented. However, the choices in these two steps, have strong implications on all the other. Prepare data and select model have - next to the theoretical aspect- also practical implications on the conduction of the simulation.

4.1. Define Evalution Scope

The Evaluation Scope utilizes the purpose (long-term goal) and objectives (short-term goal) of a safety measure effectiveness assessment to formulate one or more precise evaluation question(s). In a more research-oriented perspective, the evaluation questions may also be called research questions. However, research questions usually involve the formulation of hypothesis, which may not be relevant in a general perspective. The evaluation question(s) should take current scientific knowledge and state-of-the-art in road traffic safety into account and point out the gap that is addressed by the evaluation.
The evaluation scope should in general cover several aspects:

- Function of the system measure under test including design limitations, intended effect and market penetration,
- Area of application and addressed situation(s),
- Target region and point in time of assessment,
- Evaluation metrics

Other aspects may also be covered, if they have an impact on the selection of data and method or induce specific requirements on validation and verification of utilized simulation models or the complete framework.

4.1.1. Visualization

The Evaluation Scope can be structured in three layers covering five scope clusters (Figure 4.9).

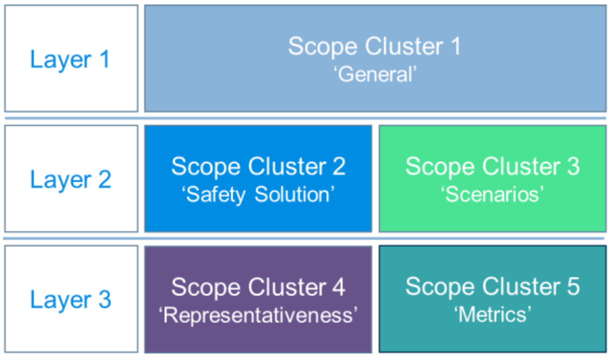


Figure 4.1: Layers and clusters of the Evaluation Scope.

Each of the clusters in the layers is defined by a set of questions that support to define the evaluation scope. The questions are related to each other, and their answers may have consequences on the answering of subsequent questions.

Figure 4.2 exemplifies such a relation, where x is the cluster, (l, m, n) are the (Q)uestion indices, and (a, b) are the alternatives to (A)nswer the question.

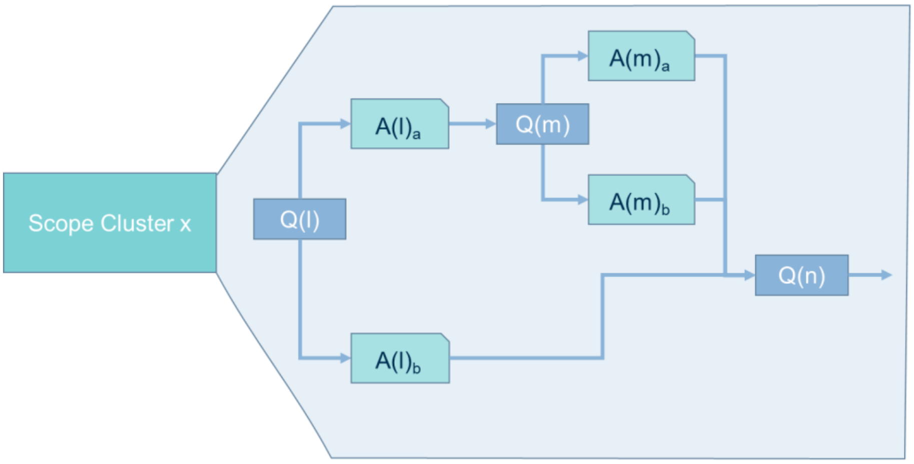


Figure 4.2: Exemplary relation of questions to each other.

The detailed description will guide through the consequences of each of the answers.

4.1.2. Detailed Description

The initial step of the assessment preparation is the definition of the Evaluation Scope. The Evaluation Scope should be based on the purpose and/or objective of the planned safety system measure assessment. Whereas the purpose sets the focus on long-term goals, the objective(s) may cover more precise short-term goals.

V4SAFETY proposes a three-layer approach, whereas each layer covers one or more Scope Clusters (Figure 4.1). A Scope Cluster summarizes questions that address similar aspects for setting the scope. By answering the questions, the user will be guided through a process that ensures the coverage of all relevant cornerstones of a scope. However, it cannot be guaranteed that the provided questions can cover the diversity of every possible scope setting. Thus, the user should always carefully review the answers to the questions to complete the scope with relevant information!

The answers to the questions are then the basis to formulation evaluation questions. In case of a scientific-oriented safety measure assessment, the evaluation question(s) is/are equivalent to research question(s). In addition to the research questions, hypothesis might be formulated that either might be confirmed or rejected based on the assessment results.

The proposed set of questions should provide a guide to the user. Figure 4.3 to Figure 4.6 show a checklist layout of these questions. They should not be answered with yes or no but formulate as short as possible and as long as necessary all aspects that explain general boundaries, the system measure and addressed scenarios, the representativeness of the effectiveness analysis and how the effectiveness should be measured.

Considering the possible variations of objectives, the formulated questions should be rather seen as a minimum set and further questions to be answered may be deemed necessary. For example, the safety measure may invoke conflicts by itself, and these are expected to have a specific magnitude. Thus, the selection of scenarios needs to be adapted to the expectations.

A final step to validate the evaluation scope and the derived evaluation questions is a review of both together with all relevant stakeholders. The stakeholders should confirm that the scope represents the objective of the assessment and that their specific needs are considered.

In case the scope becomes very extensive and, thus, unspecific, it should be considered to split the objectives in sub-objectives and conduct an evaluation scope setting of each of the sub-objectives.

4.1.2.1. First layer

<u>Scope cluster 'General'</u>

The answers to the questions may substantially influence data and method selection. The first layer describes general aspects of the evaluation (Figure 4.3).

Figure 4.3

Scope Cluster 1 'General'	
○	Is the objective to validate a concept or a product?
○	Is it necessary to evaluate false positives?
○	Is it necessary to evaluate true negatives?

Figure 4.3: Checklist for the first layer.

The *first question* identifies if the objective of the assessment is about concept or product validation. In general, concept validation allows less stringent requirements on data, model representativeness & accuracy, and documentation. On the other hand, product validation requires utilization of unbiased data, a proven representativeness and accuracy of all utilized models, and a complete documentation.

The *second* question sets requirements to the data sources. If a dataset only contains traffic conflicts and/or crashes, false positives (activations when not necessary) cannot be evaluated. The false positive rate is important if the precision (true positives / (true positives + false positives)) of a measure should be evaluated. System measure functions that have not been tested for false positives may show high effectiveness with respect to a crash-only dataset, but their application to real-world is limited because a high false-positive rate may have effects on the general effectiveness of the system (e.g. deactivation by the user due to many false activations).The false positive rate might also be tested on a different dataset that, for example, covers diverse normal driving situations.

The *third* question also sets requirements to the data sources. If a dataset only contains traffic conflicts and/or crashes, true negatives (no-activations when not necessary) cannot be evaluated, In addition to false positives the number of true negatives is necessary, if the specificity (true negatives / (false positives + true negatives)) of a system measure should be estimated.

4.1.2.2. Second layer

The second layer formulates the scope of the safety measure and the relevant scenarios (Figure 4.4).

	Scope Cluster 2 'Safety Solution'		Scope Cluster 3 'Scenarios'
○	Does your safety solution involve vehicle, infrastructure, regulatory, or behavioral aspects?	○	Are the boundaries of application defined?
○	Does the assessment involve a continuous operation solution?	○	Are the scenarios that are addressed by the safety system/function clearly identifiable/describable (selection of scenarios)?
○	Is the safety solution clearly defined (function, design limitations, intended effects)?		
○	Is the safety solution implementation rate defined?		

Figure 4.4: Checklist for the second layer.

<u>Scope cluster 'Safety Measure'</u>

The *first* question identifies if the safety measures involve vehicle, infrastructure, regulatory, or behavioural aspects. Regulatory measures or behaviour nudging address situations that change the exposure to a conflict or a crash. Thus, the utilization of situations immediately preceding a conflict or a crash without the regulative or behavioural measures may not be representative for the occurrence in real-world. For example, a speed limit will change the distribution of speeds and critical situations that occur due to high speeds may have a lower probability. Similarly, a system that warns car drivers when they are approaching a school or playground area ay change their behaviour and lower exposure to critical situations. For vehicle or infrastructure measures it needs to be evaluated if the utilized measures induce a change of exposure.

The *second* question evaluates if the assessment involves a continuous operating function. If a system measure is not continuously operating but supporting in crash-relevant conflicts, then its evaluation may also only consider the simulation of situations immediately preceding a conflict or crash. For the evaluation of continuous operating safety measures, however, situations must be considered a changed exposure to a conflict or crash. For example, an active cruise control function may increase time-headway to lead vehicles and thus, the exposure to crashes is different compared to human drivers that select a shorter time-headway.

The *third* question evaluates if the safety measure is clearly defined. Part of this description are the considered safety mechanisms including intended effects and under which conditions they are expected. Additionally, operational limitations should be stated if they are not inherently considered in the simulation set-up. Known side-effects should be considered when selecting relevant scenarios. For concept evaluations, a range of parameters instead of fixed parameters might be utilized.

The *fourth* question defines the implementation rate (market penetration) of the safety measure. It may range from greater zero to hundred percent and might be linked to the time of evaluation. For vehicle and infrastructure safety measures the implementation rates have different effects, as infrastructure measures are available for all vehicle in specific locations whereas vehicle measures are available to specific vehicles in all locations.

<u>Scope cluster 'Scenario'</u>

The *first* question incentives to formulate boundaries of the application. This means, for which type of road traffic participants the measure is relevant and if there are constraints to either the motion of conflict participants or environmental and infrastructure conditions. The more stringent the area of application can be defined the more the diversity, and thus, the number of simulated scenarios can be defined and limited. This, however, also implies that there is a good understanding of side-effects. Correspondingly, the lesser the area of application is defined, the more variance in scenarios might be considered or even just a broader area may be defined for complex areas.

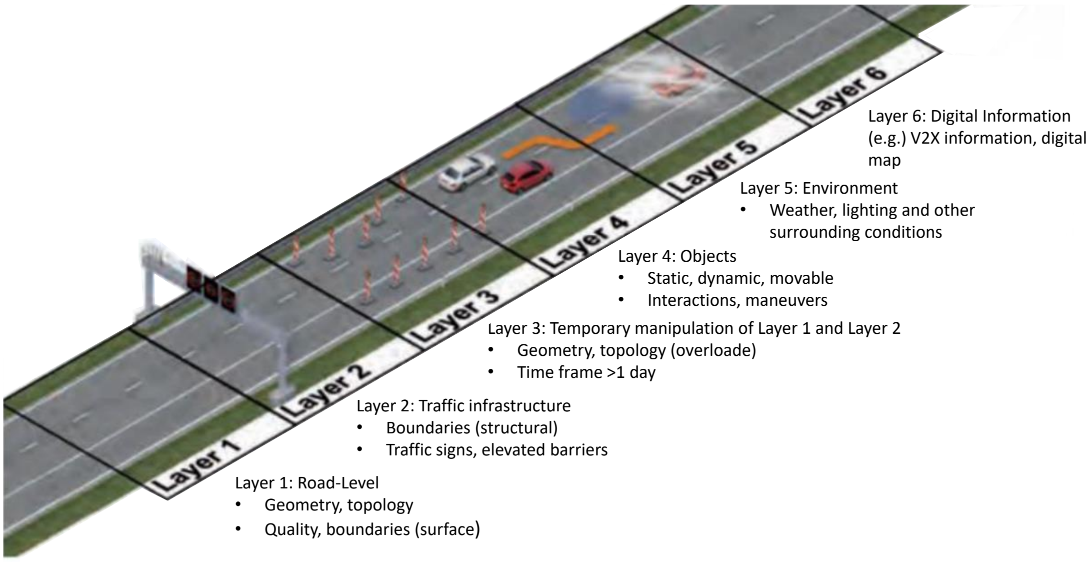


Figure 4.5: Six-layer approach proposed by the PEGASUS project (PEGASUS Project 2019).

The *second* question reveals if the scenarios addressed by the safety measure can be clearly identified. Here, a sanity check of the constraints from the first questions is conducted. In this part the selection criteria for scenarios should be formulated to allow later on to put the results of an effectiveness evaluation in context to other exposure measures and make them comparable.

Figure 4.5 shows the six-layer approach proposed by the PEGASUS project that can support to define the area of application.

4.1.2.3. Third layer

The third and last layer provides the scope of the representativeness of the evaluation results and how the results should be measured (Figure 4.6).

Scope Cluster 4 'Representativeness'	
◯	Is the evaluation region defined (data selection & scaling, interpretation)?
◯	Does the available data represent/support the region of interest?
◯	Is the point in time for the validity of the assessment effect described?
◯	Have you formulated the conditions that describe the expected change of exposure?

Scope Cluster 5 'Metrics'	
◯	Are the evaluation metrics determined (utilization of additional data & post-processing)?
◯	Is there a need for a socio-economic assessment?
◯	Do the evaluation metrics require a certain accuracy?

Figure 4.6: Checklist for the third layer.

Scope cluster 'Representativeness'

The *first* question clarifies the region for which the results should be generated. Crash occurrences and contributing factors are different for various regions, therefore the evaluation region must be stated.

The *second* question is a sanity check if available data represent the selected region. Even if the available data is a sample from the population of the region, it should be checked for potential bias. Such a bias can be induced, for example, by non-random sampling or outcome-based sampling. If data is not available for the region of interest, weighting of data from a similar region might be appropriate. Here, it should be verified that relevant data categories (such as variation of road infrastructure, weather and light conditions) and range of continuous variables (such as driving speeds) are available and comparable in both datasets.

The *third* question points out the point in time for the assessment results. For a priori assessments such a point in time is most often in the future, when a specific market penetration of the safety measure has been reached. However, other changes that may come along with a future point in time such as changes of infrastructure or traffic mode distribution are not necessarily considered.

The *fourth* question evaluates exactly the changes of exposure that are expected at a current or future point in time. Such assumptions or models have to be clearly formulated to understand the effect on the scenarios that are utilized for the virtual simulation or its pre- or post-processing. If there are no expected changes, this should also be stated. It should be also scrutinized if characteristics such as vehicle configurations, infrastructure layout, environmental conditions derived from older data sets are still representing the point in time of evaluation.

Scope cluster 'Metrics'

The *first* question of the final scope cluster supports the definition of evaluation metrics. These metric(s) may either be derived directly from the simulation results or involve further post-processing. A typical example for a metric derived from simulation results is the occurrence of a collision or the collision speed. Pre-crash simulations usually stop when a collision occurs, meaning the conflict opponents contact each other. To derive crash metrics, for example the change of velocity (delta-v) or the principal direction of force, a crash computation needs to be conducted. For an injury-related metric usually an injury risk function or a human body model simulation is applied. For concept evaluations, metrics derived from model outputs can be of interest, for example, object tracking and identification, the point in time when an object enters the field of view, or the time-to collision when a system measure is activated.

The *second* question specifically addresses the need for a socio-economic assessment. These assessments require usually various metrics as input which are not only derived from pre-crash simulation and subsequent processing. Therefore, the feasibility to derive and assess these metrics should be clarified beforehand. An example of such a metric is the probability of an injury with a specific severity leading to long-term consequences.

The *third* question and final question is maybe the most difficult to answer and, thus, may not be answered at all. Evaluation metrics have an uncertainty which often is expressed as a confidence interval. It describes the range in which the true value is expected given a specific level of confidence. Let's assume, the evaluation question aims to identify a significant difference between an evaluation metric using different safety measures. The closer the expected metric outputs are to each other and the bigger their uncertainty, the less likely it will be to identify a significant difference. Similarly, if a regulation requires that a metric does not surpass a certain threshold, the uncertainty of the expected value has to be taken into account. Thus, the requirement of an accuracy of a metric can have a substantial effect on the accuracy of the utilized models and their interconnection. A common way to estimate confidence intervals is the usage of bootstrapping, which uses random sampling with replacements from approximating distributions.

4.1.2.4. Evaluation questions

One or more evaluation questions are derived from previous formulated answers. They usually follow the schema of "What is the <effectiveness measure> of <safety measure> on <assessment metric> in <region> at <time point> given a <implementation rate>?".
Further it is necessary to formulate limitations and considerations that are relevant to interpret the answer to a such an evaluation question. The are key extracts from the answers formulated above.

Based on the drafted evaluation questions, data and methods needs and requirements will be identified. In the case that is not possible to satisfy them, the user may decide to find measures such as collecting new data, or applying new methods, or simply to go ahead while documenting the missing data or methods. However, another option for the user would be to adapt the evaluation questions in order to meet the requirements. Thus, defining the evaluation questions might be an iterative process that should involve all relevant stakeholders.

4.1.3. Input Output

The Input to the Evaluation Scope depends on the purpose of the safety assessment (Figure 4.7). In a top-down approach the assessment is driven by one or more real-world traffic safety issues that have been identified, for example, by analysing collected traffic conflict and/or crash data. Therefore, the input often, but not necessarily, comprises a literature review, which puts the traffic safety issue(s) in the context of current knowledge, identifies gap(s) in the existing knowledge, and formulates research questions to address the gap(s). Here, a gap analysis may include different types of treatments such as vehicle or infrastructure safety measure, vehicle or infrastructure design, nudging of road traffic participant behaviour, or regulation on a more abstract and generalized level. This may, for example, support authorities and policy makers to define a strategy for reducing road traffic crashes and their consequences or vehicle manufacturers to choose a more effective safety concept or approach. Regulatory aspects are relevant as all possible approaches have to comply with them.

On the other hand, in a bottom-up approach the assessment is focused on a specific measure to mitigate crash occurrence and their consequences [Pears 2015]. In this case the research question is formulated on one or more aspects of the measure such as quantification of effectiveness, parameter optimization and selection, cost effectiveness, or desired and undesired side effects. Also here, regulatory aspects are important, either for compliance or for feedback to authorities.

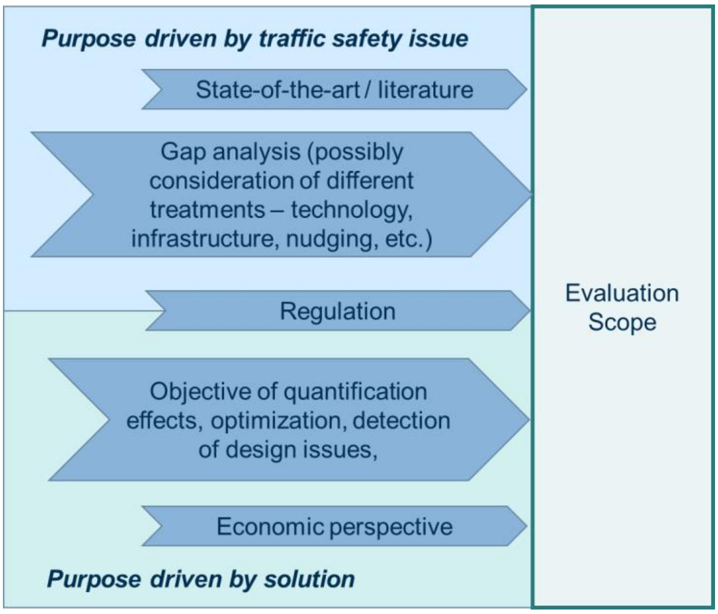


Figure 4.7: Input to Evaluation Scope.

Output of the Evaluation scope should be a clear description of the safety system under evaluation and its intended effects (Figure 4.8). From both, the field of application should be derived, which means the scenarios that are aimed to be addressed. Also, the area (both regional and contextual), point in time, and/or relevant market penetration should be defined. The effectiveness metrics should be aligned with the expected change that the system is intended to provoke. An estimation of the expected effect size(s) might help to specify either input data or data generation sizes. Besides requirements on the data, also requirements on the methods may be derived from the definition of the evaluation scope.

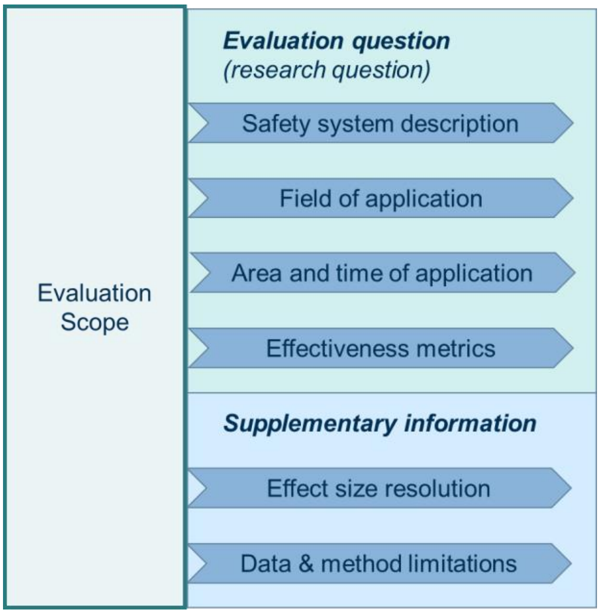


Figure 4.8: Output of Evaluation Scope.

4.1.4. Consequences

The definition of the evaluation scope is an essential process step to guide all subsequent steps. If relevant aspects have not or not sufficiently been considered in the scope the results of the assessment may not align with the assessment objective. There is also the risk that internal or external stakeholders do not trust in the results of the assessment. In a worst case, the output of the assessment will guide towards ineffective safety measures which could have substantial impact on the injury occurrence and loss of live.

4.1.5. Examples

The first example is the use case CRT/BSD (Car Right Turn / Bicycle Same Direction) with AEB and FCW from WP6:

Layer 1:

Q: Is the objective to validate a concept of a product?

A: The objective is to validate the concept of the evaluation framework.

Q: Is it necessary to evaluate false positives?

A: It is not necessary to evaluate false positives as the system measure is a simplified and generic function that has been derived in another project. Further, there is no intention to assess the precision of the safety measure.

Q: Is it necessary to evaluate true negatives?

A: The evaluation of true negatives is not necessary as the specificity of the safety measure is relevant for the assessment.

Conclusions: Because of the concept evaluation, the is no need to validate and verify system functionality against a physical test. Limitation of the data source to crashes (true positive and false negative events) is reasonable.

Layer 2:

Q: Does your safety measure involve vehicle, infrastructure, regulatory, or behavioural aspects?
A: The safety measure involves an in-vehicle system only. Therefore, it should be evaluated, if the measure is operational continuously or only gives intermittent support.

Q: Does the assessment involve continuous operating function?
A: The safety measure is only activated when a normal driving situation evolves into a critical situation with a pending conflict. Therefore, the effect of the measure is only intermittent and will not directly influence the exposure. However, they might be long-term effects in case drivers adapt their diving behaviour to the system support.

Q: Is the safety measure clearly defined?
A: The safety measures is a Bicyclist AEB with FCW. The sensor(s) track and classify a bicyclist (rider on a bicycle) in case it is in the field of view. When the car and the bicyclist are on collision course, the FCW is issued at a certain criticality threshold by providing an alarm sound and pull on the seat belt. In case the driver does not react by braking or steering, the AEB is activated at second threshold. The vehicle will then brake with the maximum available deceleration until standstill. The system measure assumes conservative a high coefficient of friction between the road surface and the tire of 0.x. The sensors are operational at day and night-time and there is no speed range that limits FCW or AEB activation.

Q: Is the safety measure implementation rate defined?
A: The implementation rate is 100% (every vehicle in road traffic is equipped).

Q: Is the area of application defined?
A: The application area represents both urban and interurban junction / intersections. The safety measure may even work on a parking area where similar vehicle motions prior to a collision may occur. However, the lack of road markings and presence of sight obstructions due to parked vehicles does not make these situations suitable for virtual simulation.

Q: Are the scenarios that are addressed by the safety measure clearly identifiable?
A: The scenario is defined as ‘car turning left and bicycle coming from the same direction’, independent from the right of way regulation, road markings, and presence of traffic lights.

Conclusions: The selection of immediate pre-conflict and pre-crash scenarios is in line with the system measure. All conflict and crash scenarios that involve a car and a bicycle with the described relative trajectories are suitable for scenario generation. Aspects such as road markings and traffic regulation are not considered. As only one car with the in-vehicle system is involved, interaction effects do not need to be considered.

Layer 3:

Q: Is the evaluation region defined?
A: The evaluation region is Europe.

Q: Does the available data represent the region of interest?
A: Available data is from the German In-Depth Accident Study (GIDAS). This data is biased due to outcome-based sampling. The data needs to be weighted to represent the German national statistics. Some characteristics are only available from the Pre-Crash Matrix (PCM), a subset of the GIDAS data. Weights should be computed to make the PCM data representative to GIDAS. The German national statistics are not representative for Europe. Thus, weights have to be applied to

make German crash statistics as similar as possible to the European statistics. Therefore, in total three different weight have to be considered.

Q: Is the point in time for the validity of the assessment effect described?
A: The point in time is of today. Therefore, the GIDAS / PCM data must be weighted to the today's European crash statistics. Further, today's vehicles are mostly equipped with ABS and Blind Spot Warning, therefore they are considered as available in-vehicle systems in the baseline generation.

Q: Are the conditions that describe the expected change of exposure formulated?
A: There is no expected change of exposure.

Q: Are the evaluation metrics determined?
A: The evaluation metrics are selected as following:

- Effectiveness in crash avoidance
- Effectiveness in injury mitigation (various levels)
- Reduction of socio-economic costs
- Cost-benefit ratio

Q: Is there a need for a socio-economic assessment?
A: A socio-economic assessment is part of the metrics. The definition of required injury severity levels is not yet available.

Q: Do the evaluation metrics require a certain accuracy?
A: As the use case analysis is done for a concept evaluation, there is no specific requirement on the accuracy of the evaluation metrics.

Evaluation question(s):

- *What is the percentage of avoided CRT/BSD crashes, if 100 percent of today's vehicle fleet in Europe would be equipped with the Bicycle AEB/FCW under evaluation.*
- *What is the percentage of injury mitigation (level x) in CRT/BSD crashes, if 100 percent of today's vehicle fleet in Europe would be equipped with a Bicycle AEB/FCW.*
- *How high is the reduction of socio-economic costs caused by CRT/BSD crashes, if 100 percent of today's vehicle fleet in Europe would be equipped with a Bicycle AEB/FCW.*
- *What is the cost-benefit ratio of Bicycle AEB/FCW systems in CRT/BSD crashes, if 100 percent of today's vehicle fleet in Europe would be equipped with it.*

4.1.6. Q&A

Not relevant for the draft.

4.2. Select Baseline Approach

The topic "Select Baseline Approach" describes different approaches to set up the baseline in the prospective traffic safety assessment by simulation. The baseline is a set of data without the road safety measure under study, to be compared when performing prospective assessments of a measure's safety performance. Based on these data the individual cases that are simulated are generated. Following the work in ISO21934-2 (n.d.) and Wimmer et al. 2023 three different main approaches are described: A) Simulation of original cases without. modification, B) Simulation of original cases with modifications and C) Simulation of synthetically generated cases. These main approaches are further subdivided in terms of their case generation and instantiation process. The case generation process covers relations between the original cases and the later simulated cases while the instantiation process addresses the use of pre- or in-simulation models for the case generation. The introduction of the different baseline approaches is followed by describing aspect to consider when selecting an appropriate baseline approach.

4.2.1. Visualization

The interplay of the baseline approach selection with the other relevant topics is given in Figure 4.9. The selection of the baseline approach clearly requires the input from the evaluation scope. At the same time the limitations in terms of available data sources and models for generating the baseline need to be considered. Hence, the process that appears to be a straightforward process in theory is in practice rather an iterative process that needs to find the best solution within the given constraints from models, data, and evaluation scope. Although the objective is to clarify the taken baseline approach in the stage of the assessment preparation, often refinements need to be done during the simulation execution. Also, the order of the simulation preparation (select baseline – prepare data – select model) is not necessarily a strict order and can vary for different evaluation scopes. The simulation execution is the stage in which the baseline approach is later applied. The practical implication will be discussed further in the upcoming deliverable D2.1.

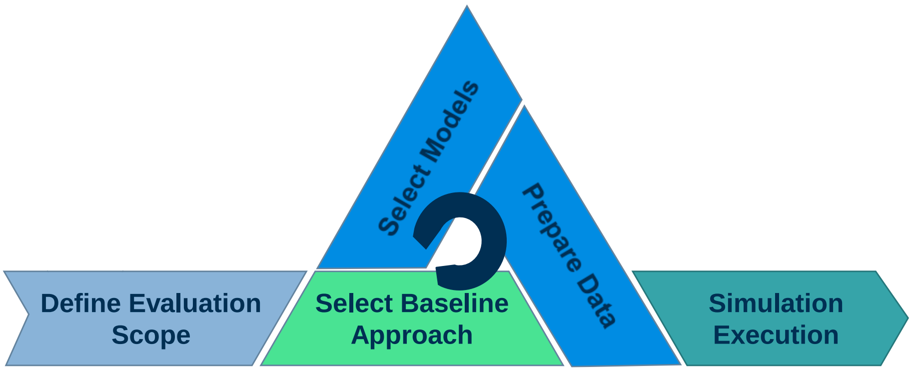


Figure 4.9 I nterplay of the topic "select baseline approach" with the other topics of the V4SAFETY framework.

Figure 4.9 provides an overview about the different "baseline approaches". The main baseline approaches build up on the work in ISO21934-2 (n.d.) and Wimmer et al. 2023.

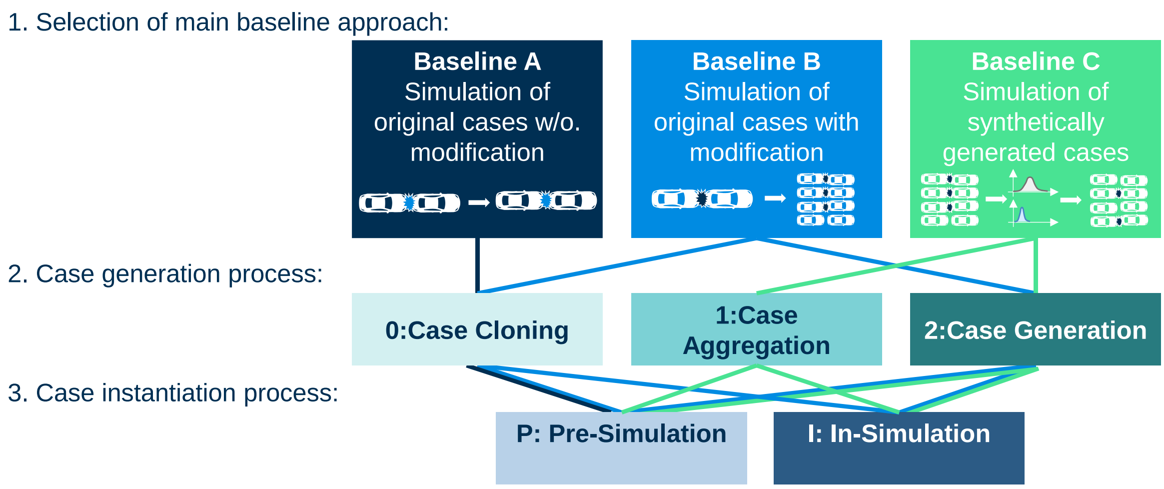


Figure 4.10 Baseline Selection Approach

4.2.2. Detailed Description

The assessment of the safety performance of a safety measure is done by a (relative) comparison of the situation without the safety measure (baseline) with the situation with safety measure present (treatment). Thus, establishing the baseline is an essential process in the safety performance assessment. Hence, this step requires special care in the preparation of the assessment.

First, a clarification of the central term "case" for the baseline selection is done. A concrete scenario is defined for V4Safety as scenario that is "always anchored in the real world and may contain a substantial amount of information". This can be crashes, critical or normal driving scenarios. The word "Case" is used synonymous for concrete scenarios. In this context, an original case is a case from the primary data source used as input for the simulation. A simulated case is a case created through simulation(s) as part of the assessment. A simulated case can be the outcome from either baseline or treatment simulation.

The selection of an appropriate baseline approach – i.e., method to derive cases from input data – depends strongly on the evaluation scope and the considered safety measure. Before any selection of the baseline approach, it is important that the evaluation scope and the safety measure to be assessed (see section 4.1) are known. Next to this, constraints in terms of models and data (e.g., available models, access to data) are the main input for selecting a baseline approach for an assessment.

Following the work of ISO21934-2 (n.d.) and Wimmer et al. 2023, there are three different main baseline approaches that can be used to create cases for the simulation-based safety performance assessment:

- Baseline approach A - Simulation of original cases without variation,
- Baseline approach B - Simulation of modified original cases,
- Baseline approach C - Simulation of synthetically[1] generated cases.

These three main baseline approaches can be further subdivided by three different ways of generating cases.

- Case Cloning (0): A process, in which only the original cases that have been identified as relevant are simulated. Thus, the number of the resulting simulated case in the baseline equals the number of original cases.
- Case Aggregation (1): A process in which the information of multiple original cases is used to derive one or more cases, which aims to represent a substantial proportion of the original cases. The resulting number of simulated cases is smaller than the number of original cases.
- Case Generation (2): A process in which new simulated cases are generated based on information that is derived based on (multiple) original cases (directly or indirectly via e.g., distributions). There are two main types of case generation. First, there is the in-case generation in which multiple variations of each original case are generated. Second, there are the synthetic cases, in which cases are generated based on sampling from parameter distributions or other mathematical approaches without a direct link to a single original case. The in-case generation can include variations that are intended to enhance robustness, but also variations that change some relevant aspects of the original case (e.g., simulating a wide range of possible behaviours as part of scenario generation).

[1] The synthetically generated cases need to have a link to the real world (e.g., via distributions derived from real-world cases).

Typically, the number of simulated cases is larger than the number of original cases. It needs to be noted that for baseline C there is not a direct link to original cases. Therefore, the original number of cases might be unknown. Thus, case generation for baseline C means that a high number of cases (typically > 100) is generated. Consequently, case aggregation for C means that a low number of cases (typically < 100) is derived.

In addition, the main baseline approaches can also be differentiated by the instantiation approach that is later used in the simulation execution to get to the baseline cases:

- Pre-simulation instantiation (P): The process of instantiation of a simulation case, in which the core of the simulated case is generated prior to the simulation. The core of the case is the movement of the individual traffic participants. In this type of instantiation in the baseline simulation, the agents follow pre-defined trajectories.
- In-simulation instantiation (I): The process of instantiation of a simulation case – at least related to the core movement of traffic participant – is done during the simulation. Here, the case is complete first once the simulation is finished. The initial conditions are pre-defined. In each simulation step, a model defines (at least partially) the state (incl. movement) for at least one agent in the simulation.

In the following the baseline approaches are discussed in more detail.

Baseline approach A is given if concrete real-world cases are simulated without any change (i.e., replay of the original case). Thus, the original and simulated cases are the same up to the technical feasibility of the simulation. The key aspect in this context is the intent to replicate the original case without modifications. The real-world scenario can come from various sources. Typical examples are reconstructed crash scenarios or driving scenarios recorded during a field trial (e.g., naturalistic driving study). For baseline approach A it is characterising that the scenarios are not changed in their parameters in any way. However, if missing information is added to the case to ensure that the simulation can be run, the baseline approach is still A. If any information is modified, then the baseline approach becomes B. Hence, every change of the scenarios automatically turns into baseline approach B.

For baseline approach A the number of investigated cases is always equal to the number of cases that have been identified as relevant in the data source. Baseline approach A is typically associated with a case cloning (number of cases is equal to the identified ones). If new cases are generated from an original case or variations of the original case are made, the baseline approach turns into B. Consequently, case generation or aggregation are not possible for this baseline approach.

Regarding the case instantiation, baseline approach A is typically associated with pre-simulation instantiation. All in-simulations models running for baseline approach A must be designed with the aim to replicate the original case. An in-simulation instantiation in which models are running during simulation, is theoretically feasible, but rather unnecessary since the kinematic parameters of the assessed scenarios are given by the original case.

The baseline approach A is an appropriate approach if

- The evaluation scope explicitly asked to analyse non-modified real world / original cases.
- enough cases are available in the used databases to allow for assessment according to the assessment scope.

- the duration of the case is sufficient to cover impact on traffic safety of the safety measure under assessment.
- [List will be continued during the work in WP2.1]

A limitation for the baseline approach A can occur:

- if the evaluation scope asks for assessing true positive (i.e., check whether crashes are avoided) and false positive behaviour (i.e., check whether the safety measure does not lead to new crashes) of the safety measure at the same time. This is only feasible if the database with the original case combines crash and no-crash cases. Typically, this is not the case for single data sources since they often include only one type of scenarios. If multiple data sources are combined this is again feasible. If a data source with only original no-crash cases is available, baseline approach A would allow only to investigate safety performance in terms of safety surrogate measures (e.g., time-to-collision), but not the investigation of safety performance in terms of crash avoidance.
- [List will be continued during the work in WP2.1]

Baseline approach B "Simulation of original cases with modification" describes an approach that uses similar to baseline approach A original cases but allows for the modification of these cases[2]. The basis for the assessments in baseline approach B is always concrete real-world scenarios. Most often these are concrete crash scenarios reconstructed in crash investigations, concrete scenarios from recordings (e.g., event data recorders or site-based crash recordings), or driving scenarios recorded during a field trial (e.g., naturalistic driving study or site-based data collection).

Different types of modifications in baseline approach B are feasible. The kinematic parameters (e.g., higher velocity or shorter distances) or the equipped systems of the vehicle under assessment (e.g., consider an ESC system for vehicle which had not ESC before to adapt for today's traffic) can be modified. Baseline approaches B can be combined with all three generation processes (B0 case cloning, B1: case aggregation and B2: case generation). In practical sense, baseline approach B0 and B2 play a major role in prospective safety assessment of safety measures that act prior to a crash.
In case missing information is added to the simulation, baseline approach B can be applied together with case cloning. However, to deviate from A0 further modification(s) need to be added to the original scenario.

A case aggregation (baseline approach B1) is also feasible but plays in real assessments a minor role. A use case for baseline approach B1 would be simulations that require a high computation effort and takes a long time until the results are available (e.g., simulation considering passive safety simulations). In this use case, the number of simulated cases needs to be limited for practicality reasons. If baseline approach B1 is chosen, it is likely that the case will go along with an intermediate baseline for which first cases are generated (B2). Since the baseline selection is considered an iterative process, this combination of baselines is feasible.

If the kinematic parameters vary or the modification considers replaying the original behaviour of road user, by computational driver behaviour model, the number of original cases typically stays constant. However, the number of simulated cases typically increases, i.e., this approach is associated with case generation (baseline approach B2). Although there are different possibilities for variations, it is important that a link to original cases remains. If the case generation gets

[2] The modification can also include complementing missing data that is relevant for the simulation of an original cases in conjunction with modifying the cases.

disconnected from the original cases (i.e., nothing is kept from an original case), it turns into the baseline approach C. In terms of case instantiation, both “pre-simulation” and “in-simulation” are feasible in this approach. However, for the in-simulation instantiation, it must be noted that the applied models – in particular computational driver behaviour models – should generate a realistic movement of the traffic agents in the simulation. Hence, the quality of the relevant models should be checked (see section **Error! Reference source not found.**).

The baseline approach B is an appropriate approach if,

- original real-world cases are available.
- the evaluation scope asks for adaptations of the real-world scenarios, e.g. to examine what-if variants
- models for a representative variation of the original case for a specific population are available.
- the duration of the case is sufficient to cover impact on traffic safety of the safety measure under assessment.
- there is missing information in the real-world data which does not allow to use baseline approach A. Therefore, complementary data can be used to have enough information to generate the baseline.
- [List will be continued during the work in WP2.1]

A limitation for the baseline approach B can be:

- if the evaluation scope asks for assessing true positives (i.e., check whether crashes are avoided) and false positives behaviour (i.e., check whether the safety measure does not lead to new crashes) of the safety measure at the same time. This is only feasible if the database with the original case combines crash and no-crash cases. Typically, this is not the case for single data sources since they often included only one type of scenarios. If multiple data sources are combined, this is again feasible. If only one data source with original no-crash cases is available, baseline approach B would allow only to investigate safety performance in terms of safety surrogate measures (e.g., time-to-collision), but not the investigation of safety performance in terms of crash avoidance.
- [List will be continued during the work in WP2.1]

For **Baseline approach C** the strict link towards an original case as in baseline approach A & B is not required. Here, the link is established in an indirect manner via the mechanisms to generate the simulated cases. Often these mechanisms include sampling approaches combined with distributions that have been derived from real-world scenarios and /or the use of appropriate models.

Baseline approach C enables a case aggregation (C1 – e.g., test cases used in a virtual Euro NCAP test[3] or selecting representative case to be considered in a time-consuming simulation like passive safety measures) as well as the case generation (C2 – e.g., stochastic generation of cases). In application, the case aggregation and generation are quite different and associated with quite different requirements for the simulation and the required models. Both can be executed by pre- and in-simulation instantiation. The in-simulation instantiation plays an important role when the baseline approach C with case generation (baseline approach C2) is utilized. A typical use case is that the initial parameters are sampled from the pre-calculated distributions, but a driver behaviour

[3] Euro NCAP selects representative concrete scenarios based on real world accident cases to assess the performance of safety measure in real-world test. To limit the effort to feasible extent. The example refers to the use cases that these tests are now done the same way but virtually.

model is used to derive the trajectories of relevant traffic participants (simulated agents) in the case to get to the conflict.

It must be noted that the applied models – in particular computational driver behaviour models – to generate the simulated cases should generate a realistic movement of the traffic agents in the simulation to achieve realistic traffic scenarios. Therefore, the quality of the relevant models should be checked (see section 7.1 Conduct Validation & Verification).

The baseline approach C is an appropriate approach if

- Enough data is available to derive the input parameter distributions or if the distributions for the input parameters are given.
- Only mechanism of the scenario (i.e., either statistics or parameter distributions or models that describe the conflict type) is available, but not necessarily the original cases. The data can come also from several data sources.
- Models are available to generate, in terms of traffic safety, realistic simulated baseline cases (e.g., valid computational driver behaviour models).
- Cases with a longer duration than in reconstructed original case are required for the assessment (e.g., assessment should also cover the avoidance of traffic safety related conflicts).
- the evaluation scope requires to consider traffic context within the simulated case. This covers for instance the case that surrounding traffic participants should be considered and information about them is not available in the data sources. Furthermore, the penetration rate of certain safety measures in traffic should be considered for the assessment.
- [List will be continued during the work in WP2.1]

A limitation for the baseline approach C can be:

- If adequate models or data are not available to describe the mechanism of the crash respectively the scenario flow that leads to the crash scenario properly.
- [List will be continued during the work in WP2.1]

Examples for the different baseline approaches are provided in section 4.2.5.

The remaining question for the baseline selection is “what is the right baseline approach for my assessment”. Obviously, there is no general answer to this question. It rather depends on the evaluation scope, including the safety measure to be assessed and the individual constraints. However, to provide some guidance, the following questions should help to identify the most appropriate baseline approach.

- Do you have access to real-world cases? Yes: all approaches are feasible; No: the only possible baseline approach is C.
- If you want to apply baseline approach A or B: Do your real-world cases include crashes? Yes: Both approaches are feasible; No: Only baseline approach B is feasible, since only by variation you will get crashes. If the used metric only investigates surrogate measures (e.g., safety critical scenarios) both approaches are possible.
- Do you need to consider a different scenario exposure than what is available in the given data? Yes, you will either need to select C, or find alternative sources for exposure estimates. However, note that if baseline approach C is chosen for exposure estimation, validation of exposure prediction is important; No: you can use any approach.
- Do you have validated and appropriate computational behaviour models to generate the baseline as described in the assessment scope? Yes: In-simulation based baseline

approach B and C are possible; No: Only pre-simulation-based baseline approaches of A, B and C can be selected.

- [List will be continued during the work in WP2.1]

Baseline approach A0I, A2P and A2I cannot exist, since either the modification by means of in-simulation models or the increase of number of cases leads to the situation that the baseline approach becomes B. A1I and A1P are not feasible baseline approaches since the aggregation process would allow for A only a sub-selection of cases. Any other aggregation process requires modification of the cases which is not A. However, a sub-selection of cases is equal to a different selection in the first place. Therefore, a sub-selection would not be considered for A and B as a case aggregation. C0P and C0I are not considered as baseline approaches since approach C has no direct link to original cases and consequently case cloning is not possible.

Note: updates of the section will consider iterative approaches in the selecting the baseline approach (interplay between scope, data and models).

4.2.3. Input Output

Inputs for the baseline selection are the evaluation scope as well as the available data and models:

- Evaluation Scope (purpose of the assessment; safety measure under assessment etc.)
- Constraints in conducting the assessment due to available models and access to data source.

Outputs of the baseline selection are:

- The most appropriate baseline approach
- Requirements in terms of data and models

4.2.4. Consequences

Not relevant for the draft.

4.2.5. Examples

In the following Table 4.1 examples for the different feasible baseline approaches are given. The examples are based on the publication Wimmer et al. 2023.

Table 4.1: Examples for the baseline selection based on Wimmer et al. 2023.

Baseline Approach	Example
A0P	• In Saadé et al. 2019, an AEB-Pedestrian system was assessed by simulating crashes selected from the French accident database, VOIESUR (Lesire et al. 2015). The database covers all fatal and 5% of all injury crashes that occurred in France in the year 2011. For the simulation only crashes were considered in which a pedestrian was hit by the front of a passenger car and the trajectory of the traffic participant and impact speed could be determined. • SIMPATO (Safety IMPact Assessment Tool) (Van Noort et al. 2015): SIMPATO was developed and used in the EU-funded project “interactIVe” that dealt with different active safety measures for multiple conflict types. The SIMPATO tool focuses on rear-end and run-off conflict situations. In the simulation analysis, 364 real-world rear-end crashes and 150 run-off road crashes of the GIDAS database were considered. The safety measures for the rear-end conflict that were analysed included warning in-car systems as well as intervening systems (braking and/or performing evasive manoeuvres). The systems’ simulation models were build based on the

		interactIVe test track tests. For the run-off road conflicts the interactIVe system reaction was always a steering manoeuvre.
B0I | B0P	•	Original data did not have all the required information → Information is added (from whichever means, assumptions, distributions…) • Original data had all relevant information, but it is modified → Information is modified (from whichever means, assumptions, distributions…)
B2P	•	Urban et al. 2020-1 provides an example for adding missing as required for baseline approach B. The input are real-world crashes from a police reports accident database in Germany. The data contains information such as the accident conflict situation, collision configuration, geo-coordinates of accident, participants involved as well as injury level of each participant. The data does not contain any time series information such as trajectories. Therefore, the trajectory is derived for these cases on data from GIDAS database. The GIDAS database contains information such as participant manoeuvre, accident location, participant type as well as driving speed, collision speed and deceleration value. The last three parameters are used to determine the speed profile to determine the likely trajectory of the traffic participant in the original case. Since the trajectories are pre-calculated the approach is considered B2P. • Building on the previous example Urban et al. 2020 – 2 provides an example for additionally creating variations. The first steps are basically the same. However, the initial trajectory in these cases is derived based on the description of the police report. In a second step the variation of the speed profile is added based on a statistical analysis of the GIDAS database, by considering variation for the driving speed, the collision speed and deceleration value. Considering the 2 participants involved in the accident, a maximum of 729 variations is generated per original case. Since the trajectories are pre-calculated the approach is considered B2P. • In the L3Pilot safety impact assessment (Bjorvatn et al. 2023) the AD systems that were developed in the project were evaluated. For the evaluation of rear-end and cut-in conflicts the baseline approach B2P was used by using a counterfactual simulation. The real-world cases came from the dataset involving crashes with Volvos (VCTAD see Isaksson-Hellman et al. 2005]), crashes from the Traffic Accident Scenario Community (TASC see Urban et al. 2020-1) database and critical situations from the SHRP2 database (Hedlund 2015). The latter database was used to assess the false positive behaviour of the systems, while the first two were used to assess the true positive behaviour.
B2I	•	Fries et al. 2023 describes a study in which the performance of a driver behaviour model in certain driving scenarios is assessed. Thus, it is not a full assessment of safety measure, but it describes the principles of the baseline approach well. In this study the starting condition of original scenario that is a passive cut-in manoeuvre is taken from a field operational test. The trajectory of the cut-in vehicle is taken directly from the recorded data. For movement of the ego vehicle the situation is simulated with the driver behaviour model “Stochastic Cognitive Model”. Hence, each simulation case contains a different drive and a different movement and reaction of the ego-vehicle. These cases could serve as a baseline for the assessment of safety measure.
C1P	•	The CATS project (Op den Camp et al. 2017) provided a proposal for Euro NCAP tests to rate Autonomous Emergency Braking (AEB) Systems that address conflicts with cyclist. Basis for the test were an analysis of car-to-cyclist crashes in different EU countries, namely France, Germany, Italy, the Netherlands, Sweden, and the United Kingdom. After processing and analysing the data, the three most relevant crash scenarios for this conflict type were identified. Based on in-depth accident study and observations, the relevant parameters (speed and presence of view-blocking obstruction) and ranges of these parameters for test scenarios were defined. By means of these analyses the thousands of accidents were reduced to a couple of baseline tests to assess AEB cyclist systems.

C1I	•	No example found yet.
C2P	•	In Denk et al. 2022, the effect of a simplified AD system with / without information of an external infrastructure-based LiDAR sensor is assessed. The relevant conflict type is a right-turn scenario, in which a passenger car is turning right in which it has yield to straight driving cyclists. In the scenarios, occlusion by parked cars and a construction site was considered. In the study the crash causation mechanisms were first repliced. In a second step the mechanism is applied to create the simulation cases. Overall, 200 million cyclist crossings were simulated in the baseline and for the 3 different treatment conditions.
C2I	•	In Helmer et al. 2014, the safety performance of a safety system addressing conflicts with pedestrian was analysed. The simulated scenario dealt with a jaywalking pedestrian. The analysed system covered warning and or an AEB. The systems were simulated with different parametrizations. The objective of the simulation was to replicate the risk in the described scenario as precisely as possible. To achieve a statistically significant result, 18 million crossings cases were simulated in the baseline and 100 million crossings were simulated for the treatment. The pedestrian and the vehicle driver are both represented by models in the simulation, which decided on their action during the simulation.
	•	L3Pilot safety impact assessment (Bjorvatn et al. 2023): Besides the B2P/I baseline approach, the L3Pilot project also applied the baseline approach C2I to assess the safety impact of automated driving. The approach for the motorway scenarios, namely lane change conflict, conflict with VRU, minimum risk manoeuvre, wrong activation, end of lane, obstacle in the lane, lower speed limit and passing a motorway entrance. The number of analysed cases per scenario varied depending on the considered infrastructure and traffic parameters. Overall, more than 25 000 cases were simulated. The parameters for the scenarios were defined by means of a stochastic approach in which the initial conditions were sampled from distribution for the relevant parameters that based on traffic data and crash databases. In the simulation, the trajectories of the conflict partners (e.g., crossing VRU) were predefined. However, the reaction of ego-vehicle as well as the surrounding traffic was determined in the simulation by means of driver behaviour model. The C2 approach was also applied for the urban AD systems. Here, all scenarios were generated with a stochastic sampling approach using copulas, which was presented in (Rösener 2020). Input to the generation of the simulation cases were different sources including accident data, traffic data and data from L3Pilot pilot studies. Also, in these simulations a driver behaviour model has been used to derive the reactions of the host vehicle's driver in the baseline cases.

4.2.6. Q&A

Not relevant for the draft.

4.3. Prepare Data

The fundament of a prospective safety assessment is the retrieval, processing and documentation of data because they form the basis for essential process steps such as the generation of scenarios, the parametrization of simulation models or the validation. Therefore, the topic "prepare data" deals with the **retrieval**, **processing** and **documentation** of all data that is used in the safety assessment process. Typically, data needs must be identified from the provided scope, the evaluation questions and data requirements from the baseline approach and models. According to Select Baseline, preparing data can not necessarily be seen as a step after the selection process of the baseline, but interdependencies have to be considered.

For the **retrieval** of data, the identified needs are linked to available data sources stated in Deliverable 4.1. In case appropriate data sources cannot be identified, alternatives are assessed for further data generation and utilization. Data sources are usually heterogeneous and not necessarily available in a form that can be directly used for performing safety effectiveness assessments. Therefore, most often data has to be **processed** to be adequately utilized to the simulation framework and its contained models, assess and ensure representativeness, and validate the processes and output. Another important role is the **documentation** of data sources and applied processing (incl. data grading and limitations).

4.3.1. Visualization

Error! Reference source not found. provides a summary of the process step “Prepare data”.

Figure 4.11 provides a summary of the process step “Prepare data”. Based on four different inputs, retrieval, processing and documentation are distinguished. More detailed descriptions for retrieval (Figure 4.12) and processing (Figure 4.13) are shown in the figures below describing the sub processes.

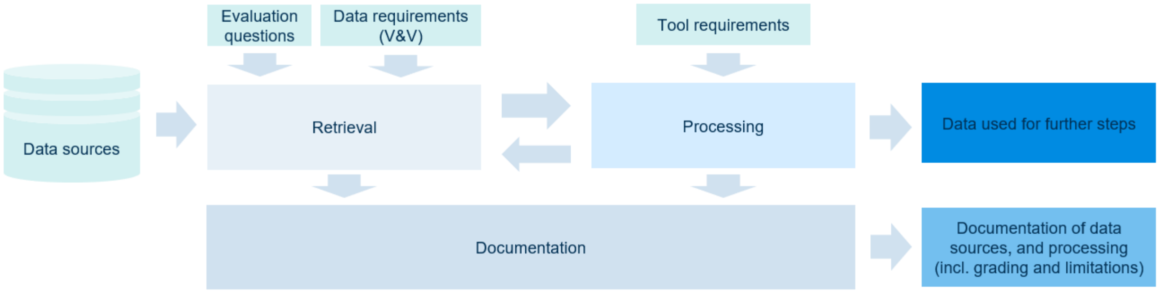


Figure 4.11: Flowchart to evaluate and prepare data

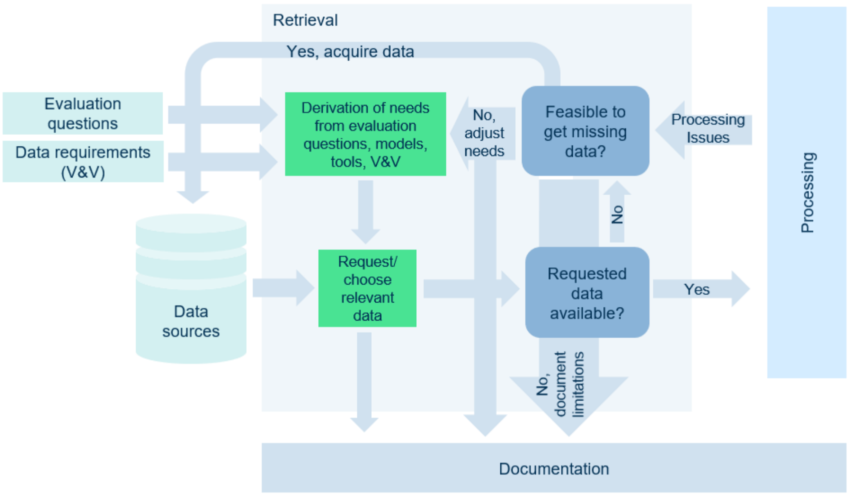


Figure 4.12: Data subprocess data retrieval

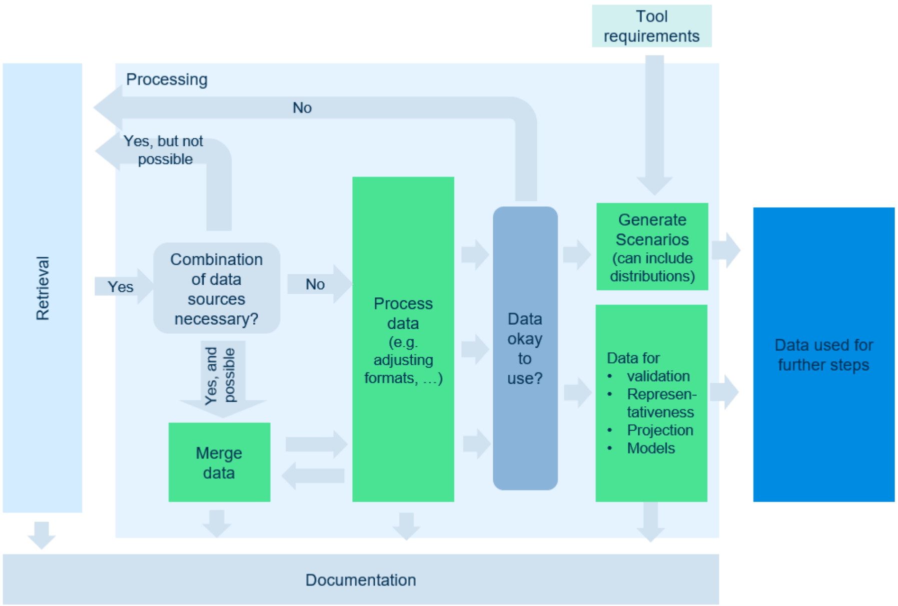


Figure 4.13: Data subprocess processing

4.3.2. Detailed Description

The selection and preparation of data is an essential part of the virtual safety assessment in order to retrieve reliable results that are representative for real-world. This implies usually the following steps.

The data preparation consists out of three main steps retrieval, processing and documentation. Those are not covered in a linear workflow but have interdependencies and can require feedback loops. Thereby, those are steered by the availability of data, methods to be used and limitations which can be stated. Therefore, the individual steps are described below.

4.3.2.1. Retrieval

The **retrieval** is the first and an essential step in the data preparation for the safety assessment. Within this step, data requirements and needs are first be identified and derived from the definition of the evaluation scope, models and tools. Based on these, the next step is to check whether the data meet the derived requirements and, if necessary, to find a compromise between adjusting the evaluation scope and derived requirements, documenting limitations or initiating further data retrieval.

Data requirements

The need to provide data arises from requests and requirements of different steps of the safety assessment framework. Among other things, the evaluation questions, the evaluation scope, and planned actions for V&V are decisive for this. In particular, the requirements that result explicitly from this must be identified and translated into specific needs for necessary data. Data requirements may result in the need to identify specific types of data such as:

- Data to parametrize models
- Data for verification and validation

- Data for baseline creation
- Data for projection
- Data for cost benefit analysis
- …

The need to obtain specific data must be decided based on the consideration of requirements with regard to scope and needed quality to be met. Evaluation questions shall contain minimum information from which data requirements can be identified, as mentioned under **Error! Reference source not found.**. This contains qualitative as well as quantitative aspects further elaborated in Deliverable 4.1.

Qualitatively, different aspects have to be covered. Data should be transparent meaning that it has to be reliable and usable to enable a transparent usage and traceability of data. Furthermore, it has to be representative according to the given evaluation scope. Therefore, data has to fit this and biases should be clearly addressed. Furthermore, quality aspects as accuracy, consistency, completeness, continuity and timeliness have to be considered.

Quantitively, a key requirement is the accessibility of enough data. "Enough" is relative. Sample size calculations and completeness assessments have to be performed to assess the necessary sample size to show a certain effect and potential limitations should be documented. An example for this is a power calculation (Hickey et. al (2018))Key factors influencing the calculation include the data variability of and potential covariances within the accessible data, the desired significance level of the results, and the approximated effect size. Especially the last two, thereby, have to be extracted from the evaluation questions. If more data is accessible than requested for a minimal confidence, the reliability of the data due to its amount increases. If less data is available, further in the process, this has to be checked and limitations have to be documented or if possible, adjustments of the needs could be made.

Besides the quality and quantity of the data, restrictions are set by other requirements, such as the compatibility with the available tools and models to guarantee further processing. This includes, in particular, requirements for the format of the data or can also include interface requirements for tools. The format must be either directly usable in the processing pipeline with its simulation tools or be transferable into them through further process steps. For example, scenarios do not have to be directly available in an OpenX standard (ASAM e.V. (2024)), but it must be possible to derive simulatable scenarios by means of further process steps. An additional requirement in case of retrieving data in order to merge with other data, is that the data is matching each other regarding space and time (e.g., same type of area like urban or same country or same time range).

Next to technical factors stated above, further requirements arise according to availability and costs must be considered. Furthermore, ethical (data collected with low or no ethical standards) and legal aspects (as data protection laws, and access rights to data) should also be considered when selecting data sources.

Data selection

Based on the identified and evaluated data sources, the next step is to acquire the data from available sources. Data sources can thereby be classified in different sub-categories. Generally, we differentiate between primary data, which are data collected from real-world events/reality, and secondary data, which are data that are e.g. acquired through simulations. Within each, multiple categories can be distinguished which are described in Deliverable 4.1. This can e.g. be reported crashed, real-world driving data or scenario-specific data. The data source categories serve different needs such as being the basis for generating baseline scenarios or for developing and parameterizing models. In addition to data sources where the measures you after are readily available, there are also data sources where substantial processing is needed for it to be useful for safety assessment. Examples include videos and police reports. If expertese is available to process these into useful measures, that an be done, but often resources, time and expertise is lacking.. Generally, it is likely that not all of the requested data for the whole simulation framework come from different data sources. In the case of the use of different data sources the data must be queried and homogenized within data processing (4.3.2.2).

Handling data limitations

After the identification of needs and the acquisition of data, data have to be checked according to set requirements (see **Error! Reference source not found.**). If all requirements are met, data can be processed further and may be merged in a next step. However, it is possible that the requirements cannot be fully covered by available data. This can be the case if data is either not available, not accessible, or no available data meet the derivedrequirements. Another reason can be that ethical or legal reasons prevent the use of possible accessible data. In this case, there are various options for the further procedure depending on available ressources and external constraints:

1. If not enough data is available or this data does not fit the derived requirements, new data may be acquired if the ressources are available. This can include the collection of new data as well as buying existing data.
2. If it is not feasible to acquire new data, the limitations of the data should be documented, and the data should be used for further processing. However, if the limitations have substantial impact on either the methodology or the outcome, the data should not be used.
3. If limitations are too significant to be acceptable, the evaluation scope should be revised. An adaption of the question can lead to different requirements for data so that the available data may be sufficient for the adapted research question.
4. If none of the options above are feasible, the available data is not sufficient for an assessment and no assessment should be made.

Especially the third option can lead to recursion loops with select baseline, setting the evaluation scope and select models. Adjustments of needs may thereby occur because of different unavailabilties of data e.g., data for specific models or scenarios. This has then a direct impact on chosen models and may have even bigger impacts on the choice of the baseline approach or the evaluation scope, which means a revision of the research questions may have to be considered.

4.3.2.2. Processing

Once data is collected it may be forwarded directly to the framework steps. Typically, the data has to be tailored and processed for their usage. Thereby, different steps as e.g. data understanding, data formatting, data quality handling, bias checks and handling and merging can be distinguished. An Overview over individual steps is given below. However, more details are given Deliverable 4.2 (REFERENCE).

Since processing of data can implicit changes in the data itself, before using the data, it has to be checked if the requirements set before are still met. The processed data may differ from initial crosschecks against derived requirements due to multiple process steps and concretization of data. Processing and merging individual data to fit further process steps may lead to relevant transformation of data, which may affect data quality. Furthermore, potential drawbacks in the process and potential assumptions about the fulfilment of requirements to data in initial steps of the data processing may arise. Since an assessment if requirements can be met may be vague or need assumptions in the beginning, they can be more concrete after processing when applying them directly. So, if data does not fulfill the requirements after processing, actions as described in data provision have to been taken to improve data or to adjust needs.

Data understanding, formatting and quality handling

To work with selected data, understanding variables as well as handling formats and quality of this data for further processing is essential. Starting with the processing, it has to be clear what variables should represent and how they are created. Therefore, things as units, categorizations and representations of e.g. null have to be clarified. Although the data may be understood, it may not be processable so that is has to be formatted. Usually, thinks as units and precisions has to be aligned. Furthermore, data may have to be cleaned. Cleaning means the handling and adjustment of different data aspects such as inconsistencies, outliers and the alignment of data ranges. Data content should not be changed, but those points have to be addressed to come to a consistent and processable dataset.

Bias checks and handling

There are many different types of biases in data as stated in Deliverable 4.1. These have to be addressed to consider and potentially to reduce the impact of these biases in the assessment results. According to Deliverable 4.2 these can have multiple reasons and impacts on further data processing. Those may e.g. be differences in variable definitions, incompleteness, missing data, demographics or biases in distribution shapes. For this identification, individual variables should not be studied in isolation, but combinations and multivariate dependencies should be taken into account. If a clear dependency can be seen and is relevant for the given data and evaluation scope. Multiple methods exist to handle such biases. Those are highly dependent on the bias and its impact. One common method are weighting procedures to reduce biases in one data source or when combining data sources.

Merging of data sources

If multiple data sources are chosen as relevant a specific (e.g., baseline generation or model generation), they have to be merged for further processing. Merging and enriching can be relevant since further models and procedures normally use one aligned data source to work with in an individual process step, but individual data sources may lack characteristics, include biases, or have not enough samples. However, it may not always possible to merge or enrich data and it highly depends on the data which should be processed.

Enrichment of data is thereby based on common identifiers so that to distinct data sources (e.g. trajectory data and weather data) can be merged based on a common variables (as e.g. time is absolute timings are used).
Similarly, the process of data merging highly depends on the data that should be merged. Different types of merging can be necessary when processing data for a safety impact assessment:

- Merging of different datasets which have similar content
- Different datasets which have different content
- Different datasets which are used at different tasks in the process

Many points have to be considered to enable merging, but also to make sure that merging data does not lead to inconsistencies or lower data quality. Therefore, the processing steps stated before have to been performed to align the different data sources to allow a proper merging. Data merging itself is discussed in more detail in Deliverable 4.2.
If data sources are preprocessed and aligned, they can be merged into a broader dataset. Thereby, common key variables have to be identified. Afterward, they must be aligned to allow a joining. If e.g. timestamps are asynchronous, data may be aligned or interpolated to fit the key variables and to align timestamps. Once key variables are matched, data sources can be joined or unified. Thereby, merging conflicts can occur when the same characteristic (as e.g. a crash constellation) is described by a variable or a set of variables similarly. However, since the data must be unambiguous in their significance, such a conflict must be resolved. It is either possible to take one of the values directly and not consider the others, or to use a (potentially weighted) combination of the values. The method should consider the reliability of the data as well as the accuracy, quality, and granularity. The granularity of data also has to be considered for other variables. If for example the sampling rate of two data sets is different, gaps have to be addressed, documented and either data has to be deleted or filled by, for example, interpolation.

Once data is merged into one dataset, data can be processed for further usage. Data merging and processing is not always a clear sequential process but may be iterative due to the type of data. As a result, different cleaning and alignment steps may be necessary as described above.
.

Further processing steps

Several other processing steps highly depend on the data and what the data is used for. Next to the separation of data according to initial requests for projection, economic assessment, and scenario representativeness and weighting information, models for parameters have to be fit and scenarios have to be created for simulation. In particular, individual application of those steps highly depends on set inputs and requirements for the data, so those are described in detail in Deliverable 4.2. A main driver for differences in processing is the choice of baseline approach. Depending on the baseline and model setup distributions or actual values have to be assigned to models. For scenario generation, different scenario designs have to be created. Whereas for baseline approach A concrete trajectories have to be created from individual data points in a one-to-one relation, scenarios for approach B and C have to be handled differently. For approach B, scenarios still rely on actual cases and are transformed and slightly modified to new scenarios. For approach C, distributions have to be calculated first to sample completely new and independent scenarios out of it and to hand them over in the simulation. Another opportunity for C would be the usage of traffic flow simulations. Therefore, models still have to be parametrized and values for infrastructure layout, initial states and more abstract values for e.g., traffic density have to be defined. For all such steps, including scenario generation, model fitting and processing of data for V&V or projection, besides the already set requirements from models and evaluation questions, requirements from the tools have to be considered.

4.3.2.3. Documentation

Another important step is the documentation of data and processing steps. Even if this step does not directly contribute to the Prepare data, it is important to be able to assess the output data. On the one hand, the selected data sources and their quality (e.g., representativeness and accuracy) must be described or links to descriptions have to be given. Further details are described in Deliverable 4.1. This includes relevant meta data and important further information such as trend breaks as e.g. different recording methods or other external factors within a data source. On the other hand, the selection of the data sources must be justified. This is particularly important if the quality of the data has not already been proven good. This may be necessary but should be made clear for transparency. Furthermore, it must be made clear within the documentation whether the selected data meets the data requirements (4.3.2.1). If this is not the case, the gaps as well as limitations arising from them should be discussed. This may also have implications on other parts of the framework.

Since data changes during processing, these changes and underlying assumptions of the processing shall also be documented since the data quality can change within this process. Highly relevant is to document any assumptions (e.g., when different data source are merged or filling in missing information). Stored pre-simulation models and the fitting of data for models shall also be documented. The level of detail of the documentation (see 7.2) depends on the stakeholders addressed (see 3.2) and should allow a repetition of the process.

4.3.3. Input Output

The preparation of data has several in- and outputs. Input wise four categories can be differentiated which are either related to the request of specific data or requirements for those data:

- Evaluation questions: Based on the evaluation questions many data requests arise which have to be taken from data sources within the process step. Those can be subdivided in area and time of application of interest in the safety assessment, choice of baseline approach, selected models and further data requirements resulting from e.g., the formulation of the questions.
- V&V Requirements: Independent of evaluation questions the methodology itself has to be validated. Therefore, the validation step has requirements and requests for data.
- Tool requirements: Further requirements set by the utilized tools, for example on the data output for further processing.
- Preliminary data sources: These data sources are based on an initial selection of potentially relevant (primary or secondary) data sources for the safety assessment. Those data sources are described in detail in D4.1.

Furthermore, the following documentation of the data sources and processing has to be provided:

- Scenarios for simulation execution which can include concrete scenarios directly or scenarios with parameter distributions.
- Concrete values, distributions or ranges requested to specify models.
- Data to generate weights or change distributions to make the utilized sample representative for a specified region.
- Data to project the simulation results in time and space.
- Data to validate the methodology and verify the model representation.
- Documentation of used data sources and any processing steps including limitations and grading of data.

In addition to the outputs mentioned above, there may be some recurring steps after checking availability of data which then may serve again as an input for requirements and needed data.

4.3.4. Consequences

Not relevant for the draft.

4.3.5. Examples

Note: section under discussion with respect to the format in which the examples are presented

The preparation of data highly depends on the availability of data, the scope and further constraints and requests sets by the framework. So, also examples are quite different, and applications can differ. In Table 4-2 examples are given for each of the individual process steps.

Table 4-2: Examples for preparing data steps.

Process step	Example
Extraction of needs	• TODO • (Example here how to come from question to needed data) • (Power calculations)
Choice of relevant data	• TODO (Request including data source description)
Merging of data	• Bärgman, Victor 2020 uses SHRP2-data for road user kinematics and glance data from on-road experiments for further baseline generation. • Kalayci et al. 2020 combines results from various studies. Thereby, questionnaires as well as quantitative data is mapped using triangulations.
Processing of data	• TODO
Generation of scenarios	• Rösener 2020 defines relevant scenario categories based on its defined evaluation scope according to the Challenger model and take GIDAS data for a parametrization. This is used to create probability functions. From those distributions, concrete values are sampled and interpreted within the predefined scenario. • Glasmacher et al. 2023 creates scenarios based on limited amount of real-world data. Therefore, the parameter space is logically constrained by mathematical constraints and causal relations using a graph representation for the scenario. From this representation, scenarios are sampled and can be used for further simulation. • Scanlon et al. 2021 reconstruct accidents from police reports, scene diagrams and photographs. Thereby, trajectories are reconstructed based on pre-crash kinematic, traffic sign states and relevant initial states of the actors
Fitting of model parameters	• Weber et al 2023 using driving simulators to generate data and fit a driver behaviour model. Thereby, the extracted parameters from simulations are mapped to the predefined model.
Description and grading of data	• Leo et al 2019 describes Swedish accident data on bicycle injuries detailed for its assessment. Thereby, the years of accident occurrence, the number of data points and the injury level is described as an overview first. Furthermore, filtering steps are described and argued. Furthermore, that data is compared to another data source from the Netherlands.

Alternative example:
Given a safety assessment evaluation scope which aims to assess the effectiveness of a AEB system for urban following scenarios in Germany at bad weather conditions with a good accuracy, requirements can be extracted directly from the scope since scenarios, the region, conditions and indications about accuracy can be directly derived directly from the scope. Furthermore, baseline approach C2I is assumed from the select baseline step. So, it can be derived, that it should be possible to create distributions for scenario generations and parameters for models with data. Especially in terms of accuracy, those have to be refined to get a proper value for the needed sample size.

With those requirements, data can be searched. GIDAS may be a data source which can fit to the set requirements. Data is requested from GIDAS and it is checked if enough data is available for the given following scenario in Germany. Having the actual sample size, the accuracy can be calculated using power calculations. Furthermore, it turns out that data on weather is not good enough for the given data. Therefore, weather data from a weather institute may be used to say if the rain was strong enough for the given assessment at the accident. To check this, GIDAS data has to be enriched with this weather data first. Therefore, both data types can be synchronized via the timestamps of datapoints. After merging those, the actual number of data points meeting the requirements can be revised and potential accuracy calculations can be adjusted.

Those data points have to be transferred into a parametrizable scenario description. From this one, distributions can be made and scenarios can be sampled out of it for further simulations.

Furthermore, that data may be also used for validation of those scenarios.

Beside the scenario generation data stream, more can be necessary. Since a C2I approach is used, data is needed for agent models. Those may also come from the scenario data, but can also come from e.g. traffic simulator studies to parametrize according to the use case. These data do not have to be merged with the baseline data, but may be used directly for fitting scenarios.

4.3.6. Q&A

Not relevant for the draft.

4.4. Select Models

In the “Model selection” step, the overall simulation structure is defined. This structure consists of models of relevant elements and their interactions during runtime. For each of these models, the required minimum level of detail as well as requirements on the sources of all input signals in terms of quantity and quality should be defined. Based on these requirements, existing models should be provided. If some models do not exist, it must be ensured that those models are generated according to the requirements. Furthermore, model- and structure-related requirements for the V&V process should be defined.
The overall structure, the models within and their required minimum level of detail depend on several influencing factors. Some of these factors are an output of the evaluation scope. Another factor is the chosen baseline approach. And finally, some data aspects need to be considered, like the actual baseline cases and available data for model parameterisation. It might be necessary to perform the “select models” step multiple times for different combinations of influencing factors to determine the best option among several possible (like using two different baseline approaches).

4.4.1. Visualization

Figure 4.14 provides a summary of the process step “select models”.

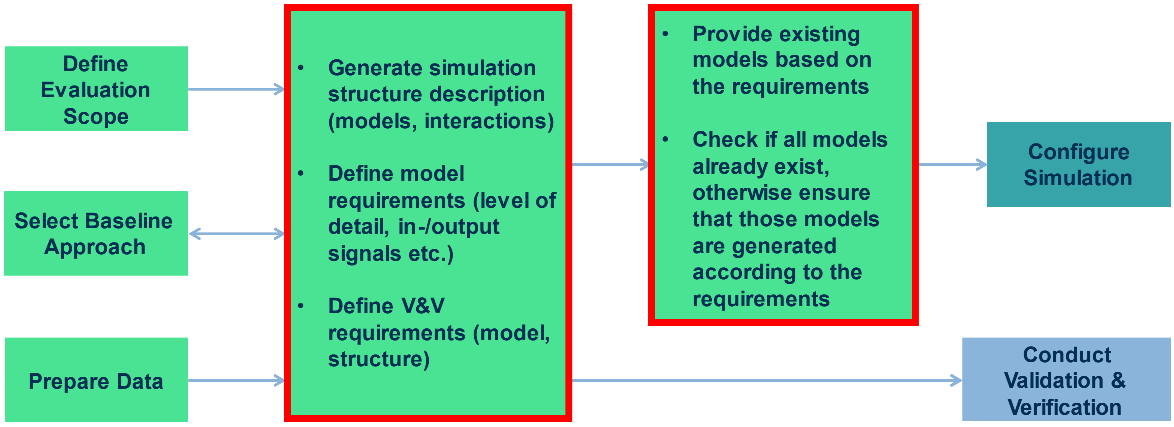


Figure 4.14: Select models process step.

4.4.2. Detailed Description

The detailed descriptions of the various sub-steps in the model selection step are spread over several other V4SAFETY deliverables:

- D4.2 describes how to come to model and V&V requirements related to the baseline simulation.
- D3.1 describes how to select models of behaviour of drivers and other road users, and how to select in-crash models.
- D5.1 describes how to come to a simulation structure and how to select all other models apart from the ones dealt with in D3.X.

Note: V4SAFETY Deliverable that are named above are still under preparation.

4.4.3. Input Output

Input from process step “define Evaluation scope” are:

- Safety measure description (function, intended effects),
- Field of application (addressed scenarios),
- Area and time of application,
- Metrics.

Input from process step “select baseline approach” are:

- Chosen baseline approach.

Input from process step “prepare data” are:

- Actual baseline data,
- Model data.

Output to process step “configure simulation” are:

- Overall simulation structure description,
- (Required level of detail for all models in the structure),
- (Input signals sources requirements),
- Actual models or model components if available.

Output to process step “conduct validation & verification” are:

- Requirements for model & structure V&V.

4.4.4. Consequences

Not relevant for the draft.

4.4.5. Examples

The following example inputs from previous steps are used:

- Input from process step “define Evaluation scope”:
  - Safety measure description (function, intended effects): **AEB pedestrian.**
  - Field of application (addressed scenarios): **Pedestrian crossing from near- and farside, with and without visual obstruction.**
  - Area and time of application: **Urban area, varying weather conditions, current situation, 100% penetration rate.**
  - Metrics: **AIS2+ injury risk.**
- Input from process step “select baseline approach”:
  - Chosen baseline approach: **C2P with stochastically generated trajectories.**
- Input from process step “prepare data”:
  - Actual baseline data: **10000 test cases as xosc files.**
  - Model data: parameters for the individual models, e-g-VuT parameters.

Based on these inputs and the given IT-infrastructure including available tools, a co-simulation-based approach is chosen. The simulation structure is shown in Figure 4.15.

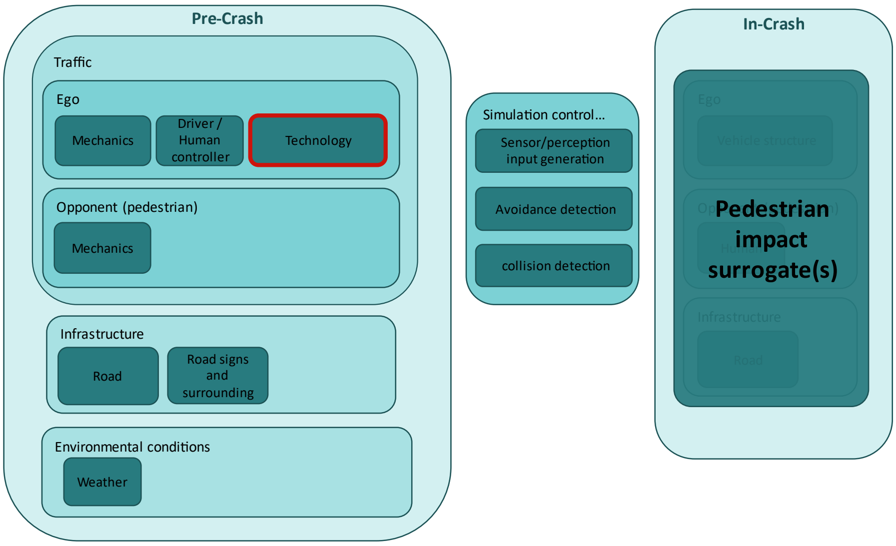


Figure 4.15: Example simulation structure.

Table lists the models in the simulation structure, their requirements, the chosen modelling approaches and (fictitious) links to the respective models.

Table 4-3: Example models and their properties.

Model	Requirements	Chosen modelling approach	Model instance

VuT dynamics	Longitudinal dynamics, controlled by driver and AEB; consider friction provided by environment; Vehicle hull geometry for collision calculation	Point-mass model with brake/throttle input, 2D geometry	Path1/Model1
Sensor	FoV: 90° opening angle, range 80m, no sensor errors to be considered (measurement errors, misclassifications)	Ideal, 2D geometry-based sensor model	Path2/Model2
AEB	If pedestrian is detected and VuT and pedestrian are on collision paths, brake at TTC<1s with maximum possible deceleration	Function logic model	Path3/Model3
VuT driver	Follow the given trajectory	PID controller	Path4/Model4
Pedestrian	Move along given trajectory; bounding box for visibility and collision calculation	Point-mass model, 2D geometry	Path5/Model5
Visual obstruction	Fixed position, hindering sight from sensor to pedestrian	2D rectangle geometry with given length and width	Path6/Model6
Infrastructure	Consider friction provided by environment	2D-plane	Path7/Model7
Environment	Provide tire-road friction coefficient based on weather conditions	Look-up table matching weather conditions with friction coefficients	Path8/Model8
Pedestrian impact surrogate(s)	Determine the head and leg injury severity based on crash configuration and vehicle geometry; needs to be capable of performing calculation in short time (10000 test cases!)	Black-box surrogate model	Path9/Model9
Simulation control	Observe simulation state; check for collisions and start in-crash simulation in that case; terminate simulation if end criteria are met		Path10/Model10

This information (simulation structure and model details) will be used in the “configure simulation” simulation step to set-up the overall simulation model.

4.4.6. Q&A

Not relevant for the draft.

5. Simulation Execution

This main topic deals with execution of the simulation in a practical sense. For this purpose, it uses the input of the assessment preparation to generate the outcome that is further analysed in the next main topic. The four topics of the execute simulation are structured along two dimensions. The horizontal dimension of conducting the entire simulation process – staring form preparing the input to storage of the simulation – is covered by the topics “configure simulation” and “manage simulation”. The topics “Simulated Base” and “Simulate Treatment” cover the vertical dimension by describing the specific aspects for these simulations.

5.1. Configure Simulation

The topic "Configure Simulation" takes place before the simulation runs are started. The pre-simulation work deals with configuration and parametrization of the simulation data and models. Based on the information and requirements from the previous preparation steps, the parametrization of the simulation models must be made. On the one hand, the parameters define the initial and boundary conditions (e.g., velocity, trajectory, road infrastructure) and they also determine the course of the simulation itself (e.g., time step, simulation time, stop criteria). The other part of this task is to configure the simulation structure. This structure and the model settings are important for the later interaction and the flawless running of the simulations. At the end the "Configure Simulation" step leads to a data set for a defined simulation structure that can be executed by the chosen simulation tool.

5.1.1. Visualization

The main process steps of “Configure simulation” are shown in Figure 5.1.

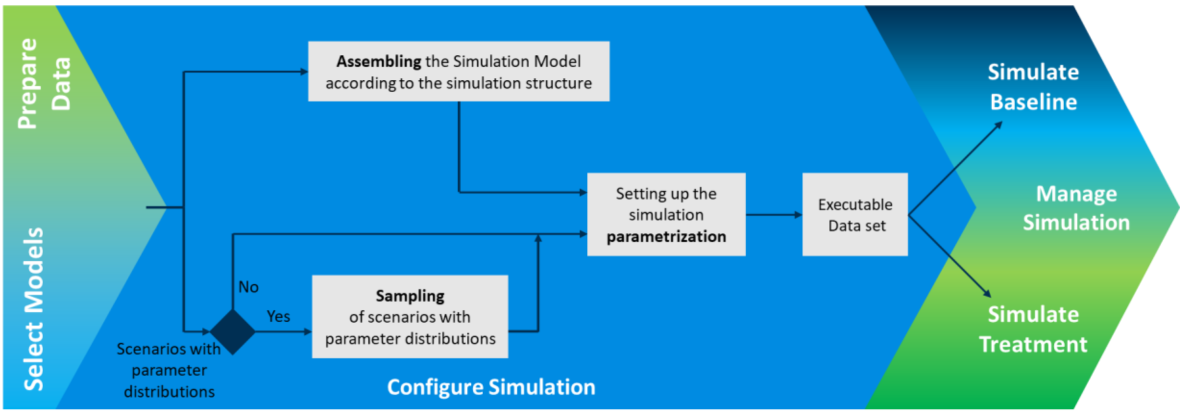


Figure 5.1: Process of Configure Simulation.

5.1.2. Detailed Description

The main tasks of “configure simulation” are to bring the specifications and information from Assessment Preparation into the simulation models and to create the necessary simulation structure.

The simulation structure is defined in “Select Models” and in “Configure Simulation” the simulation models are assembled according to that structure. It is not only the road user models - vehicles, pedestrians, cyclists, etc. - in general that belong to the simulation structure. Detailed sensor models or behavior models of road users can also be included. Furthermore, the infrastructure, the environment conditions, and the selected simulation tool itself are also part of the simulation structure. Depending on the evaluation scope, in-crash models and simulation controls to switch from pre-crash to in-crash simulation might be necessary. More details about the simulation structure can be found in the V4SAFETY deliverable D5.1.

According to Figure 5.1 the parameters of the concrete scenarios used must be obtained. The question of how many cases should be simulated is answered by the power calculation which is part of the "prepare data" step. If the scenario description from "prepare data" is available with concrete values, no intermediate step is necessary to parameterize the simulation models. However, if the scenario description is available as a parameter distribution, the concrete scenarios must be generated using suitable sampling methods. If random number generators are used, the seeds should be stored to obtain reproducible results. This applies to the sampling methods to obtain concrete scenarios but also to the application in in-simulation models, e.g., behavior models for pedestrians.

After the simulation structure is created, the number of concrete scenarios is defined and the corresponding parameter values are determined, the parametrization of the simulation will be set up. This means that all static parameter values like initial- and boundary conditions or parameters defining the simulation run itself are set. If the simulation tool does not independently keep the units consistent in all sub models, care must be taken when setting the parameters to ensure that the identical units, for example for force or time, are used everywhere. For reasons of reproducibility, it also might make sense to specify the number of processors to be used in the simulations.
When talking about initial and boundary conditions, think about

- time step, simulation time, stop criteria, …
- velocity and trajectory of a traffic participant
- number of lanes, curvature of the road, traffic light, sidewalk, …
- weather, trees, objects, …
- parameter(s) defining the behavior model(s)

In the parametrization step also the results which should be written out and the output format must be defined.
At the end of the “configure simulation” step the result will be a data set (consisting of one or multiple files) defining the simulation structure, models, parametrization, outputs, initial conditions and boundary conditions which can be executed by the chosen simulation tool. Afterwards it is possible to start the baseline and treatment simulations using the process defined in “manage simulation”.

5.1.3. Input Output

Input

- Defined evaluation scope.

Input from Prepare Data

- Data matching to the baseline approach:

        - Scenarios for simulation execution which can include concrete scenarios directly or scenarios with parameter distributions.
        - Data for representativeness of scenarios / Power Calculation.
- Concrete values, distributions or ranges requested to specify models.

Input from Select Models

- Overall simulation structure
    - Simulation tool
- (Required level of detail for all models in the structure)
- Input signals sources requirements
- Actual models or model components if available

Output

A data set (consisting of one or multiple files) defining the simulation structure, models, parametrization, outputs, initial conditions, and boundary conditions which can be executed by the chosen simulation tool.

5.1.4. Consequences

Not relevant for the draft.

5.1.5. Examples

Will be added at a later stage of the project once the V4SAFETY examples are ready.

5.1.6. Q&A

Not relevant for the draft.

5.2. Manage Simulation

In contrast to the topic "configure simulation" which addresses the pre-simulation work, this topic deals with the execution and post-processing of simulation in terms of the process. While the execution of the simulation deals mainly with monitoring of the process, the post-process covers the handling of the resulting simulation data. The major objective of this step is to ensure that the simulation runs flawless and the data for the post-process is provided correctly.

Note: An update of the section will consider more explicit the input of configure simulation (including related simulation setup and decisions) in the data storage part. The update is intended to underline that the results must been seen always in the light of the input.

5.2.1. Visualization

The process of the topic manage simulation which is split into two subprocess – namely simulation execution and data storage – is shown in Figure 5.2

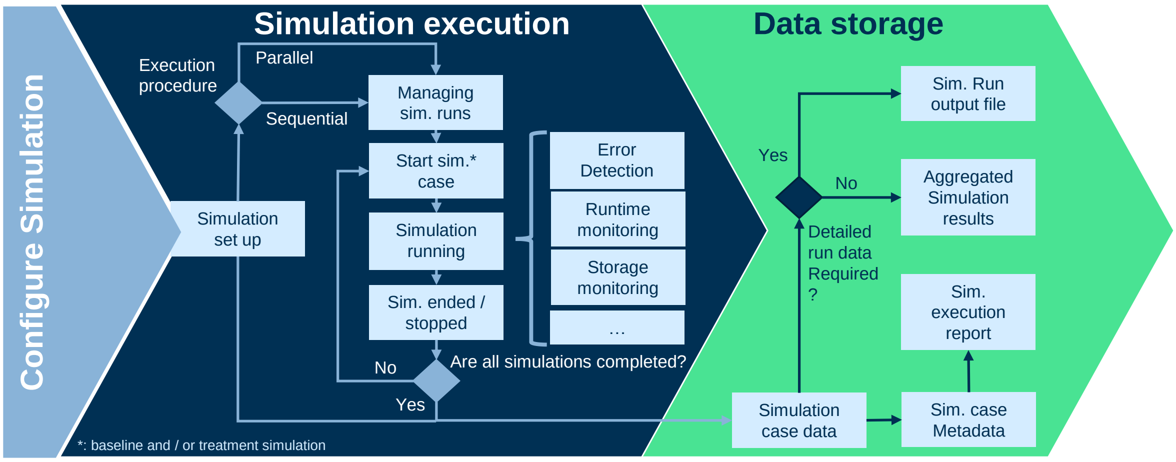


Figure 5.2: Process of the topic "Manage Simulation".

5.2.2. Detailed Description

The "manage simulation" topics deals with the simulation execution and the storage of the simulation output. The process that is described in the following applies for the baseline simulation as well as the treatment simulation.

The initial point of the process is the simulation configuration and the work in the simulation preparation. In the latter one defines beside other aspects how many cases should be simulated. Next to this, the available IT-infrastructure as well as the simulation tool are additional obligatory inputs. All three components define the simulation setup. Depending on the simulation setup it needs to be decided whether the simulation of the individual cases should be done in sequential order or in parallel. The first option is in terms of implementation effort the less demanding measure. However, it requires typically more time. For the simulation execution in parallel different implementation measure are feasible. Also, combinations of the sequential and parallel execution are feasible.

In the next step, the simulation of the individual cases is triggered according to the decided process. During the simulation it is important to monitor the simulation constantly. The monitoring should be an automated process. The objective is to ensure that the simulation runs flawless. Therefore, it is reasonable to implement checks which could be dynamic or static checks. Examples for these checks are:

- Error Detection: Should check and identify any calculation error that occurs during the simulation. In particular cases in which calculations end up with values that are out of the specified range or provide an infinite result need to be detected and reported.
- Runtime control: This module should monitor the simulation during runtime to identify cases in which a simulation takes much more time than expected. This can be an indication for errors in the simulation.
- Storage monitoring: The module should check the usage of the internal storage constantly to prevent an overflow of the storage. This could lead to a breakdown of the simulation or even the whole computer.
- [List will be continued based on later discussions]

If the simulation is finished (successful or aborted due to an error), it shall be checked whether all simulation cases are completed. If not, the simulation needs either to wait for the other simulation cases to be finished or trigger the execution of the next simulation case. In case the simulation was aborted due to an error it should be decided after the complete simulation, whether the simulation should be repeated (with necessary modifications).

The second stage of this topic is the handling of the simulation output. Here, it needs to be decided which output is required. If only a complete simulation report is required, the individual simulation case outputs need to be merged. This can be done within the simulation tool or in a separate step in the post-processing (potentially with a different tool). In the latter case the simulation output per case needs to be provided. Here, it is important to ensure that the output is provided in a commonly used format that is compatible with other tools. Proprietary formats are in this sense not ideal since they limit the exchange possibilities between partners. However, such format could be needed due to specific purpose (e.g., reduce storage). Besides the simulation results that provide the results in the request metric or the trajectories of the simulation agents, a report on the simulation execution should be provided to summarize information such as number of successful simulations, detected error(s) etc.

5.2.3. Input Output

Inputs for this topic are:

- Configuration for the simulation (incl. number of baseline cases).
- Simulation tool.
- Available IT-Infrastructure.

Outputs of this topic are (depending on requested reporting level):

- Report on simulation execution (aggregated).
- Aggregated simulation results.
- Simulation results per simulation run in defined format.

5.2.4. Consequences

Not relevant for the draft.

5.2.5. Examples

Will be added at a later stage of the project once the V4SAFETY examples are ready.

5.2.6. Q&A

Not relevant for the draft.

5.3. Simulate Baseline

All preprocessing and preparation topics in the framework end up in the topic “Simulate Baseline”, where configured simulation is executed, to produce simulated cases for later analysis. This topic involves on the one hand (obviously) starting the simulation execution, using the pre-defined baseline simulation configuration. On the other hand, checks are relevant within this topic. The latter are relevant both before and after the simulation execution, to assure sufficient precision and accuracy of the simulation output, for the evaluation of the safety assessment of treatment. Furthermore, if issues with the simulation are detected prior to execution of the (at times) long simulation durations, these process steps also enable efficient progress towards the safety assessment results.

5.3.1. Visualization

See Figure 5.1 and Figure 5.2.

5.3.2. Detailed Description

The "Simulate Baseline" topic in the V4SAFETY framework has the objective to create reference data, known as the baseline, by means of simulation. The baseline data are what the treatment is compared to, to estimate the safety benefit of the safety measure under assessment. The main goal of the topic ”Simulate baseline” is to produce accurate and trustworthy results based on the parameterized and configured simulation structure from the “Configure simulation” framework component. Valid and trustworthy simulation data in this topic ensures that the baseline data can be effectively compared with the results from the "Simulate Treatment" topic. Since the data are prepared and models configured already in prior topics, the primary focus here is to ensure the simulation meets the Key Performance Indicators (KPIs) set during the assessment preparation. To assist users in achieving a high-quality simulation, checklists based on literature, past experiences and common mistakes are provided. Another main aspect of this step is to initiate the execution of the baseline simulation based on a simulation configuration performed in previous process steps.

Some example checks relevant for the “Simulate Baseline” topic are shown below:

- Are the timings between road-users interaction in the simulation as in the original data?
- Is the scenario generation creating crashes across the entire spectrum (low/property damage, to high severity), similar to reality?
- Does the baseline simulation reproduce the given baseline data (trajectories, crash configurations, ...) accurately enough?
- Are traffic safety related parameters close to real world data? (Crash rate, injury severity, angle of impact, crash velocity etc...)
- If random processes are included in the simulation, are the seeds for the random generators set for each case so that the simulation becomes reproducible?
- Do the traffic flow simulations lead to realistic/expected traffic flows?

5.3.3. Input Output

Input:

- A data set (consisting of one or multiple files) defining the simulation structure, models, parametrization, outputs, initial conditions and boundary conditions which can be executed by the chosen simulation tool for the Baseline from Configure Simulation

Output:

- Simulation data for the defined KPI metrics
- Results of Simulate Baseline checks

5.3.4. Consequences

Not relevant for the draft.

5.3.5. Examples

Will be added at a later stage of the project once the V4SAFETY examples are ready.

5.3.6. Q&A

Not relevant for the draft.

5.4. Simulate Treatment

The safety performance of a road safety measure is determined by comparing outcomes from the baseline and treatment simulations. The baseline for the assessment is the simulation without the measure present as part of the simulation model while the treatment is the simulation with the measure present as part of the simulation model.

There exist different types of safety measures as treatment which can be simulated like in-vehicle technologies, infrastructure measures and behavioural, for examples look into “Examples” section.

5.4.1. Visualization

See Figure 5.1 and Figure 5.2.

5.4.2. Detailed Description

The "Simulate Treatment " activity in the V4SAFETY framework has the objective to generate results by application of treatment, that are comparable to the baseline as reference data, by means of simulation. The treatment data are compared to the baseline data, to estimate the safety performance of the safety measure under assessment. The main goal of this step ”Simulate treatment” is to produce accurate and trustworthy results based on the parameterized and configured and parametrized simulation structure of earlier settings in the assessment process from the “Configure simulation” framework component. This ensures that the treatment data can be effectively compared with the results from the "Simulate Baseline" phase. The main aspect of this step is to actually start the execution of the treatment simulation using the process described in “Manage Simulation”, based on a simulation configuration set in the “Configure Simulation” step.

Depending on which type of measure is addressed, simulation of treatment may involve alterations ("alterations" should be understood as "with regard to what was used in simulation of baseline") of the driver model (regulations, behaviour-targeting measures, perception-targeting measures), alterations of the infrastructure or traffic models (e.g., measures targeting other traffic such as protective equipment for VRU). The most common type of simulation of treatment, however, will concern vehicle-related measures (whether involving full on-board technologies or not) and involve vehicle-infrastructure and/or vehicle-driver interaction models - and in many cases models for actual actions on the vehicle’s commands.

5.4.3. Input Output

The input is the data set (consisting of one or multiple files) defining the simulation structure, models including treatment, parametrization, outputs, initial conditions and boundary conditions which can be executed by the chosen simulation tool from Configure Simulation.

The output are simulation results containing all relevant information (including safety performance metrics defined by Configure Simulation) which are comparable to the simulation results of “Baseline Simulation” and additional metrics for the verification of the treatment.

5.4.4. Consequences

Not relevant for the draft.

5.4.5. Examples

Will be added at a later stage of the project once the V4SAFETY examples are ready.

5.4.6. Q&A

Not relevant for the draft.

6. Assessment Analysis

The analyse assessment presents the topics that follow process-wise once the simulations are conducted. These topics evaluate the safety performance that discussed possible metrics and their calculation, the time- and space-wise project of results as well as the conduction of cost-benefit analysis which is often the final step of the virtual evaluation. However, it must also be noted that projection and cost benefit analysis not always required step in virtual assessment.

6.1. Evaluate Safety Performance

Evaluating the safety performance of a safety measure consists in comparing safety outcomes in a situation where the safety measure is present – to a given extent – in the field, to safety outcomes in a reference situation (baseline) where it is not present at all. Safety outcomes are computed through metrics defined in the research question, which can then be summarized as crash reduction rate, or injury reduction rate. These can also serve for further cost or socio-economic assessment when those are in the scope of the study.

Depending on the safety measure under scope the analysis could be related to crashes or to near crashes. Appropriate metrics for assessment shall be considered for each, making sure that both are available in baseline and treatment results, so they can be compared. In some cases, e.g., if surrogate safety measures for near crashes are in scope, it is common that metrics need to be calculated from the baseline and treatment results. Differently, in the case that injury probability is in the scope and changed crash parameters are observed from baseline and treatment results, Injury risk functions may need to be considered alongside the set of relevant crash related metrics.

6.1.1. Visualization

Figure 4.9 provides a summary of the process step “Evaluate Safety Performance”.

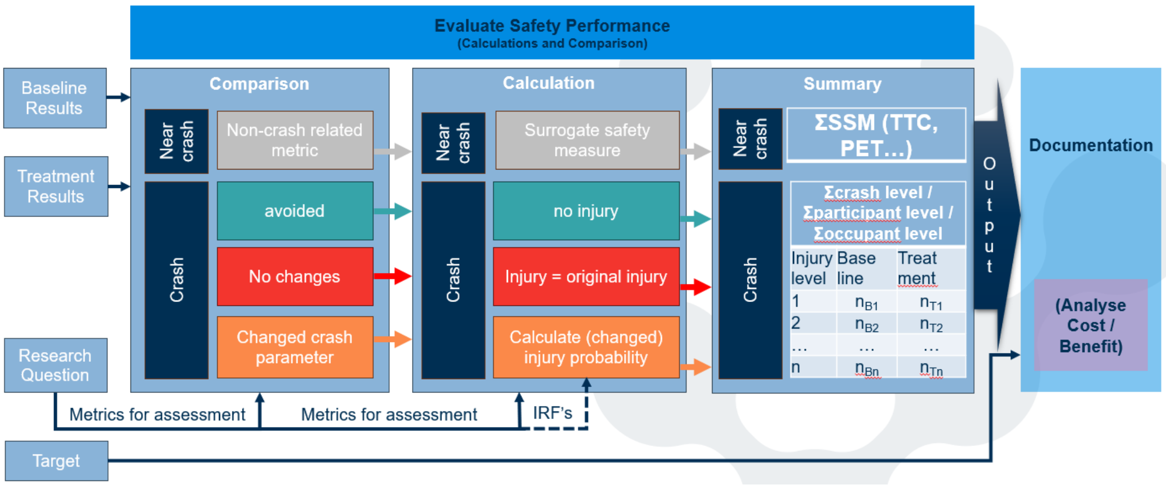


Figure 6.1: Evaluate Safety Performance.

6.1.2. Detailed Description

The evaluation of safety performance deals with the assessment results of the safety measure under study. This steps uses the baseline simulation and treatment simulation results, and compares them based on metrics that are defined in the research questions. Some of these metrics may be obtained directly from the simulation results. For example, an assessment of a safety measure aiming at crash avoidance could be quantified from the amount of crashes avoided when the safety measure is considered against the quantity of crashes without the safety measure. However, other metrics may need to be calculated and therefore require post-processing of the simulation results obtained.
In this step, it is necessary to distinguish that the evaluation of safety performance of a safety measure may be related to crashes or to non-crashes or consider both.

The mandatory input data of the evaluation of safety performance are the output data of the baseline simulation and the treatment simulation. Additionally, based on the research question, relevant metrics for safety assessment have to be identified. In the case that the assessment is related to injury level, injury risk function as input parameter is required. If the documentation of the safety performance evaluation should consider a target out of the research question, the target is required as input parameter as well.

6.1.2.1. Crash related assessment

The crash assessment focuses on crash related metrics, such as collision speed or energy equivalent speed (EES). These metrics are typically compared between baseline and treatment. The metrics itself are derived from the kinematic data of the simulation (position, speed, orientation etc. of crash involved traffic participants). The assessment delivers the three possible results / outcomes:
avoided,
no changes,
changed crash parameter.
Depending on outcome of the comparison step, in the calculation step the following operations are defined
If the treatment simulation avoids the collision, there is no injury / crash anymore and no injured person expected
If there are no changes between baseline and treatment simulation, there is no change for the crash outcome (injury level, material damage) expectable
If the crash parameters changes, a recalculation of the injury probability is required. If this part is required, injury risk function(s) as input parameter are mandatory
In the summary step, a summary of the number of crashes / participants / occupants on several injury level for the baseline and the treatment simulation is operated. The injury level definition can be “fatal”, “sever injured”, “slightly injured”, “not injured” or related to the MAIS scale or free chosen / defined by the research question or the used injury risk function.

6.1.2.2. Non-crash related assessment

In the non-crash assessment, it is less obvious to find a meaningful safety benefit of a function, such as crash avoidance, injury avoidance or injury reduction. Therefore, it is needed to use surrogate measures, that can allow understanding the effectiveness of a safety measure. In terms of road safety, a surrogate measure should allow converting non-crash events into a corresponding crash frequency and /or severity (Tarko et. 2009 Ref). Some examples of these measures can be time to collision (TTC) and Post-encroachment time (PET). Such measures may need to be derived

or calculated from non-crash related metrics such as distance travelled or speed. The comparison between baseline and treatment for such surrogate safety measures is performed. The fields of “avoided”, “no changes” and “changed crash parameters” are skipped.
Examples for crash metrics and safety measures of both types of assessments are given in the example section of this chapter (see 1.18.5).
After having calculated the safety metrics, they shall be summarized and presented together with the evaluation objective. Generally, the safety performance (P) of a safety measure will be addressed considering its Injury severity (I) and exposure, also denoted as frequency of occurrence (f), for the addressed scenarios.

$$P_{\mathrm{i}} = I_{\mathrm{Treatement,i}} \cdot f_{\mathrm{Treatment,i}} - I_{\mathrm{Baseline,i}} \cdot f_{\mathrm{Baseline,i}}$$

For crash related assessment, the change in severity may be obtained using Injury Risk Functions that can use as input both baseline and treatment simulation results (e.g., collision speed, or EES). When it comes to non-crash related assessment, the change in severity may be obtained by comparing the non-crash related metrics such as TTC, with the same metric measurement in crash data, so that a relation with injury can be obtained.

The exposure can be analysed for crash related assessment in terms of amount of crashes per driven kilometre. This measure shall be treated with special care, as crash number as well as driven kilometres should be representative of the same addressed scenario. For non-crash related assessment, the exposure can be also determined based on the frequency of occurrence of the addressed scenarios per driver kilometres.

The output of the crash and non-crash related assessment is – depending on the type of assessment – presented as result or used for further analysis steps. These analysis steps are the project of results (see chapter 6.2) and / or a cost benefit analysis (see chapter 6.3).
The last step of the evaluate safety performance step is the mandatory documentation. The documentation should consider the target of the research question, the target values shall be an input for the documentation. If the research question required a cost benefit analysis, the output data of the summary are required as input parameters for the following cost benefit analysis.

6.1.3. Input Output

The input data of the step “evaluate safety performance” is derived from the output of steps “Simulate baseline” (see chapter 1.16), “Simulate treatment” (see chapter 1.17) and “Define Evaluation Scope (see chapter 1.10). Both “Simulate baseline” and “Simulate treatment” will provide simulation results that can be treated under this step. Additionally, “Define Evaluation Scope” will provide the Research Questions from which, relevant metrics for safety assessment shall be identified. Depending on the research question, the metrics could be related to either:

- Crash parameter (e.g., EES, DV, impact speed, etc.)
- Non-crash parameter (e.g., TTC, PET)

and shall be identified for both baseline and treatment simulation. It shall be noted that the metrics for safety assessment may not need to be included in the baseline and treatment simulation results, so it may be required to calculate them from the simulation results.

If it is required to calculate a changed injury probability based on changed crash parameters, injury risk functions are required as additional input data.

As output, this step shall document:
The identified metrics for safety assessment, including the reason for its selection.
The outcome of comparison between baseline and simulation results, highlighting the main difference between both.
The outcome of comparison between safety assessment metrics for both baseline and treatment results. These results may be documented based on the appropriate metrics for the scope of the safety measure under assessment, which means either crash related metrics or non-crash related metrics.
The assumptions and limitations identified in this process, making reference to those also documented in other steps which are part of the sections "Prepare Assessment" and "Execute Simulation", since this step is showing the results, therefore clear messages of under which conditions these have been obtained shall be clearly stated.

Additionally, when cost/benefit assessment is part of the evaluation scope, the results obtained will be an output of this step that will serve as input for the step "Analyse Cost/Benefit (see chapter 1.20).

6.1.4. Consequences

Not relevant for the draft.

6.1.5. Examples

Example for the metrics used from ISO21934-2 (ISO 21934) are:

- **Time to collision (category: none crash related metric)**
  The TTC describes the remaining time until an imminent collision is going to happen if the movement (direction, velocity) of the involved road users does not change. The metric, which is often also used as a trigger criterium in in-vehicle safety measures, aims to determine the criticality of a driving scenario.
- **Number of collisions (category: crash avoidance metric)**
  The number of collisions calculated the number of simulations runs that led to a collision between two traffic participants. Typically, also a road departure of a single traffic participant is counted as a collision. This metric is typically used when the collision avoidance potential of a safety measure should be assessed. Oftent the number of collisions is set in relation to the number of simulation runs to obtain a collision rate.
- **Number of victims (category: crash injury metric)**
  The number of victims (n_vict) counts the number of simulations for which collision was detected and for which at least one for one of the traffic participants a specified injury criteria (e.g., seriously injured or MAIS 3+) has been reached. The definition of victim must be the same in the baseline and in the treatment. To determine the consequence of a collision often injury risk functions used that link the kinematic condition to a probability of a to be expected injury outcome (for example impact velocities, impact location, age, or height of involved vulnerable road users, Energy Equivalent Speed of the vehicles involved, or change of velocity during collision). Injury risk functions are commonly derived from statistical models, such as regressions, over populations of collisions. There are different injury risk functions available in the literature. They differ in terms of used data source, addressed collision type, considered road user, predicted injury level. Therefore, the choice of the applicable IRF depends on the research question, and the consistency of this choice should be checked.
- **Impact velocity (category: change crash parameter metric)**

The impact velocity is the velocity of a certain road user immediately prior before time point of collision to impact that is describe by time point of the collision (i.e., in simulations this is typically the last simulation step prior to the collision).

6.1.6. Q&A

Not relevant for the draft.

6.2. Project the Results

Prospective safety performance assessment using virtual simulation aims at predicting the impact of safety measures before their introduction or before sufficient data is available to conduct a retrospective assessment.

Safety performance assessment is usually done on samples that have been selected / collected using a set of criteria. These criteria may induce biases towards the target population. Therefore, the representativeness of these samples regarding the target population must be checked and eventually methods have to be applied to correct for these biases. For this process those characteristics are used that show a significant difference between the sample and the target population. Examples of these characteristics are traffic composition, conflict or crash type, pre-crash maneuvers, injury severity, and traffic participant speeds. This, however, requires that the selected characteristics in samples and target population follow a common definition.

The market penetration of different types of safety measures (in-vehicle, infrastructure, regulatory behavior) and their observed effects on road safety can vary greatly: in-vehicle safety measures may take up to decades to spread over the whole fleet whereas infrastructure-related measures may have an immediate effect on all traffic participants in the relevant areas of applications. Regulations may have an immediate or a stepwise effect, depending on their type, enforcement, and acceptance. The implementation and effects of measure may also depend on country economy profiles, and on whether they belong to the high-, middle- or low-income countries group. Measure costs and discounts, influential on penetration rates, will also vary over time.

Since traffic is subject to constant change, projecting the sample data to a future point in time using current distributions of characteristics may not be sufficient. The methodology must then incorporate factors that may come along with future changes such as variations in fleet size, the introduction of new means of transport (e.g., micro-mobility, autonomous shuttles, or higher levels of automation for passenger vehicles) or other reasons for changes in the modal choice by road users (e.g., fiscal incentives).

6.2.1. Visualization

Figure 6.2 provides a summary of the time projection of results process.

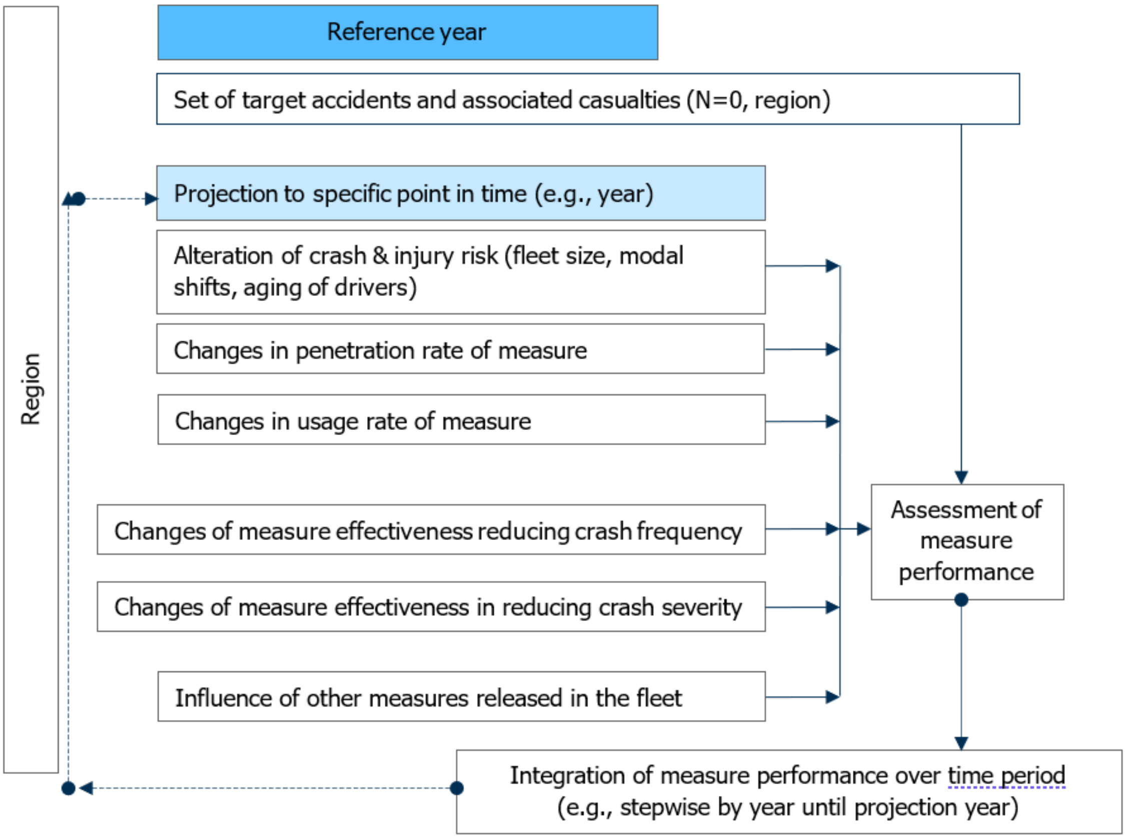


Figure 6.2: Time projection of results.

Estimation of performance might be independent of point in time and just consider a given market penetration rate (e.g., 100%) regardless of the number of years necessary to reach it. If assessment of market penetration over time is available, the projected performance estimation might be linked to specific points in time and done stepwise e.g., year by year.

Figure 6.3 provides a summary of the space projection of results process.

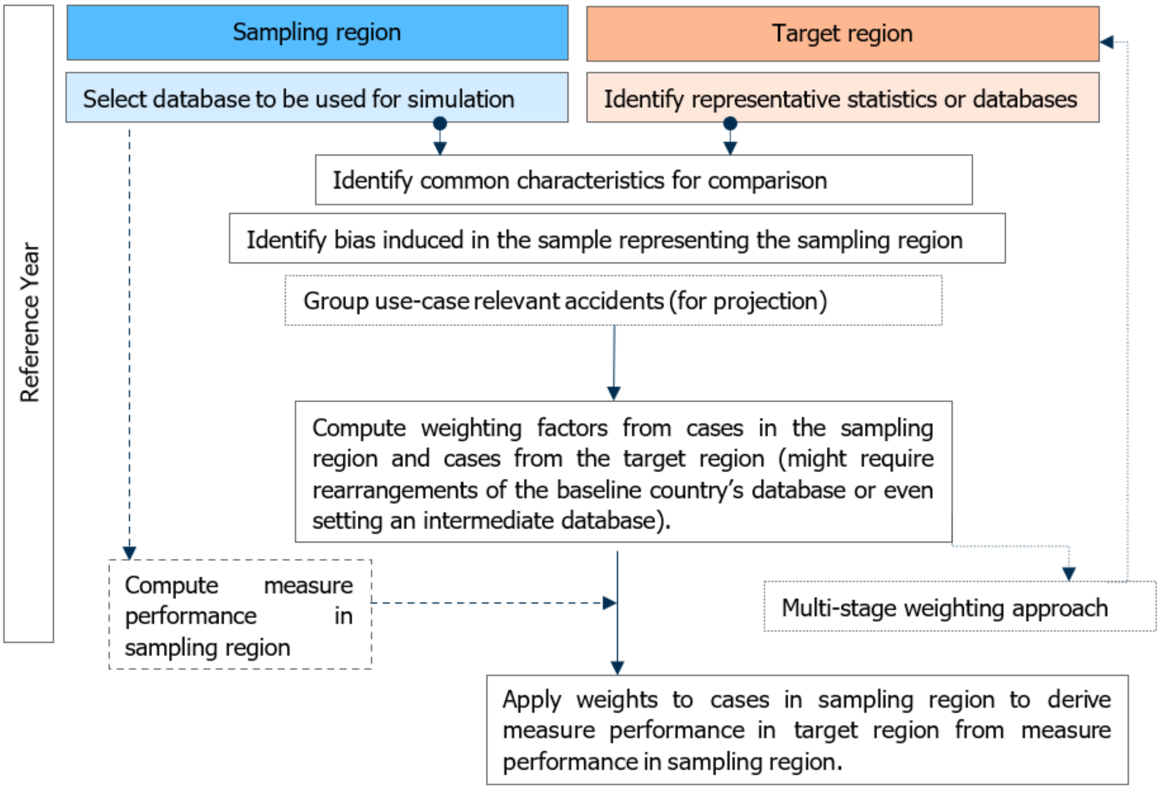


Figure 6.3: Space projection of results.

6.2.2. Detailed Description

In the above visualization schemes, measures can consist of in-vehicle systems, communication systems between the vehicle and its environment (e.g., other traffic, infrastructure), infrastructure alterations or regulations.

Active safety in-vehicle systems aim to reduce crash frequency (preventive effect) but may also provide a mitigating or protective effect (unavoidable crashes will in most cases be less violent). Passive and tertiary safety (e.g., e-call) in-vehicle systems have a protective effect as they aim to reduce the consequences of a crash, although for of tertiary safety measures the effect is not a direct one as they only aim in having rescue arrive quicker on site and thus avoid injury severity to degrade.

Crash prevention and injury mitigation can also be achieved by infrastructure related measures: improved lighting or VRU dedicated lanes can be seen as active, crash-preventing systems ; guardrails or padded guardrails are examples of passive, user-protecting systems.

Regarding regulations, one can think of similar examples: prescribing that all new vehicles go through a new type of crash test will have effects on the vehicles structure, thus have a protective effect (influence on crash severity). Enforcing regulations on driving under the influence will have a preventive effect on crash frequency.

Recalling that the benefit of measures is composed of:

Relevance x **Effectiveness** x **Usage Rate** x **Market Penetration** = Benefits
(Source: GDV)

one can identify the different effects to be taken into account when projecting results.

- **Relevance** covers the crash and injury risk relevant to the measure(s) under study, that can be seen as the **size of the accident scene theoretically addressed by the measure** (and the associated injuries); when projecting over time, it is necessary to think of alterations of relevance due to changes in fleet sizes (due to e.g. government fiscal incentives, catastrophic events such as pandemics) and modal shifts to other means of travel – the latter also being a source for new types of accidents and new measures; the relevance of a given measure can also be influenced by the penetration of other measures : for instance one can think of a generalized alcohol-lock device, that would effectively suppress a fair proportion of accidents due to speeding or of single-vehicle run-off roads. When projecting over space, assessing relevance in both the sampling and the target region(s) is one of the primary tasks.

- The intrinsic **effectiveness** of the measure on its intended domain of operation. Typically, this is the result of a **simulation of treatment** activity. Over time, effectiveness can be influenced by technical upgrades of the measure or adaptation of the driver behaviour to the presence of measures. Examples of the latter include better compliance to speed limits in the presence of automated speed cameras or less attention paid to driving task on account on more automated systems on board. Behavioral factors such as aforementioned compliance or complacency might also be considered when projecting over space, as regional attitudes may have an influence on effectiveness. The effectiveness of a given measure can also be influenced by the penetration of other measures: for instance, one can think of the influence of a constraining ISA over AEB for Vulnerable Road Users. By reducing the number of conflicts in which the driver has no time to react when faced with a VRU, ISA would influence AEB effectiveness. For regulations concerning human behaviours, it is the quality of law enforcement that will have an effect (this is covered by "upgrading technical characteristics" in figure 6.2's flow). As an example, one can think of posted speed limit being enforced by police forces at a few daily varying points of the road network, then – when safety measure became available and reliable - by thousands of efficient automated speed cameras.
- The driver **usage** for some in-vehicle measures (e.g., LKA that can be switched off by drivers), which may very over time, over regions and even over age classes in the same region at reference point in time.
- The **penetration or installation rates of the measure**, which will vary over time. Mandating systems (e.g., ADAS) will influence their penetration rates over time. Differing socio-economic conditions between sampling and target regions also have an influence on penetration rates and have to be taken into account when projecting over space.

For reduction of bias, weights can be computed that are assigned to each case (e.g., hypercube weighting, decision-tree weighting, inverse probability weighting, etc.) – this goes mainly for baseline approaches A and B, but also C – but also by altering distributions from which cases are derived (approach C). This applies for either weighting over time or over space. Often a multi-stage weighting is a suitable approach if direct projection is not possible.

Projection over time and regions cumulates the difficulties of both type of projection, especially in terms of inputs (see corresponding section hereafter). Although many attempts have been made in that direction, we recommend not to attempt this type of projection at "averaged" level but to keep it at country level, on account of possible wide differences in economic status, hence of fleet renewal. The EU status in this regard is highlighted in figure 6.4.

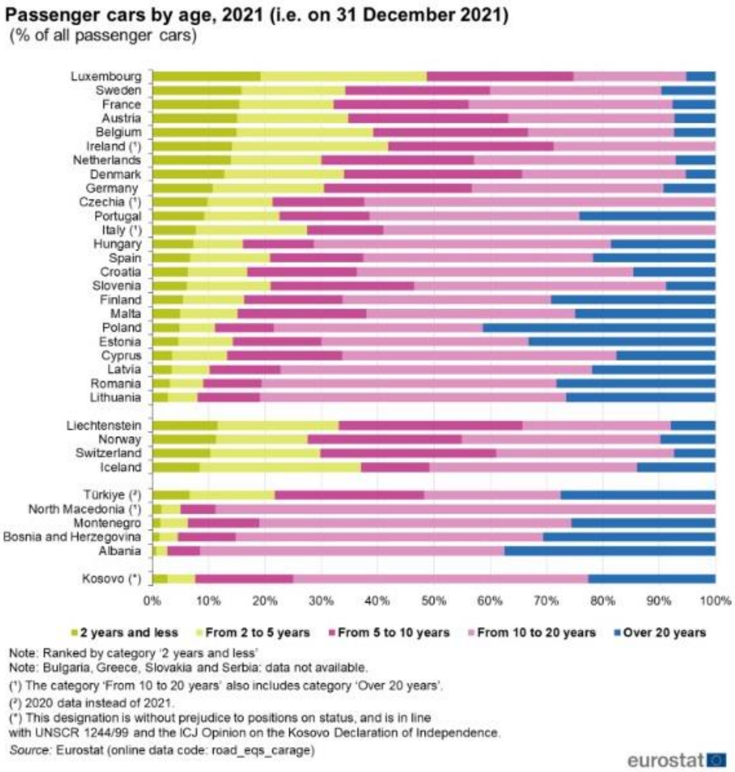


Figure 6.4: EU fleets by age (Source EuroStat 2021).

6.2.3. Input Output

Naturally, the main input to the "projection of results" activity is the result of measure performance in the reference region, at the reference point in time. Other necessary inputs are listed hereafter.

For time projection, inputs depend on the evaluation scope and usually include:

- Car registrations and fleet size over time. At European level, this is readily made accessible online, by e.g., by the "Association des Constructeurs Européens d'Automobiles" (ACEA).
- Penetration rates of the targeted measure over time. This is not easily recovered, the main sources for in-vehicle systems being suppliers thereof. Penetration rates are often given as a percentage of equipped new registrations, therefore necessitating a "bridge" function between new registrations and the global fleet.
- At European level, the "bridge" mentioned here above can be derived from the ACEA published "vehicles in use" tables. These can be used directly or used to derive a vehicular survival curve (giving the probability of a vehicle's survival in the fleet, given its age).

For more precise time projection assessment, one will need, as inputs:

- Modal shifts models, that will give either the proportion of trips or distance travelled that is substituted from "old" travel modes (e.g., cars, public transport) towards new mobilities (e.g., bicycles or e-scooters)
- Data on driver behavior regarding the usage of e.g., switchable in-vehicle systems and also on driver adaptation to measures (e.g., compliance to posted speed limits, less attention paid to driving task on account on more automated systems on board etc.) that may have an influence on measure effectiveness.

For space projection, the main inputs are:

- Distributions of crash configurations (type classification),
- Distributions of victims by crash configurations
- Possibly distributions of accident locations

for both sampling and target regions cases. Regarding the latter, it is important to get target use cases categorized in the same way as in the (in-depth) baseline database that will be used for the simulation. At European level, CARE is the only available target database containing aggregated characteristics of the European Union. However, it has shortcomings, as not all countries use the same accident categorization or do not provide the same level of detail on crashes. Projection to some countries might then be very approximative or not be feasible at all.

Secondary inputs to space projection consist, depending on the research question, of:

- Statistics on law enforcement, attitude to road safety (e.g., from ESRA study)
- Socio-economic statistics (e.g., fleet size, renewal rate of the fleet, segmentation thereof)
- Demography statistics (e.g., age structure of the driving population)
- Road type statistics (e.g., length of road network by road type)

for both sampling and target regions cases. These may influence the case weightings from sampling to target region.

Output to projection of results is, generally speaking, the assessment of the injury burden over time or space. Those can be coupled with an economic valuation to get the socio-economic benefits of measures.

6.2.4. Consequences

Not relevant for the draft.

6.2.5. Examples

Examples for different projection approaches are given in the following table.

Table 6-1: Examples for different projection approaches.

Projection scope	Reference
Projection over space	• Niebuhr et al. (2013) described a methodology, which leads to a GIDAS-aided quantification of the effectiveness of traffic safety measures for passenger vehicles for countries in EU 27. Even based on rather limited accident information from national statistics and under the assumption that comparable accidents in different countries lead to comparable accident outcomes, the described procedure allows defining weighting factors. By weighting each single accident of GIDAS a modified GIDAS database could be established which imitates the accident situation in the region or country of interest to some extent. The main ingredients of the proposal are a proper clustering of European countries according to their accident occurrence and a statistical procedure (Iterated Proportional Fitting) which allows the prediction of the common distribution (high dimensional) of accident data of the region or country of interest based on available lower dimensional marginal distributions (even one-dimensional). • Chanove et al. (2022) used simulation files based on police-recorded accident data to recreate pre-crash accident scenarios. While this dataset is limited to Saxony, Germany, this paper proposes an extrapolation method to adapt it for broader application, such as assessing safety systems at a European scale. The approach involves reviewing the simulation files and linking them to European accident data, followed by extrapolation using weighting factors. Challenges arise due to differences in data categorization among countries' accident statistics. To overcome

	this, a harmonized database of police-recorded accidents is created to facilitate grouping based on similar variables. This process enables extrapolation to a European or country-specific level, providing a framework for calculating weighting factors by establishing consistent input data and variable groups.
Projection over time	• GDV, Themenschwerpunkt, „Automatisiertes Fahren, Auswirkungen auf den Schadenaufwand bis 2040“ (Redlich et al. 2022) exposes a simplified method to derive time evolutions of benefits induced by the introduction of active in-vehicle systems.
Modal shifts	• Stipdonk et al. (2012) described and applied a method to assess the effect on road safety of a modal shift from cars to bicycles. Ten percent of all car trips shorter than 7.5 km were assumed to be replaced by bicycle trips. Assuming constant risk (casualties per distance travelled), the expected number of accidents is proportional to the mobility shift. The results indicated that the total gain of the modal shift was negative for fatalities, which means that there was a net increase in the number of fatalities. For hospitalized casualties, due to the strong influence of the many hospitalized cyclists in nonmotorized vehicle crashes, there was a strong negative overall effect, and the modal shift resulted in a positive effect for 18- and 19-year-old males only. Overall, a small increase (up to 1%) in the number of cyclist fatalities and a greater increase of 3.5 percent in the number of inpatients was expected. The increase in casualties was mainly due to the proportion of single-vehicle bicycle crashes with serious injuries in relation to the total number of injured cyclists. The study provides a first approximation of the effect on road safety of a mobility shift from cars to bicycles. This approximation indicates that, in general, road safety does not benefit from this modal shift. • Christoforou et al. (2021) studied the usage of micromobility vehicles in the city of Paris (France), and especially electric scooters (ES) and free-floating electric scooters (FFES) that have been thriving there over the past couple of years. A survey showed that ES users rarely own their proper vehicle, are mostly men, aged 18–29, and have a high educational level. They are not less motorized than the general population and use ES occasionally. Their main motivation is travel time savings followed by playfulness and money savings. They shifted mainly from walking and public transportation (72%) and few have increased their total mobility by making new trips (6%).
Projection over time and space	• In the L3 Pilot, Deliverable D7.4 - Impact Evaluation Results, by Bjorvatn et al. 2021, in order to assess the future potential of automated driving system (ADS) to avoid accidents, three different approaches were used: ○ for parking-ADFs accident data was analyzed and checked if they correspond to ODDs of parking ADFs; from these the share of accidents was calculated which does not show any possible limitations, that would reduce the efficiency of the parking-ADFs. Finally, assumptions on possible market penetration rates were taken into account. ○ for motorway/urban ADFs two different scenario-based approaches were used ▪ Counterfactual simulation: original events were re-simulated as baseline and compared to other identical simulations where the ego-vehicle was firstly equipped with an AEB and secondly with a designed L3Pilot motorway ADF. ▪ Monte-Carlo traffic simulation: synthetic scenarios were sampled from distributions found in naturalistic driving studies and applied to ego-vehicles, agents and surrounding traffic using driver behavior models or pre-defined manoeuvres (note: urban scenarios did not consider surrounding traffic). For the scenario-based approaches, subsequently, injury risk functions (IRF) were used to estimate the probability of a specific injury severity level for pedestrians, cyclists, and belted car occupants (by type of collision). Both, changes in accident frequency and severity, were then used to calculate the impact of ADFs. To scale up the results on EU-wide level, the EriC-method was used. Thereby, the impact was calculated by injury severity level, assumed market penetration, and weighted by differences in traffic volumes and motorway types by vehicle kilometers travelled and by driving scenario occurrence in EU (these weighting factors were not applied for accidents in urban areas due to missing data availability).

6.2.6. Q&A

Not relevant for the draft.

6.3. Analyse Cost / Benefit

Cost-benefit analysis (CBA) is a method that considers the potential benefits and costs of the introduction of a safety measure to prioritize different options and evaluate a decision from the societal perspective. All costs and benefits are expressed in terms of money as much as possible. The V4SAFETY CBA takes the vehicle perspective, i.e., those safety measures are considered that influence the interaction of vehicles (car, van, truck) with other traffic participants. Consequently, besides in-vehicle safety measures, changes in the infrastructure and regulatory part are considered as well.

The cost part considers the development, manufacturing/ implementation, and service costs per unit (vehicle system, infrastructure measure) as well as the development of the number of units and the penetration rate.

The benefits part considers the crash risk, the injury probability in a crash, and the vehicle fleet per year and in time projection as well as the effectiveness of a system to calculate the number of prevented or mitigated crashes. Based on these numbers, the monetary benefit can be calculated considering the injury severity level.

In the last step of the analysis, the total costs and the total monetary benefits deliver the cost-benefit ratio and balance considering the discount rate.

Definitions for the Analyse Cost Benefit

- Cost: The monetary valuation of all resources1[4] needed to develop, produce, implement, and maintain and repair a safety measure or combination of safety measures.
- Benefit: The monetary valuation of the positive and negative2[5] societal impacts of the safety measure or combination of safety measures.
- Discount rate: an interest rate applied to benefits and costs that are expected to occur in the future to convert them into a present value.
- Cost-benefit analysis is a systematic method for calculating and comparing the benefits and costs of different safety measures.

Note: The “Analyse cost / benefit” will be updated by adding a detailed description based on the progress in the corresponding WP and the overlap with the topic of project wil be solved. The foreseen measure is that projection will be covered only in the topic “Project the Results”.

6.3.1. Visualization

The visualization of the cost-benefit analysis (CBA) is presented below in separate figures for in-vehicle safety systems, infrastructure measures and combined in-vehicle and infrastructure measures. This is because the CBA of each type of treatment includes different elements, although the basic structure of the CBA and most elements are the same for each type of treatment.

[4] Resources include all efforts, materials, human resources.
[5] Negative impacts are considered as negative benefits.

Figure 4.9 depicts the structure of the CBA methodology for in-vehicle safety systems. The right-hand side shows the calculation of the benefits. Firstly, the benefits in one country or area are calculated, based on the number of target casualties in the base year, projection to future years and the effectiveness of the safety system. Next, the benefits are projected to other countries. The same procedure is applied to the calculation of the costs, which is shown at the left-hand side. Both the costs and the benefits are calculated on a per-vehicle basis firstly and in the next step the total costs and benefits are calculated using vehicle fleet data.

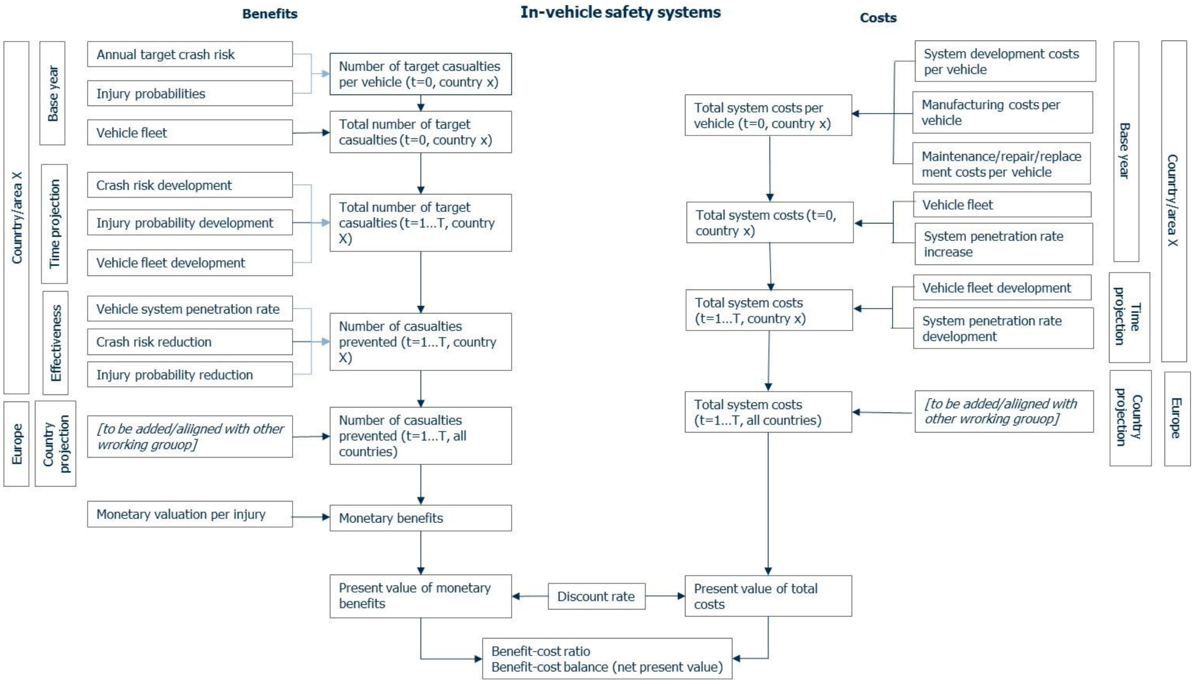


Figure 6.4: Structure of the CBA of in-vehicle safety systems.

Figure 6.5 visualizes the CBA of infrastructure measures. It is mostly identical to the picture for in-vehicle safety systems, but different variables are used on several occasions (depicted in bold).

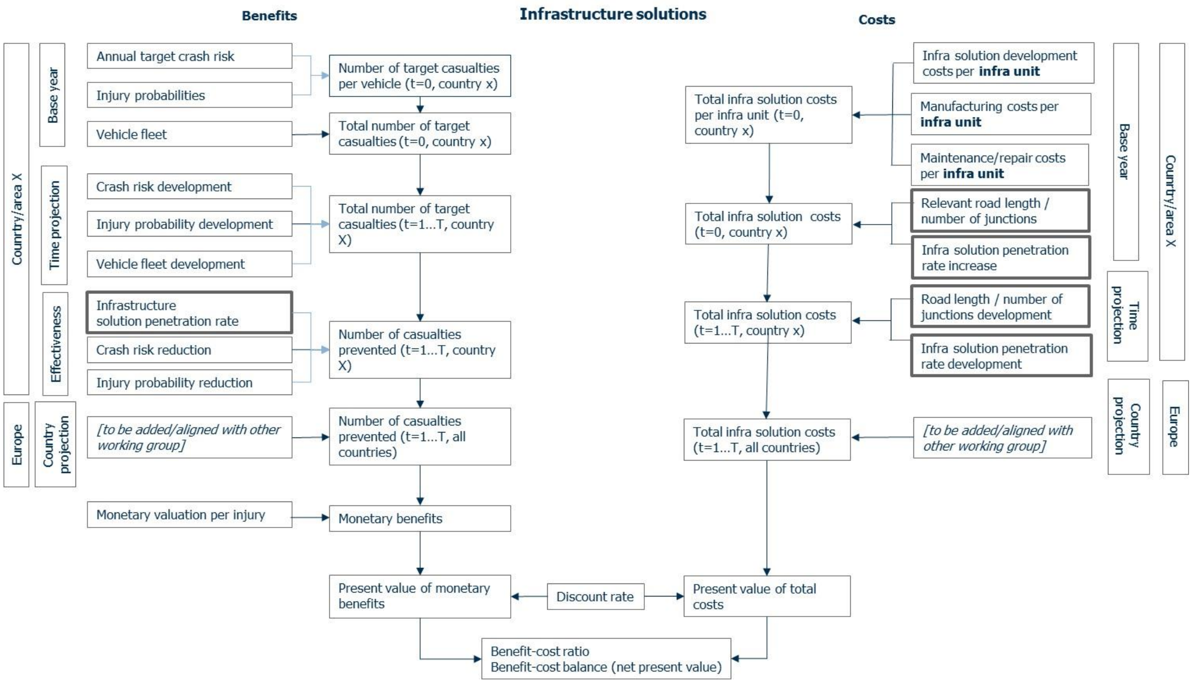


Figure 6.5: Structure of the CBA of infrastructure measures.

The structure of the CBA of combined in-vehicle and infrastructure measures is presented by Figure 6.6 (benefit side) and Figure 6.7 (cost side). The main difference on the benefit side is the fact that the number of casualties prevented is determined by the number of vehicles affected by the combined measure, which in turn depends on penetration rates of both the vehicle system and the infrastructure measure (see bold blocks). The cost side is a combination of the cost side of in-vehicle safety systems and infrastructure measures since the costs of combined measures are the sum of the costs of both measures.

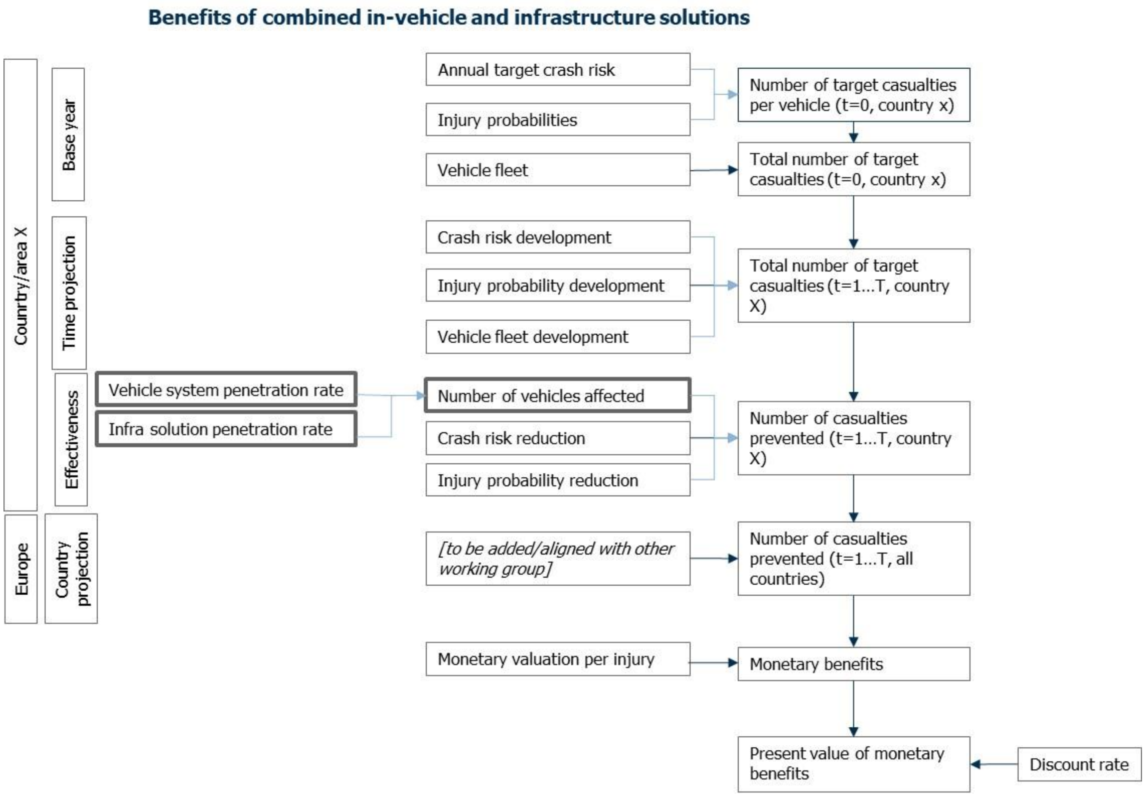


Figure 6.6: Benefit side of the CBA of combined in-vehicle and infrastructure measures.

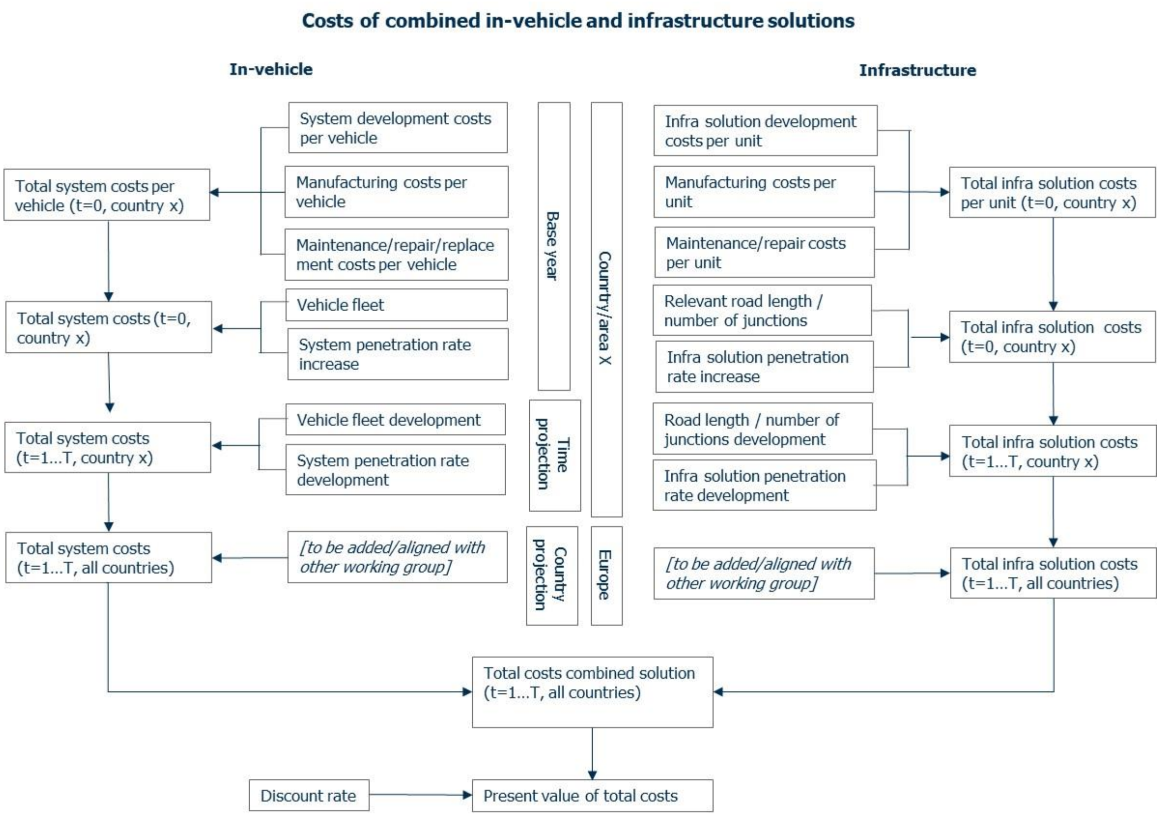


Figure 6.7: Cost side of the CBA of combined in-vehicle and infrastructure measures.

6.3.2. Detailed Description

A detailed description for this topic will be added aq later stage once work in the related work packages has carried out.

6.3.3. Input Output

Not relevant for the draft.

6.3.4. Consequences

Not relevant for the draft.

6.3.5. Examples

Will be added at a later stage of the project once the V4SAFETY examples are ready.

6.3.6. Q&A

Not relevant for the draft.

7. Cross Assessment Topics

The cross-assessment topics address aspects that are relevant to (nearly) all assessment related topics. Thus, the input to these topics should come from all topics of the chapters four, five and six. The two cross assessment topics are the conduction of validation and verification and the documentation. Obviously, the output of the validation and verification is also an input for the documentation.

7.1. Conduct Validation & Verification

The results of a prospective safety assessment can serve as a foundational basis for decisions that may lead to substantial societal or economic consequences. For example, when a legislation obliges manufacturers to equip their vehicles with a specific safety system, or when a manufacturer elects to formulate their system development strategy based on the simulation results.

Concurrently, performing a safety assessment simulation is a rather complex task, since it comprises the interaction of data processing, simulation models and tools. This complexity thereby offers extensive potential for inaccuracies, which may consequently yield biased or non-credible results that sufficiently accurate represent the real-world.

To address this concern, a validation & verification (V&V) process is established within the V4SAFETY framework. It aims to counteract potential inaccuracies by continuously accompanying the assessment process and providing measures that attest to the credibility, accuracy, and precision of the simulation results. Additionally, it shall ensure that all phases of the simulation process are tailored to align with the specific safety assessment objective and thus effectively address the underlying evaluation scope.

7.1.1. Visualization

Figure 7.1 shows how the V&V activities are arranged with regards to the assessment projects. It should highlight that V&V activities continuously accompany the simulation.

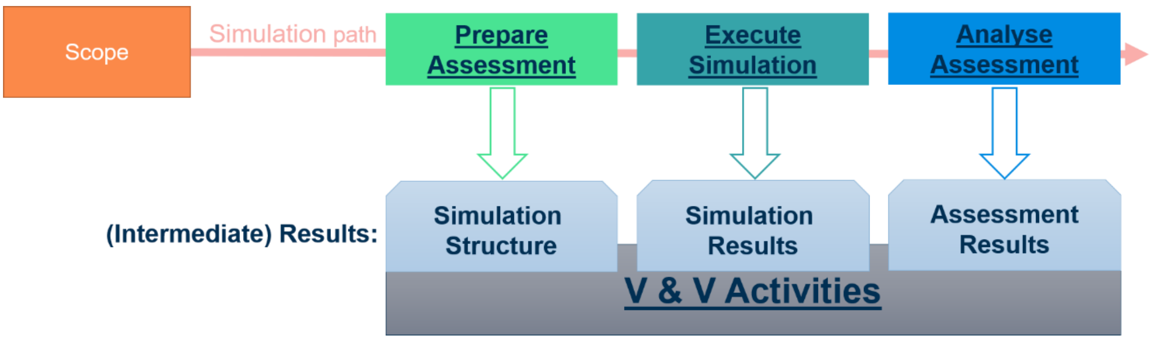


Figure 7.1: Integration of the V&V activities into the V4SAFETY process.

7.1.2. Detailed Description

Within the V4SAFETY framework the topic of conducting validation and verification activities has a special role. It covers the entire assessment process since it has to consider the outputs of the assessment preparation, the simulation execution and the assessment analysis. These outputs are the basis for evaluating the assessments credibility by applying different V&V approaches.

Therefore, V&V procedures in the assessment process are pivotal for ensuring the assessments credibility, particularly when stakeholders depend on them for informed decision-making. The aim should be to provide a systematic basis for this decision of acceptance and to document it in a readable and transparent way. By doing so, V&V not only bolsters the quality of the simulation but also mitigates risks from incorrect predictions and elicits confidence among stakeholders. Thereby, a key aspect of V&V should basically be the practical applicability of the methods used. A V&V procedure that is too extensive and ties up too much capacity is just as ineffective as a desultory executed V&V process step at the end of the assessment.

The term “validation” is often subject to different understandings when used in quantitative and qualitative research (Dellinger & Leech, 2007). The same holds true for related terms. To have a uniform understanding of the underlying terms for the following explanation of the V&V process, the ones that are considered as most important are defined in Table 7-1.

Table 7-1: Definitions used for the V4SAFETY V&V process.

Term	Definition
Safety assessment validation	Validation of a prospective safety assessment is an evaluation of simulation results, demonstrating them to be credible, accurate, precise, and adequate for purpose with regard to the evaluation question.
Verification	Confirmation through the provision of objective evidence that (internal) requirements of the safety performance assessment process have been fulfilled.
Credibility	Credibility in simulation results is the quality to elicit belief or trust.
Accuracy	Accuracy is the proximity of simulation results to a reference value.
Precision	The consistency of the virtual simulation in reproducing the same safety assessment results under the same conditions.
Adequacy for purpose	The simulation structure is aligned to the specific safety assessment scope and objectives and can effectively address the underlying evaluation questions.

V&V is an ongoing activity that spans the entire assessment procedure, rather than being a specific phase or step within it (Balci, 1994) By embedding V&V throughout the entire process, potential inaccuracies or flaws in intermediate results can be identified and corrected early on. Therefore, it is required to interlock the V&V process with the procedure of the overall V4SAFETY assessment Framework.

The procedure of the V4SAFETY assessment Framework is depicted on a very high level in Figure 7.1. Starting from the scope, the three top-level phases prepare assessment, execute simulation and analyse assessment are run through. Each of these phases produce an intermediate result, which serves as input for the next phase. The intermediate results are the stages at which the dedicated V&V process interlocks with the overall assessment process, because each intermediate result can be validated or verified using various V&V techniques. The assessment result represents the final results, which ideally should also be validated. However, in case of a prospective assessment this is due to the nature of the prospective assessment and due to missing data typically not possible. Therefore, emphasis should be put on the V&V of the intermediate results with specific focus on the baseline.

It has to be noted that analogous to the top-level process steps, only the abstracted intermediate results are shown in Figure 7.1. If, for example, the intermediate result "simulation structure", which emerges from the prepare assessment, is broken down, this also includes the evaluation question, the generated baseline cases and selected or created models.

Figure 7.2 illustrates the V&V process. It is a condensed and abstracted version of the V&V procedure model that Rabe, Spieckermann, and Wenzel (2008) have described. The figure is divided into two parts: on the left side the procedure is depicted that is run from the bottom left along the simulation path to the top right. On the right side a list of V&V approaches is shown that encompasses a range of potential V&V approaches.

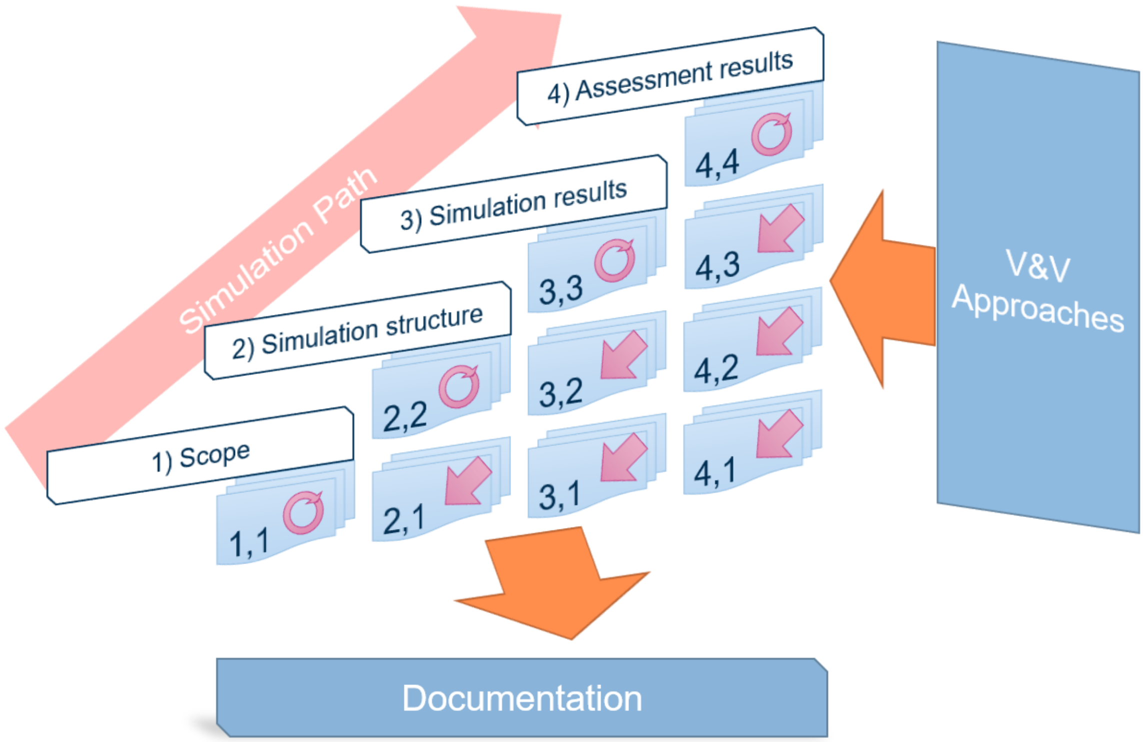


Figure 7.2: V&V process along the simulation path.

Following the V&V process, each column represents an intermediate result that originates from the entire process. Every line is made up of V&V elements, represented by the blocks with a number that is denoted by two indices. The first index identifies the phase result that is validated by the activities of this V&V element. The second index identifies the phase result that is used as the reference for the V&V of this V&V element. As an example of the intermediate result "simulation results" – the third column of the process graphic – it would be as follows: Starting from the top, the first V&V element is numbered 3,3. This means that an intrinsic V&V activity is carried out, i.e. the intermediate result is evaluated against itself. This is also indicated by the circle in the right-hand corner of the V&V element. Within each element of the V&V process different methods can be applied. A concrete example of such an intrinsic evaluation could be the application of the V&V procedure of a sensitivity analysis. A sensitivity analysis is a systematic method to evaluate the impact of uncertainties in the input parameters of a model or simulation on its output. If small changes in input parameters lead to significant changes in outputs, the used models or results might not be representative of the real-world.

Going down the column according to the Figure 7.2, the next V&V element would be number 3,2. This means that the simulation results are now validated with respect to the simulation structure. A possible V&V approach for this element could be the comparison of results. Thereby, the results of the simulation are compared with another simulation study that has a similar simulation structure and is already validated. It would also be possible to compare the results of the simulation with different methods like test track tests or the retrospective analysis of accident data. A reference for this method could be found in Fahrenkrog (2016). Note that great care has to be taken to selection biases in validation datasets.

Lastly, the simulation results are evaluated against the scope of the assessment. This is reflected in the V&V-element 3,1. A possible approach could be the face validation as described in (Balci, 1998). In the V4SAFETY context face validation involves a collective evaluation carried out by project team members, potential users and individuals possessing substantial knowledge about the system being examined. Through their collective insight and intuitive judgments, they subjectively analyse the simulation results and check on whether they reasonably mirror the anticipated system behaviour for the underlying scope. With regards to the definition provided in Table 7-1, this can be rather considered as a plausibility check.

A mapping of recommended V&V procedures to intermediate results is presented in Table 7-2.

Table 7-2: Mapping of V&V approaches to process results (initial version that will be updated during the project).

V&V approach	Validation or verification	Scope	Simulation structure	Simulation results	Assessment results
Comparison of results	Verification & Validation		(X)	X	(X)
Review of the whole method or the simulation results	Verification		X		
Check by means of test data	Verification		X	X	X
Back-to-back tests	Validation & Verification		X		
Model fit of statistical models	Verification		X	X	
Review of the whole method or simulation models	Validation	X	X		
Test validity	Validation	X			

Propagation of uncertainties	Validation		X	X	X
Statistical testing of differences in model vs input data	Verification		X		
Sensitivity analysis	Validation		X	X	X
Face Validation	Validation	X	X	X	X
Animation	Validation		X	X	X
Historical Data Validation	Validation		X	X	X
Turing Test	Validation		X	X	X
…					

When all V&V activities of the process have been carried out, the findings are documented. The proper documentation of the V&V activities is an inevitable step. It provides a record of the V&V process, which can be used to review the validation activities at a later time and to assess the credibility of the simulation study. The V&V documentation should include detailed information about the V&V activities that were performed, the results of those activities, and any findings or recommendations. It should be clear, concise, and well-organized so that readers can easily understand the V&V process and its results. A short report should be created for each V&V-element, that will be part of the longer document that encompasses the whole V&V process.

7.1.3. Input Output

As outlined in the previous detailed description, the V&V process is strictly interlinked with each phase of the entire simulation procedure – accompanying it from the beginning to the very end. Due to the parallel execution of the two processes, the type of inputs and outputs are differentiated between process related inputs (non iterative) and simulation related inputs (iterative). In Figure 7.3 the first ones are depicted as arrows pointing towards the V&V process, while the latter one is integrated into the V&V “circle”.

Simulation related inputs are the results of the simulation phase under investigation. They are fed into the V&V process to determine whether they meet the requirements that have been set for validation and verification of this specific simulation phase. In case the requirements are not met, the results of previous phases have to be reconsidered and preceding V&V actions have to be repeated. Therefore, the requirements that affect the phase of the simulation procedure that has generated insufficient results, may need to be adjusted within the V&V process and fed back to the simulation procedure.

The requirements for the V&V process itself are one of the three process related inputs. They are derived from the evaluation scope, the data that is used and processed in the V&V activities and the assessment criteria.

The second process related input are the limitations and specifications that have to be considered when performing any V&V actions. Elements of the simulation are often constrained by data, method, and model limitations. Data limitations could emanate from inconsistencies, incompleteness, or irrelevance of the used datasets. Methodological constraints may arise due to intrinsic restrictions of specific V&V or simulation approaches, potentially leading to suboptimal applicability. Models, as abstractions of reality, are inherently founded on assumptions which may not universally apply, thereby risking omission or misrepresentation of real-world elements. Further, simulation tool specifications determine the scope, fidelity, and capabilities of a simulation environment or an algorithm.

The last process related input is the reference data that is used to perform certain V&V techniques – e.g., accident data that can be used to compare synthetically generated crashes with the real world.

The output of the V&V process is twofold: The "hard" output is proper documentation. The documentation should be a structured and traceable record of all aspects related to V&V, including methodologies and data employed, tests executed, and results obtained. This not only ensures transparency but also fosters repeatability, allowing other experts or teams to understand, replicate, or build upon the V&V process. A complete and transparent documentation of an immaculate performed V&V process should evoke the second "soft" output: The trust of the stakeholder in the assessment quality. The stakeholders need to be confident that the assessment results are valid, and that the simulation demonstrated to be credible, accurate and precise.

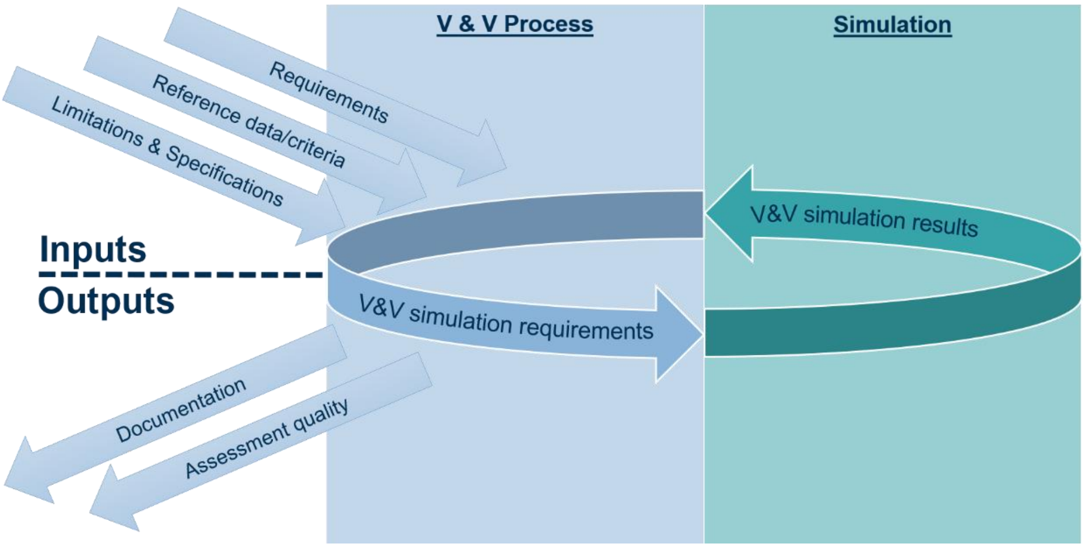


Figure 7.3: Input and outputs for the V4SAFETY V&V process.

7.1.4. Consequences

Not relevant for the draft.

7.1.5. Examples

Table 7-3 presents a selection of V&V approaches. The first column lists the name of the approach, the second column offers a brief explanation, and the third column cites a reference where this approach has been applied.

Table 7-3: Examples for the application of V&V approaches.

V&V Approach	Explanation	Reference

Sensitivity Analysis	It is a method for validating the results of a simulation or the outputs of a model, as it can help to identify the key parameters that drive the simulation and to assess the robustness of the results to changes in these parameters. In a sensitivity analysis for an AEB system, various input parameters such as sensor accuracy, reaction time, and braking force might be examined. For instance, a variation in the sensor angle might significantly impact the system's ability to timely detect and respond to obstacles, thus affecting the efficacy of the AEB system in preventing collisions. By altering these input parameters and observing the changes in the output (e.g., collision avoidance or mitigation), the sensitivity analysis helps in understanding how robust the simulation is to uncertainties in these key parameters, and which parameters are most influential in the system's performance.	Chajmowicz, Saadé, and Cuny (2019)
Check by means of test data	The V&V approach relies on comparing simulation outcomes or model outputs against a benchmark. This benchmark is derived from experimentally gathered data, such as those from test track experiments, drone recordings, or FOT data. For instance, XX contrasts the time headway distribution from simulations using the BMW cognitive driver model against data from euroFOT and HighD.	Fries & Fahrenkrog (2021)
Check baseline generation against crash database outcome distributions	This type of validation is aimed at understanding if the baseline generation produces crashes with the same (or similar enough) characteristics as those in real life. Here real-life is often represented by crash databases. That is, the validation is then a comparison between retrospective data (historic crashes) and the simulated cases. The comparison can be made on several crash characteristics (e.g., outcome severity metrics such as relative speed at impact, deltaV, injury risk, but also impact angles etc), but a starting point is often some severity metric. In this type of validation, it is crucial to consider the bias in the outcome severities in the crash database to which the simulations are compared. All crash databases have substantial selection bias across severities, and most do not include any property damage only (PDO) crashes, even if the latter can be 50-90% of all crashes (depending on scenario). As several of the baseline approaches create crashes across all severities (including PDOs), the crash database bias must be considered, or the comparison (validation) is between "apples and oranges".	Bärgman et al. (submitted)

Note: table will be continued for final for the deliverable.

7.1.6. Q&A

Not relevant for the draft.

7.2. Document Assessment

There are two main aims of documenting safety benefit assessment. One is documenting the benefit assessment results for stakeholders for decision-making. The second one is the documentation of the method – how the results were produced. Particularly important are details on assumptions and decisions that will or may impact the benefit assessment. The latter is important to provide stakeholders the right context for usage of study.

In the V4SAFETY framework we propose a tiered approach to documentation. That is, we recommend documenting a particular assessment at three different levels. Practically this may be three documents of 3, 30, and 300-pages, respectively. Note, however, that the neither the number of pages, not the inclusion of all are mandatory – they only serve as guidelines.

The 3 pages document would be an executive summary that can be used to understand the basics of the work (including scope, general on method, main results, and main assumptions) and main results. The 30-pager would typically be the most important of the three levels. A reader should get enough details about the method and results to determine the relevance and validity of the work. It would however typically not be possible to replicate a study based only on the 30-pager. The 300-page level document should contain all details of the assessment. The 300-page document should (aim) to contain enough information to enable replication of the assessment.

7.2.1. Visualization

Figure 7.4 and 7.5 provides an overview of content across the three levels of documentation, from two different perspectives. The two representations have different uses. The Proportion of Document plot (Figure 7.4) indicate to writers of documentation how much to include on each topic when actually perform the documentation. The Proportion of All Material plot indicate to readers where they would find what information, and what “completeness” to expect on the different topics.

Note that the number of categories making up the assessment has been substantially reduced compared to the overall framework (see Figure 2.1; the main framework figure). Also note that “Assumptions and limitation” have been added as a separate entity. Those are aspects of the documentation that are in all of the original components, but that are particularly important to highlight in documentation, as it more or less is the only means stakeholders have to evaluate the relevance and validity of a specific study, in relation to the scope of the study. It is thus crucial information for stakeholders in their decision making.

Both Figure 7.4 and 7.5 are spider-plots with the three levels of documentation included as three different “treads”, each with the enclosed area of different colours. The axes, however, have different meanings in the two plots. Figure 7.4 shows, for each topic, the proportion of the overall material in the specific document (i.e., the number of pages/amounts of text; in percentages of the total length of the document) across the document types. That is, a topic with a low value (close to the centre) for one of the documents, should cover much fewer pages than a topic with a high value (towards the outer edge of the spider web). For example, only a very small part of the 3-pager should be about the details of how the simulations where run, while it should contain much about the “Assessment Scope” and “Limitations and assumptions”. This can be seen as the “Proportion of document” plot.
For Figure 7.5, the axes indicate how much of the overall material (potentially) available on a topic that should be included in each of the three types of documents. For example, if the value on one axis is low (close to the centre), only a small fraction of the overall material (potentially; i.e., if all possible material on the topic was available) available for this topic (for the specific assessment) should be included in that document type. Take the example of methods, of which a very small proportion should be included both in the 3-pager and the 30-pager, but all should be included in the 300-pager. As you can see, the 300-pager almost include all the material.

Note that the two figures are complementary only to provide some guidance – to give users of this work “a feeling for” what to put the emphasis on across the three levels of documentation.

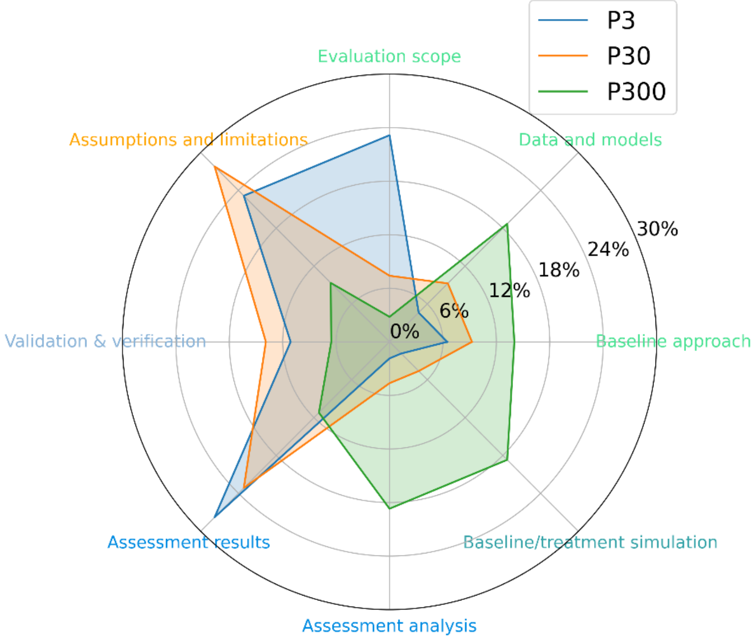


Figure 7.4: Proportion of Document: Documentation focus across the three levels of documentation. This panel shows the relative amount of material/pages for each of the document types. For the relationship between the categories in the plots and the framework categories, see Figure 7.1. The label colours correspond to the Framework colours.

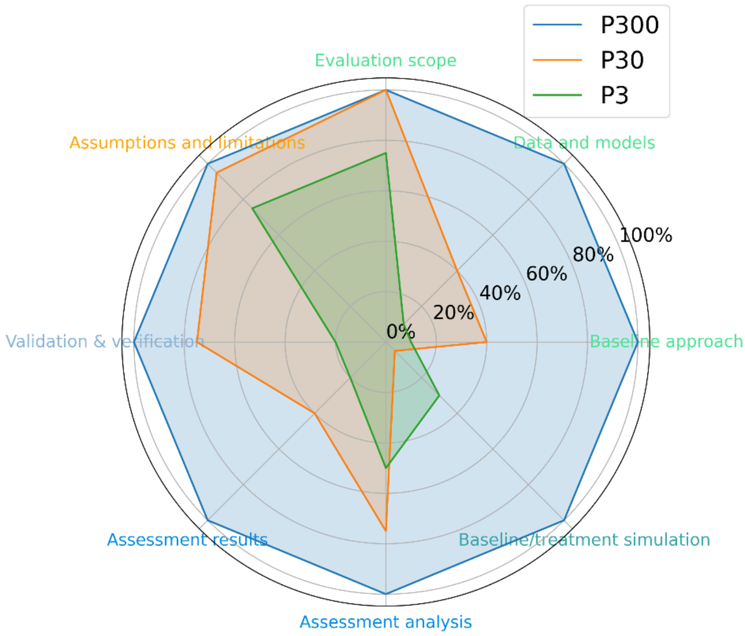


Figure 7.5: Proportion of all potentially available material: Documentation focus across the three levels of documentation. This panel shows how much of the total amount of material (potentially) available for each topic. For the relationship between the categories in the plots and the framework categories, see Figure 7.1. The label colours correspond to the Framework colours.

7.2.2. Detailed Description

The tiered assessment-documentation approach of V4SAFETY has the aim to provide different stakeholders with different levels of detail of information, as well as information with different focus.

The approach is called “3/30/300” and differentiates three levels of documentation level.

Important note: *The fact what we use 3/30/300 does not mean that it always should be exactly 3, 30 and 300 pages (it may be 2, 37 and 93), nor that there should be a factor ten between the number of pages, or that all three documents always should be produced (i.e., the “300” document may be skipped if for example company internal or documentation is available from previous projects).*

The setup of the documentation is naturally depending on many different factors, such as the overall scope (how large the study is), already existing documentation (i.e., replication of documentation of details on methods or models are not needed if references can be included - typically general overviews of methods and models should still be provided), and the targeted stakeholders/users of the assessment results (i.e., if you as the person performing the analysis are and will be the only user of the results, the documentation needs are different than if it is an assessment that may have major impacts on policy making or legislation).

The following describe each of the three levels of documentation, including a description of which items to typically include, as well as the rationale for why to include them. These are only high-level guidelines. For some items there are separate templates for documentation (e.g., how to document a vehicle or driver model). This is followed by a section on do’s and don’t’s in documentation, and what documentation is expected for each of the individual framework components. For each component the information that is recommended to be included in the 3/30/300-pages, respectively, are outlined. Examples of detailed documentation can be found in Deliverable D6.1.

The 3-pager (executive summary) provide readers with the information necessary to understand the basics of the assessment. It focuses on assessment scope, results, and provide a short general section on method – the latter with an emphasis on aspects of the assessment that affect validity and generalization of results (e.g., about assumptions and limitations that may impact validity/decision-making). Intended readers of the 3-pager include people who wants to get a quick overview of a specific assessment. Note that this document does not contain all the details concerning the validation & verification of the results as well as the limitations and assumptions of the study. Therefore, it shall be treated with care and when used for decision-making, cross check the 30-page document as well to obtain a sound overview of the study.

The 30-pager document is the intermediate level document, which provide a decent overview of the entire assessment. Intended readers of the 30-pager include anyone that is to use the results for decision making, which is why it is important that it contains enough information for the user to judge the validity and relevance of the assessment. It also means that typically many of the method details need not to be included. Instead, the focus should, in addition to scope and results, be on aspects of method that are likely to have the (main) impacts on the outcomes. If a particular assessment compares the results of the current study with previous/others’ work, information about the differences and potential biases of both studies should be compared in the 30-pager. Note that it typically would not be possible to replicate a study based only on the 30-page document – that is what the 300-pager is for. The structure and level of detail of the 30-pager may be in the direction of a scientific paper, however with more focus on assumptions and limitations that may affect decision-making based on the results.

The 300-pager level document should contain all details of the assessment and the construction of the entire assessment. It should (aim) to contain enough information to enable replication of the assessment, and understand all assumptions and decisions made. Intended readers are people who a) intends to replicate the study or perform a new assessment using the same method/framework, b) wants to understand the details of assumptions, decisions and limitations, and c) plan to further develop part of the assessment method (e.g., models).

Note that, a) the 3 and 30-pagers are what most stakeholders would have access to, while the 300-pager is often organisation-internal, b) the 3 pager can be the "executive summary" of the 30-pager, or even the 300-pager, c) typically that there may be documentation templates available for different parts (e.g., some model types).

Table 1 provides an overview of what information is recommended to be included. The documentation is divided into background/scope, method, results and discussion/conclusions, but for each of these topics documentation guidelines for each relevant framework components are described (for each of the 3/30/300 document levels).

Table 7-4: An overview of what information is recommended to be included (Note that TBD (To Be Defined) will be filled later after the draft)

Documentation structure	Framework component	3-pager	30-pager	300-pager
General:		Provide readers with the information necessary to understand the basics of the assessment. It would focus on assessment scope, results, and provide a short general section on method – the latter with an emphasis on aspects of the assessment that affect validity and generalization of results.	TBD	TBD
Background and scope				
	Define evaluation scope	This is a very important part in the 3-pager. A reader should be able to understand the entire scope.	The 30-pager should describe the scope in a more detail than the 3-pager.	The evaluation scope is already covered by the 3 and 30 pagers, not much extra to add here.
Method				
	Select baseline approach	The baseline approach should be described at a	TBD	TBD

		very high level. Decisions/ assumptions/limitations that (may) affect generalizability and validity should be mentioned, and references to the 30-pager should be made for details.		
	Prepare data	Provide a short overview of the data used for baseline generation. Add additional information if the assessment scope is an extrapolation of the data used. That is, that the data do not directly fit the assessment scope, and thus the assumptions/extrapolations may impact the results (e.g., with respect to generalizability and validity).	TBD	TBD
	Select models	Details on models should typically not be included, but if choices of models were made where one (or more) models are substantial extrapolations of what they were intended for, a short description should be made (see also Discussion)	TBD	TBD
	Configure simulation	No information needed.	TBD	TBD
	Manage simulation	No information needed.	TBD	TBD
	Simulate baseline	No information needed.	TBD	TBD
	Simulate treatment	No information needed.	TBD	TBD
	Evaluate safety performance	A high-level description on how safety performance was calculated can be included.	TBD	TBD
	Analyze cost/benefit	A high-level description on how cost/benefit was performed can be included.	TBD	TBD

	Project the results	A high-level description on how results were projected can be included.	TBD	TBD
	Conduct validation & verification	TBD	TBD	TBD
Results				
	Evaluate Safety Performance	The main safety performance results should be presented here.	All results relevant for decision-making should be included at this level.	All results can/should be included here, but with the focus on the details (e.g., in-depth plots an tables).
	Cost/benefit	Cost/benefit results should be reported at a high level.	TBD	TBD
	Project the results	Projected results should be reported at a high level.	TBD	TBD
	Conduct validation & verification	Any validation relevant to the assessment scope that has been done on, for example, baseline generation, should be mentioned. Descriptions on verifications (incl. result of such) are typically not needed in the 3-pager.	TBD	TBD
Discussion & conclusion				
	Select baseline approach	TBD	TBD	TBD
	Prepare data	TBD	TBD	TBD
	Select models	Only include a discussion about models used if the model choice/design/underlying data may substantially impact the results (e.g., generalizability and validity), and then only information relevant to that.	TBD	TBD
	Configure simulation			

	Manage simulation	Typically, no information is needed about the management of the simulation. However, if there are decisions made that directly may impact generalization and validity of the results, those and the potential impact on results should be described. This may include information of sampling methods of simulation stopping criteria.	TBD	TBD
	Simulate baseline	Typically, no information is needed about the simulation execution. However, if there are decisions made that directly may impact generalization and validity of the results, those and the potential impact on results should be described.	TBD	TBD
	Simulate treatment	Typically, a short discussion on the safety measure should be included, with focus on simplifications (e.g., ideal sensor models) that may impact the results.	TBD	TBD
	Evaluate safety performance	Safety performance analysis aspects/assumptions/ decisions that may affect generalizability and validity should be mentioned and briefly described (referring to the 30-pager for more details).	Safety performance analysis aspects/assumptions/ decisions that may affect generalizability and validity should be described in quite some detail here. All that decision makers need to make informed decisions should be included here.	This can focus on details, such as detailed sensitivity analysis etc. (see also V&V)
	Analyze cost/benefit	Methodological decisions/assumptions/ limitations that (may) affect results should be mentioned, referring to the 30-pager for more details.	TBD	TBD

	Project the results	Methodological decisions/assumptions/ limitations that (may) affect results should be mentioned, referring to the 30-pager for more details. Examples include briefly describing aspects related to weighting that are used for extrapolation/projection.	TBD	TBD
	Conduct validation & verification	TBD	TBD	TBD

Details about the 3/30/300 pager will be added later.

7.2.3. Input Output

The input to documentation is all information from all topics needed to properly document a particular assessment, while the output include the 3/30/300 documents, as well as complementary documentation (e.g., specific templates for some components, such as for measure, vehicle and behavior models).

7.2.4. Consequences

Not relevant for the draft.

7.2.5. Examples

The documentation examples are part of Deliverable D6.1.

7.2.6. Q&A

Not relevant for the draft.

8. Summary

The draft document presents the initial version of the V4SAFETY framework for the virtual evaluation of road safety. The draft is intended to forester discussion in the consortium as well as with external stakeholders. The work on the framework will continue for the entire duration of V4SAFETY. The final deliverable will present the final framework.

After describing the process and background of the V4SAFETY framework, the framework is presented. The process of the framework definition started with the definition of requirements that should be fulfilled. This has been the basis to develop the framework which consist of four main topics that are divided in different topics and two cross-topics (conduct V&V and documentation), which are relevant for all process related topics. The four main topics which have been structure along the logical process (not necessarily in chronological order) are:

- V4SAFETY Framework (topics of this main topic: Definition, User & Stakeholder, Examples and Formulate Conclusions) – addresses the general aspects of the framework and is the only not directly process related main topic.
- Prepare Assessment (topics: Define Evaluation Scope, Select Baseline Approach, Prepare Data, Select Models)
- Execute Simulation (topics: Configure Simulation, Manage Simulation, Simulate Baseline, Simulate Treatment) describes the actual simulation. Configure and manage simulation describe the general process, while simulate baseline and treatment address the specific aspect of these simulations.
- Analyse Assessment (topics: Evaluate Safety Performance, Analyse, Cost / Benefit, Project the Results) covers the handling of the simulation output including following assessment steps, such as project the data to a different region or time and cost-benefit analysis).

All topics are described along a common structure. First, a high-level introduction of the topic is given. Then a visualization of the topic's process, a detailed description of the topic and the input to as well as the outputs to the topic is provided. In addition, to this core information in the draft a few examples are provide – as far as they are already available. For the final deliverable additional examples, a Q&A section as well as the discussion of potential consequences of certain decisions in the process will be added.

The draft report will be continued to be developed during the V4SAFETY project. In this sense, it presents an intermediate result, which will be carefully reviewed, discussed, and constantly updated based on the feedback of the other V4SAFETY WP and the input of the advisory partners. A few aspects have already been identified for being updated or added to the framework. These are:

- role of in-crash simulation,
- consideration of injury risk functions,
- theoretic background for the sampling the Baseline C approach,
- etc.

References

ASAM e.V. (2024), OpenX standard, OpenScenario XML, https://www.asam.net/standards/detail/openscenario-xml/

Balci, O. (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research*, *53*(1), 121–173. https://doi.org/10.1007/BF02136828.

Balci, O. (1998). Verification, validation, and testing. *Handbook of Simulation*, *10*(8), 335–393.

Bärgman, J., Svärd, M., Lundell, S., & Hartelius, E. (Submitted). Methodological challenges of scenario generation validation: a rear-end crash-causation model for virtual safety assessment (in publcaiton process).

Bärgman, J. Victor, T. (2020) Holistic assessment of driver assistance systems: How can systems be assessed with respect to how they impact glance behaviour and collision avoidance? IET Intelligent Transport Systems. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85091932127&doi=10.1049%2fiet-its.2018.5550&partnerID=40&md5=ce357501cba57978c60eb147552475c3.

Bjorvatn, A., Page, Y., Fahrenkrog, F., Weber, H., Aittoniemi, E., Heum, P., Lehtonen, E., Silla, A., & Bärgman, J. (2021). L3Pilot D7.4-Impact Evaluation Results. L3Pilot Deliverable, https://l3pilot.eu/fileadmin/user_upload/Downloads/Deliverables/Update_14102021/L3Pilot-SP7-D7.4-Impact_Evaluation_Results-v1.0-for_website.pdf.

Chanove, A., Pohle, M., Urban, M., Lorente Mallada, J. (2022) *An extrapolation method on European accident data based on weighting and data harmonization*, 8th Road Safety and Digitalization Conference, Athens, Greece, https://www.nrso.ntua.gr/rss2022/wp-content/uploads/2022/06/RSS2022_paper_217.pdf.

Chajmowicz, H., Saadé, J., & Cuny, S. (2019). Prospective assessment of the effectiveness of autonomous emergency braking in car-to-cyclist accidents in France. *Traffic Injury Prevention*, *20*(sup2), S20-S25. https://doi.org/10.1080/15389588.2019.1679797.

Christoforou Z., Gioldasis, C., de Bortoli, A., Seidowsky, R. (2021). *Who is using e-scooters and how? Evidence from Paris*. Transportation Research Part D: Transport and Environment, Volume 92, 2021, 102708, ISSN 1361-9209, https://doi.org/10.1016/j.trd.2021.102708.

Dellinger, A. B., & Leech, N. L. (2007). Toward a Unified Validation Framework in Mixed Methods Research. *Journal of Mixed Methods Research*, *1*(4), 309–332. https://doi.org/10.1177/1558689807306147

Denk, F., Brunner, P., Huber, W., Margreiter, M., Bogenberger, K., & Kates, R. (2022). *Assessment of traffic safety interventions using virtual randomized controlled trials: potential of connected and automated driving including V2X for collision reduction at urban intersections*. IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), 1183–1190. https://doi.org/10.1109/ITSC55140.2022.9921764.

Fahrenkrog, F. (2016) *Wirksamkeitsanalyse von Fahrerassistenzsystemen in Bezug auf die Verkehrssicherheit* (Dissertation). Rheinisch-Westfälische Technische Hochschule Aachen; fka Forschungsgesellschaft Kraftfahrwesen.

Fries, A., & Fahrenkrog, F. (2021). Validation and Verification of the Stochastic Cognitive Driver Model. *ACIMobility Summit,* Braunschweig, Germany.

Fries, A., Lemberg, L., Fahrenkrog, F., Mai, M., Das, A. (2023). Modeling driver behavior in critical traffic scenarios for the safety assessment of automated driving, 27th Enhanced Safety of Vehicles (ESV) conference, Yokohama, Japan.

Glasmacher, C., Weber, H., Schuldes, M., Wagener, N., Eckstein, L. (2023). Generation of Concrete Parameters from Logical Driving Scenarios Based on Hybrid Graphs. VEHTIS, Prague, pp. 215-222, 2023.

Kalayci, T., Kalayci, E., Lechner, G., Neuhuber, N. Spitzer, M. Westermeier, E., Stocker, A. (2020) Triangulated Investigation of Trust in Automated Driving: Challenges and Solution Approaches for Data Intregration., Journal of Industrial Information Integration, 2020.

Hedlund, J. (2015). *SHRP 2 naturalistic driving study: The first data analysis studies*. Journal of Safety Research, 54, 3–4.

Helmer, T. (2014). *Development of a methodology for the evaluation of active safety using the example of preventive pedestrian protection.* Disseration TU Berlin, https://depositonce.tu-berlin.de/handle/11303/4270.

Hickey, G. L., Grant, S. W., Dunning, J., & Siepe, M. (2018). Statistical primer: sample size and power calculations—why, when and how?. *European journal of cardio-thoracic surgery*, *54*(1), 4-9.

ISO 21934 (n.d.) ISO/AWI TS 21934-2: Road vehicles — Prospective safety performance assessment of pre-crash technology by virtual simulation — Part 2: Guidelines for application (under development). https://www.iso.org/standard/81790.html.

Isaksson-Hellman, I., & Norin, H. (2005). *How thirty years of focused safety development has influenced injury outcome in volvo cars*. Annual Proceedings. Association for the Advancement of Automotive Medicine, 49, 63–77. https://pubmed.ncbi.nlm.nih.gov/16179140/.

Lesire, P., Léopold, F., Chauvel, C., Hervé, V., & Cuny, S. (2015). *Implication of Children in Road Accidents in France in 2011*. 24th International Technical Conference on the Enhanced Safety of Vehicles (ESV).

Niebur, T., Kreiss, J.-P., Achmus, S. (2013). GIDAS-Aided Quantification of the Effectiveness of Traffic Safety Measures in EU 27. 23rd International Technical Conference on the Enhanced Safety of Vehicles (ESV), Korea, South Korea.

Op den Camp, O., van Montfort, S., Uittenbogaard, J., & Welten, J. (2017). *Cyclist Target and Test Setup for Evaluation of Cyclist-Autonomous Emergency Braking*. International Journal of Automotive Technology, 18(6), 1085–1097. https://doi.org/10.1007/s12239.

PEGASUS Project (2019). Pegasus method – an overview. PEGASUS research project overall method. https://www.pegasusprojekt.de/files/tmpl/Pegasus-Abschlussveranstaltung/PEGASUS-Gesamtmethode.pdf.

Rabe, M., Spieckermann, S., & Wenzel, S. (2008, December 7–10). *A new procedure model for verification and validation in production and logistics simulation*. In *2008 Winter Simulation Conference* (pp. 1717–1726). IEEE. https://doi.org/10.1109/WSC.2008.4736258.

Redlich, J., Schult, J., Schurath, R. (2022). *Automatisiertes Fahren – Auswirkung auf den Schadenaufwand bis 2040*. Study of Gesamtverband der Deutschen Versicherungswirtschaft e. V., https://www.gdv.de/resource/blob/74398/a3f836ecb82a30b65ccdc320ab4a8cf0/d-studie-data.pdf.

Rösener, C. (2020). *A Traffic-based Method for Safety Impact Assessment of Road Vehicle Automation.* Disseration RWTH Aachen University, https://doi.org/10.18154/RWTH- 2020-08950.

Saadé, J., Chajmowicz, H., & Cuny, S. (2019). *Prospective Evaluation of the Effectiveness of Autonomous Emergency Braking Systems in Increasing Pedestrian Road Safety in France*. IRCOBI. http://www.ircobi.org/wordpress/downloads/irc19/pdf-files/37.pdf.

SAE J3016:2021. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, https://www.sae.org/standards/content/j3016_202104/.

Scanlon, J., Kusano, K., Daniel, T., Alderson, C., Ogle, A., Voctor, T. (2021). Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain. Traffic Injury Prevention, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85018334131&doi=10.1080%2f15389588.2017.1300257&partnerID=40&md5=98963c1cfa286856cbab40a398a1a7fa.

Stipdonk H, Reurings M. (2012) *The effect on road safety of a modal shift from car to bicycle*. Traffic Inj Prev.;13(4):412-21. doi: 10.1080/15389588.2012.660661. PMID: 22817557.

Tarko. A, Davis, G., Saunier, N., Sayed, T., Washington, S. (2009). *Surrogate measure of safety.* White paper ANB20(3) Subcommittee on Surrogate Measures of Safety, https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=30801fa815159dad645eed6f1e3dbbbba2f30150.

Urban, M., Erbsmehl, C., Lorente Mallada, J., Puente Guillen, P., & Tanigushi, S. (2020-1). *A Methodology for Building Simulation Files from Police Recorded Accident Data (for ADAS Effectiveness Assessment)*. FISITA World Congress. https://www.fisita.com/library/fisita-world-congress/2020/f2020-pif-039.

Urban, M., Erbsmehl, C., Landgraf, T., Pohle, M., Lorente Mallada, J., Puente Guillen, P., & Tanigushi, S. (2020-2). *Parameterization of Standard Test Scenarios of Automated Vehicles Using Accident Simulation Data*. FISITA World Congress. https://www.fisita.com/library/fisita-world-congress/2020/f2020-pif-038.

Van Noort, M., Bakri, T., Fahrenkrog, F., & Dobberstein, J. (2015). *SIMPATO-The safety impact assessment tool of interactive*. IEEE Intelligent Transportation Systems Magazine, 7(1), 80–90. https://doi.org/10.1109/MITS.2014.2340054.

Weber, H., Eckstein, L., Beringhoff, F., Josten, J. (2023) Towards Modeling Driver Performance Within Crash-Relecant Scenarios As Virtual Reference of the Safety of Automated Vehicles. 27th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Yokohama, 2023.

Wimmer, P, Op den Camp, O., Weber, H., Chajmowicz, H., Wagner, M., Lorente Mallada, J., Fahrenkrog, F., Denk, F., (2023). Harmonized Approaches for Baseline Creation n Prospective Safety Performance Assessment of Drivin Automation Systems. 27th Enhanced Safety of Vehicles (ESV) conference, Yokohama, Japan.

Appendix A User Roles

Table A.1 V4SAFETY User Roles and motivation.

Who	Motivation for using V4SAFETY	How will their results using the V4SAFETY assessment output be applied?	Success factors	Bottlenecks	Key requirements
Specialist in safety assessment	• This is the daily work of this user. • The specialist would like to follow standards to avoid debate on the methodology. • The specialist would like to follow reporting standards to avoid misinterpretation of the results.	• The safety assessment specialist will provide an advice to system developers and development managers within his organization about the effect of the safety measure proposed. • The specialist will use the reporting templates to show the conclusions, assumptions and uncertainties in a standardized way.	• The framework should fit the workflow of the user organisation. • Compliance with internal standards • Good level of expertise with indirect users • Experience in a-priori assessment methods	• Ambiguity in V4SAFETY framework leading to discussions on results.	• Framework should provide a way to address the unavoidable uncertainty in the assessment. • Support the creation of comprehensive safety assessment reports and documentation for internal and external stakeholders, including regulatory authorities. • Scalability to divers safety measures and complex road user interactions. • Explicit and well-structured documentation of the assessment choices and uncertainties.
Human behaviour modelling specialist	• For efficient simulation of safety measure performance, human models are very useful. Also, they have a fast and reproducible output. However, they need to be based on a solid, accepted foundation. • The human behaviour specialist is highly motivated to have	• Evaluate the consequences of different human behaviours on road safety. • Evaluate outcome-based validity of models compared to real-world. • Utilise data requirements. • Consideration of standardized interfaces • Steps in using it: • Review and utilize human behaviour model	• The available behaviour models can be integrated into the framework with limited efforts. • V4SAFETY deliverables are referenced in future road user behaviour research publications.	• Lack of experience in the usage of a complex simulation environment • A functional simulation framework is not (yet) available. • Proposed framework supports more an	• Framework description and application understandable for non-engineers • Proposed simulation structure enables integration of various behaviour models. • Right balance between engineering and behaviour science language • WP3 uses behaviour science language.

	his/her models utilised in development and V&V since human behaviour contributes substantially to crash causation. • Still lot of knowledge gaps how to model human behaviour in an appropriate way • Theoretical models need to be tested and compared to real-world data using an accepted procedure. • In-vehicle safety systems depend on human interaction, so understanding this interaction is important for safety.	documentation / literature review. • Guideline for model implementation • Identification of further research areas • Method reference for future publications		engineering approach and less a scientific approach.	• Right balance between an engineering and a scientific approach • Openness of the framework for future development.
Virtual simulation expert	• The virtual simulation expert needs to follow the state-of-the-art to strategically plan the road map for virtual simulation and V4SAFETY aims to build upon state-of-the-art. • V4SAFETY will highlight existing limitations in system behaviour that are most relevant to address in the future. • The output of V4SAFETY is a harmonized framework and communication with external collaborators will be	• To verify compliance with the V4SAFETY methodology. • Align check lists and documentation to the result format of V4SAFETY.	• Reference to V4SAFETY deliverables in future virtual simulation methodology publications • Application of some or all aspects of the V4SAFETY framework in the proprietary organizational framework.	• Transferability of the V4SAFETY framework is limited due to lack of generalization. • Time / cost efforts are too high to evaluate the V4SAFETY framework.	• Framework should cover all feasible and suitable measures. • Modularity of the framework should allow for step-by-step partial integration.

	easier for the expert if the harmonized aspects are considered.				
Virtual test engineer	• The virtual simulation engineer needs to share simulation results. Therefore, a harmonized and accepted approach reduces efforts to gain trust in simulation results. • Exchange with collaborators outside of the company is simplified due to a common approach. • Internal communication of the simulation method and the results is simplified as reference to the V4SAFETY documents can be made.	• To support and guide development of traffic safety measures. • To review necessary improvements utilizing validation and verification processes.	• Method and results of safety assessments follow the V4SAFETY framework. • Virtual Simulation Engineers are familiar with the methods of the framework.	• The alignment to the V4SAFETY framework is too time / cost intensive. • The V4SAFETY methods do not fit to the workflow of the simulation engineer set by the organization.	• Documentation is available on how to apply the framework. • Modularity of the framework allows for step-by-step integration.
Traffic and infrastructure safety specialist	• Infrastructure safety treatments are costly and simulation tools allow to select the most appropriate measures. • the complex environment requires an appropriate simulation framework. • Traffic data are extremely variable and proper data management is required. • Preventive cost effectiveness	• Infrastructure improvements • Traffic engineering interventions • Policy recommendations • Standards revisions • Targeted enforcement • Resource allocation	• The framework should be flexible, easy to use and allow also the use for non specialists is simulation. • The simulations are often conducted on a very large scale and therefore reducing computation and model construction time can increase the success.	• Lack of knowledge and expertise. • Data availability (and quality). • Limited resources • Time constraints (for building the model and for computation).	• The framework should be adaptable to very different contexts (different countries, different environments, different road code rules etc.) • It should allow the use for different levels of data availability (and quality). Reduced data availability should be allowed even though it will result in a lower reliability of the simulation. This should be made clear to the user. • Provide clear application boundaries to avoid

	evaluation allows comparison/benchmarking of safety treatments. • Validation procedures are essential for reliable results to define both “fidelity” and “validity” of the simulation context. • Proper scientific evidence can support policy formulation and standards revisions.				extrapolations to inapplicable contexts.
Infrastructure design engineer	• Estimate the potential effectiveness of different safety treatments. • Solve complex design issues that design standards. • Compare the potential effectiveness of different safety treatments. • Analyse non conventional design issues.	• To support decision makers in the selection of the most appropriate measure. • As a basis for Road Safety Impact Assessment. • To support innovative safety measures.	• The process needs to be clear and straightforward. • Provide simplified simulation frameworks.	• Simulations are often time consuming and not compatible with design schedules. • Tools and models need to be applicable to the specific type of intervention and network.	• Can handle a wide variety of safety measures. • The framework should be adaptable to very different contexts (different countries, different environments, different road code rules etc.). • Provide clear application boundaries to avoid extrapolations to inapplicable contexts. •
(Safety) System engineer	• This role can include in some companies be a pure system engineer and in other organisations a system engineer that is also responsible for the safety concept. • The system engineer wishes to design and maintain a system architecture that is robust and sufficient for the vehicle or safety	• The system engineer will specify the system under development at a high level (functions, total weight, power, etc). • They will specify the safety measure at high level (function, response curve, volume, etc). • This is then transferred to a safety assessment specialist to be analysed.	• The system engineer receives the assessment results and decides whether a redesign is needed. If not, the concept is transferred to engineers for further development.	• Availability of a specialist with experience in V4SAFETY Safety Assessment Framework • Capability of the framework to handle loosely defined systems and safety measures	• Can handle a wide variety of systems and safety measures. • Assessment result is reliable or uncertainty is quantified.

	measure under development and likely variants. This has a direct interaction with the safety concept of this vehicle: architecture choices influence isolation or propagation of failures, sensitivities or perception issues. • The system engineer wishes to explore the safety impact of a possible system to judge whether the additional development and component costs of a safety measure produce sufficient safety benefits.				
Physical test engineer	• The physical test engineer needs to understand how an assessment by virtual simulation is conducted to identify the needs for physical validation. • The validation of virtual testing against physical testing should be conducted in a combined effort of the virtual and physical test engineer as each has the best understanding of respective test specifications and method limitations.	• Comparison of the virtual simulation results against physical simulations. • Identification of limitations in transferability of assessment results. • Request results on model or sub-system level that may not be part of a virtual assessment report	• The physical test engineer understands the validation needs for a virtual simulation and can prepare testing without resource-consuming iterations.	• The organisation does not enable / stimulate cross-functional exchange between virtual simulation and physical simulation.	• The validation requirements must be formulated in a way that physical test engineers understand. • The validation requirements should be feasible and follow best practise of physical test engineers. • The metrics to describe and affirm validity should be accessible in physical testing.

	• The test engineer needs to be informed about potential future validation needs and strategies				
Requirements Engineer	• The Requirement Engineer needs to understand a priori the real-world consequences of one or more concept designs. • The Requirement Engineer wishes to update requirements if a safety assessment provides results with unsatisfactory safety. • Virtual simulation is a cornerstone for future legal and consumer rating, thus, knowledge on opportunities and limitations is essential to apply it in a productive way.	• Evaluate safety concepts and define requirements based on the outcome of virtual simulation. • Evaluate virtual assessments of competitors to observe the market and adapt to the state-of-the-art.	• Virtual simulations aligned with V4SAFETY prove to guide towards the most effective concept measure regarding the defined evaluation metrics. • A-posteriori evaluations using field data show similar results and confirm the validity and usability of virtual simulation for a priori assessment.	• Cases where the virtual simulation results show major deviations from field data analysis. • The trust in virtual simulation has vanished due to major limitations in representativeness or Misguided expectations about simulation.	• The V4SAFETY framework should highlight relevant requirements on the validity of virtual simulations. • The V4SAFETY framework should be adaptable for various types of analysis required for requirement definition.
Regulation & rating expert	• The usage of virtual simulation is currently evaluated for the assessment of legal and consumer requirements. The regulation and rating expert must understand the requirements on transparency, applicability, and validity of virtual simulations. • The results of virtual simulations have to be	• Review results and assess their representativeness for real-world safety. • Conclude on the validity of virtual simulations as physical test replacements.	• Virtual simulation following V4SAFETY guidelines provide reliable results that stakeholders share with authorities or consumer test organisations. • Scalability from few tests to diverse test population is validated.	• Acceptance of virtual simulation results by consumers. • Prevention of manipulation to virtual simulation validation & verification. • Representativeness of utilized test scenarios for real-world.	• Transparent validation and verification process of virtual simulations, for example against physical tests. • Prevention of manipulation on results. • Specific training how to utilize virtual simulation results and identify validity.

	put into context of real-world safety to decide on the suitability for replacement of or extension on physical tests.				
Type approval engineer	• UN-ECE indicates the need for including virtual simulation in type approval for CCAM. • Ensure that the national situation is considered 'sufficiently' in the type approval process, such as local infrastructure, modal split, etcetera. • Meet EC-legislation.	• Review results provided by industry as part of the type approval process. • Review assumptions and models used, with their limitations. • Follow UN ECE NATM multi-pillar approach • Follow national implementation act procedures. • Steps in using it: • Review safety assessment documentation, partially according to V4SAFETY templates • Decide on additional tests to be performed by OEM (in simulation). This may contain some random sampling for certain object types or parameter values. • Define process to judge the input, process and the results – compare. to type approval requirements • Follow this proposal. • Set up periodic technical inspection – to check the potential degradation of system performance. • Review periodic OEM reports from in-service monitoring.	• Reproducibility, comparability • Explainability – on test selection, process and results, also for consumers • Trust and cooperation between type approval authorities and the applicant of the type approval. • Sufficient resources on the side of the type approval authorities (external success factor for use of V4SAFETY framework) • Training of the type approval authorities (external success factor for use of V4SAFETY framework).	• Knowledge and expertise needed on the simulation framework/process. • Confidence/trustability on the validity of the results • Availability of data on new technologies, transportation modes • How to deal with Intellectual Property issues? • How to handle unknown scenarios that pop up during the lifetime of the vehicle type?	• Clarity on the origin of the data, data collection process (related to Rq-30). • Inspection procedure including random selection of input data and check on the results new requirement? (related to Rq-53). • Procedure is standardized (Rq-40, Rq-46, Rq-49).

		• Review OEM incident reports and corresponding mitigation proposals • Monitor period technical inspection.			
Consumer testing specialist	• Consumers benefit from a consistent and transparent approach for safety assessment. The first level of judgment for a consumer is a star rating. However, the next step would be to read a review of the assessment. Knowing that this is based on an accepted framework will confirm trust in the star rating. • The consumer testing specialist wishes to create a reliable consumer rating of a new vehicle for consumers. • The consumer testing specialist wishes to adopt a broadly accepted method for safety assessment so that the result is comparable with other NCAP bodies and there is little debate about the result.	• The V4SAFETY results will be translated into a star rating. • The V4SAFETY documentation will be used as input to a full text description of the consumer test and star rating. • Steps in using it • The consumer testing specialist will define an evaluation scope. • The consumer testing specialist will pilot an assessment to determine the range and threshold values for certain star ratings. • The consumer testing specialist will adopt a baseline approach. Traditionally this was A, but this will more often become B or C for automated driving. • The consumer testing specialist will delegate the execution of a simulation to a virtual test engineer. • He will evaluate the safety performance based on the simulation results.	• The V4SAFETY framework must be easy to explain to consumers. • The uncertainties in the assessment result shall be clear for the consumer testing specialist.		• The baseline data shall be sufficient to support application of the star rating for several years.
Developer using simulation	• Assess the safety performance of system designs in the early stages of development	• Run sensitivity analysis prior to setting design parameter values.	• Guidelines for using the framework are clear for non-experts of accident science.	• Design departments not equipped with necessary	• Short learning curve: understanding the approach of the framework in one

	– check whether the design is going in the "right" direction.	• Get guidance on how to do cost-benefit analyses (e.g., cost of a sensor/system vs safety benefit) • Steps in using it: • Identify the most influential parameters. • Assess safety performance of system being designed. • Compare costs to benefits (to society, to companies	• Applying the framework must be time efficient (application in a matter of weeks]). • Applying the framework is user-friendly and "straightforward" (no need to come back to the documentation each time framework is used): usability in a matter of days.	models or the necessary interfaces between models. It might take a high financial effort necessary to become equipped. • Manpower not available in design department with sufficient knowledge to use framework. • Risk: Minimum requirements for framework application too high	day, using it in full in two days. • and usage curve "steep"
Policy maker	• Earn money (within projects) • Saving time to not come up with new things. • Improving own tools (be inspired by framework) • Gaining credibility/ transparency/ acceptance for results • May be requirement from customer to use framework. • USP to sell application of this specific framework.	• Communicate model to customer. • Check improvement/ comparison what is the best measure. • Interpreting of results to give profound advice to customers. • Steps in using it • Translate customer question into processible within framework. • Identifying relevant steps to use according to the question. • Depends on the case.	• Flexibility to use own tools within framework. • Easy access to find information about framework. • Easy understandability • Easy applicable • Fast production of results specifically for the use case • Prevent misuse to prevent loosing credibility.	• Potentially limited knowledge of consultants to apply framework. • Policy advisors might not be aware of the framework. • Misuse of framework • Maintenance / outdated?	• Adaptability / Flexibility to framework • Clear recommendations • (Basic evidence about accuracy) • Quick start guide (How to apply V4SAFETY 101) • Application without too deep background • Maintenance