SAFESPOT INTEGRATED PROJECT - IST-4-026963-IP

DELIVERABLE 4.6.1

SP4 – SCOVA – Cooperative systems applications vehicle based

Pilot Plan

Deliverable No. (use the number indicated on technical annex)		**D4.6.1**	
SubProject No.	**SP4**	**SubProject Title**	**SCOVA - Cooperative systems applications vehicle based**
Workpackage No.	**WP6**	**Workpackage Title**	**Validation**
Task No.	**T4.6.1**	**Task Title**	**Pilot Plan**
Authors (per company, if more than one company provide it together)		**Frederik Diederichs (USTUTT), Dehlia Willemsen, Arjan van Leysen (TNO), Giulio Vivo (CRF), Maria-Carmela De Gennaro (MMSE), Harri Koskinen (VTT), Erik Nordin (VOLVO)**	
Status (F: final; D: draft; RD: revised draft):		**F**	
Version No:		**1.2**	
File Name:		**SF_D4.6.1_Pilot_Plan_v1.2.doc**	
Planned Date of submission according to TA:		**31/01/2009**	
Issue Date:		**20/04/2009**	
Project start date and duration		**01 February 2006, 48 Months**	

Revision Log

Version	Date	Reason	Name and Company
0.1.0	15.10.08	Asking partners for contribution by filling out template	F. Diederichs (USTUTT); A. van Leijsen (TNO)
0.1.6	10.02.09	Integrating Partners contribution and working on fusion.	F. Diederichs (USTUTT)
0.1.7	14.02.09	Clarifying Introduction and content structure.	F. Diederichs (USTUTT)
0.1.9	21.02.09	Comments by Dehlia (TNO) and further work on Evaluation by Frederik.	D. Willemsen (TNO; F. Diederichs (USTUTT)
0.2.1	25.02.09	Release for first internal draft review.	D. Willemsen (TNO; F. Diederichs (USTUTT)
0.2.2	03.03.09	First draft review	G. Vivo (CRF)
0.3	10.03.09	Integration of provided contributions	F. Diederichs (USTUTT)
0.5	22.03.09	Comments by VOLVO integrated	F. Diederichs (USTUTT)
0.6	22.03.09	Stable version for internal peer review	F. Diederichs (USTUTT)
0.7	25.03.09	1) added numeration to the references 2) added connections to the references within the deliverable, 3) added missing references cited in the deliverable 4) added connections between paragraphs, chapters 5) modified the chapter for evaluation methodology 6) general revision of the doc with comments added	M. C. De Gennaro (MMSE)
0.8	26.03.09	Revision by Continental Automotive	K. Belhoula, H. Pu (CA)
0.9	30.03.09	Peer review by CRF	G. Vivo (CRF)
1.0	05.04.09	Integrating all comments and release for Peer Review	F. Diederichs (USTUTT) G. Vivo (CRF)
1.1	17.04.09	Integrating Reviewers comments and contributions by F. Peyret (LCPC) Compilation of Annex document	F. Diederichs (USTUTT) G. Vivo (CRF)
1.2	20.04.09	Final Version	F. Diederichs (USTUTT) G. Vivo (CRF)

Abbreviation List

ADAS	Advanced diver assistance systems
BLADE	Business models, Legal Aspects, and Deployment
CAR	Correct Alarm Rate
COSSIB	Cooperative Safety Systems Infrastructure Based
FAR	False Alarm Rate
ESPOSITOR	SAF**ESP**OT **SY**STEM MONI**TOR**
FCW	Frontal Collision Warning
FN	False Negative (or Missed Alarms)
FP	False Positive (or False Alarms)
GW	Gateway
HF	Human Factors
HIL	Hardware In the Loop
HLO	High Level Objective
HMI	Human Machine Interface
HURR	SAFESPOT High level objectives, User needs, Requirements and Risks
HW	Hardware
IP	Integrated Project
ISO	International Standard Organisation
JDVS	Joint Driver Vehicle System
LAN	Local Area Network
LDM	Local Dynamic Map
LIVIC	Laboratoire sur les interactions vehicules-infrastructure-conducteurs: Research laboratory for advanced driving assistance systems and cooperative systems
LR	Long Range
MAR	Missed Alarm Rate
MARS	Multi Agent Real-Time Simulation Technology
NASA TLX	National American Space Agency – Task Load Index
PC	Personal Computer
PReVENT	Preventive and Active Safety Applications (EU-Project)
R&D	Research and Development
RIS	Road Intersection Safety
RSU	Road Site Unit
SAFEPROBE	In-Vehicle Sensing and Platform

SBC	Single Board Computer
SCOVA	Cooperative Systems Applications Vehicle Based
SINTECH	Innovative Technologies
SMA	Safety Margin Assistant
SP	Subproject
SUS	System Usability Scale
SW	Software
TN	True Negative
TP	True Positive
UDP	User Datagram Protocol
VANET	Vehicle Ad Hoc Network
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
V&V	Verification and Validation
WP	Work Package
WT	Work Task

Table of contents

List of Figures

List of Tables

Partners

The Deliverable was compiled by USTUTT; TNO; MMSE; VTT and VOLVO.
The authors thank also CRF and CA for valuable comments based on a detailed internal review.

EXECUTIVE SUMMARY

This deliverable describes the Pilot Plan adopted for the evaluation of the Safety Margin applications based on the SAFESPOT cooperative system. The pilot plan defines the methods and the criteria for evaluating the Safety Margin Assistance applications.

Pilot tests will be assessed through the utilisation of simulation tools, as well as in the Test Site environment and in Driving Simulators, based on the vehicles prototypes developed within SP4-SCOVA.

The evaluation based on the utilisation of the vehicle prototypes will involve experts and subjects and will be performed on the different test sites and simulators that are currently available or in development within the SAFESPOT IP. In particular, different vehicle prototypes (cars, trucks, motorcycles) will be used in the different pilot tests.

A dedicated pilot plan test case template has been produced for a common and standardized test planning of all the pilots to be conducted in SP4. The test case template is described with detail in the body of the deliverable, while **the set of the SP4 tests planned so far is collected and reported in the annex of the present document**. In order to achieve common and comparable results, a standardized reporting approach is proposed.

Concerning the methodological aspects, the ISO 9001 standard and the V - model approach have been used for the planning and for the definition of the criteria to adopt for the evaluation.

In the definition of the Pilot Plan, a specific effort has been spent to define and to clearly point out the procedure to follow in order to obtain consistent results in the pilot tests, homogeneous and harmonised at the overall IP level. Such relevant achievement will be easily reached, during the validation trials, if the tests will be conducted by complying with the following simple workflow, constantly adopted during the planning stage:

1) starting from high level objectives, user needs, requirements;
2) taking into account the corresponding demands in the evaluation process;
3) defining appropriate evaluation criteria, indicators and test cases for the evaluation.

The pilot plan test case template developed in WP4.6 has the purpose of guiding the test leaders to follow this workflow. As proposed and recommended by SP4, the adoption of the same methodology, approach and templates by the different SPs and TSs involved in the last year of activity of the SAFESPOT project, will ensure to obtain the validation activities and documentation for the entire Integrated Project consistent and well aligned.

1. Introduction

SAFESPOT aims to develop a platform for V2V (vehicle to vehicle) and V2I (vehicle to infrastructure) communication. Such a system would extend by far the possible environmental recognition of current sensor technologies. Any individual on-board sensor system will meet its limits at some extend (i.e. detection of objects around curves) so the SAFESPOT approach will enable vehicles to communicate relevant information such as ego-vehicle and surrounding environment parameters to other vehicles and to the infrastructure. In SAFESPOT SP4 (SCOVA) the consortium developed applications that use this cooperative technology, with the main focus on V2V cooperation. Together with the cooperative V2I applications developed in SP5 (COSSIB) and the basic platforms assembled in SP1 (SAFEPROBE) and SP2 (INFRASENS) which contain sensing modules, positioning, communication and the local dynamic map developed in SP3 (SINTECH), it aims to develop and evaluate the functionality of a number of safety applications such as:

- Road Intersection Safety (RIS)
- Lane Change Manoeuvres (LCM)
- Safe Overtaking (SO)
- Head On Collision Warning (HOCW)
- Rear End Collision Warning (RECW)
- Frontal Collision Warning (FCW)
- Vulnerable Road User detection and Accident Avoidance(VRUAA)
- Road Condition Status (RCS)
- Curve Warning (CUWA)
- Predictive Speed Limitation and Safety Distance (SLSD)

SAFESPOT WP4.6 is dedicated to evaluate these SP4 applications in terms of:

- Technical evaluation of the SAFESPOT vehicle to vehicle technology.
- Human factors evaluation of the applications with system users.
- The impact of a SAFESPOT system on the complete traffic situation.

The present deliverable D4.6.1 – Pilot Plan reports important testing methodology for technical testing, for testing with human participants and also methodological recommendations on how to test the effect of such a system. Consequently this deliverable aims for achieving identical results reporting for all evaluation tests and will try to connect the expected results to the SAFESPOT and SP4 objectives.

1.1. Innovation and Contribution to the SAFESPOT Objectives

This deliverable aims for evaluation of SAFESPOT SP4 applications. Its main challenge is to define the standardization of the high amount of (expected) pilots in the field of technical evaluation, human factors evaluation and traffic impact evaluation. The authors see the main innovation in the approach of focusing explicitly on the fulfilling the SAFESPOT pre-defined high level objectives in the evaluation pilots. Since the high level objectives are rather general ones, further success criteria were derived by the SAFESPOT consortium. These include the mandatory identification of the user needs and requirements relevant for the specific tests as well as the list of risks that might jeopardize the success of the applications.

1.2. SAFESPOT SP4 objectives

SAFESPOT WP4.6 has as main objective the evaluation of SP4 applications with respect to pre-defined success criteria. These success criteria are distributed in 4 clusters of pre-defined objectives:

- High Level Objectives
- User Needs
- Requirements
- Risks.

These objectives are further referred to as HURR, taking their initials as acronym. The evaluation plan aims to prove the success of SAFESPOT SP4 by complying with the HURR.

The SP4 pilots have been designed in order to gather as many results as needed to measure the success of reaching the SAFESPOT objectives. This is ensured by a common pilot plan test case template that will be used by every single pilot test case (see chapter 5). In the pilot plan test case template the SAFESPOT HURR are pointed out explicitly and a mandatory step in the execution of the test sessions is to identify and state explicitly how the pilot will contribute to the SAFESPOT HURR.

1.2.1. High Level Objectives

The following high level objectives (HLO) have been extracted from the SAFESPOT Technical Annex [1] and from the SAFESPOT web page for SP4 [2]. They are cited here as stated in the original source and as used in the template for pilot test case planning.

- The main aim of the project will be to show the feasibility and benefits of Co-operative Systems solutions in improving road safety well beyond the level which can be achieved with autonomous solutions (vehicle or infrastructure based).
- To demonstrate benefits for accident types with a calculated potential safety impact (in terms of saving of lives as well as other gains).
- To improve the range, quality and reliability of the safety-related information.
- To support drivers preventively to the proper manoeuvres in the different contexts; (to optimize the intervention of vehicle controls with respect to critical situations).
- To manage existing incidents to minimize further negative safety impact.
- To increase safety for all road users in a specific situation.
- To show that the safety impact can be achieved without affecting transport efficiency.
- To increase the Safety Margin of vehicles using in-vehicle and infrastructure information.
- To create applications for extended cooperative awareness by means of real time reconstruction of the driving context and environment.
- To open the development of new safety applications based on a cooperative approach.

The HLO are on a very general level and therefore not easy to meet with a limited number of tests. However it is possible to contribute to some aspect of each HLO with specific tests. The partners that plan the pilots are therefore asked to specify which possible result of their tests can actually contribute to the HLO. The HLO are stated in the pilot plan test case template where this assignment is specified.

1.2.2. User Needs

The User needs for SP4 have been derived from the HLO and are explained in D4.2.4 - User Needs and Requirements [3]. The user needs are defined and specified for each SP4 application. They are in most cases more detailed than the HLO and therefore it is easier to define measures that indicate if a user need is met or not.

1.2.3. Requirements

The requirements for SP4 have been derived from the HLO and user needs and are explained in D4.2.4 - User Needs and Requirements [3]. The requirements are defined and specified for each SP4 application. Additionally a list of common requirements has been defined that applies for all applications. The requirements are the most specific success criteria and the definition of measures indicating if the requirements are met or not is usually relatively easy. It should be underlined that the evaluation of the components performance and in general the compliance with the SP4 specifications is out of the scope of the pilot plan activities, since these aspects are already addressed and validated in the previous phases of the work (WP4.5 activities). Hence, only those requirements are in the scope of SP4 WP6 evaluation that can be directly related to higher level objectives (user needs, high level objectives) and that can only be tested with a fully integrated system available on the TSs for the WP6 evaluation.

1.2.4. Risks

SAFESPOT SP6 defined a list of possible risks that might influence the success of the deployment of SAFESPOT services. Consequently this could impair also meeting the SP4 success criteria in the evaluation pilots. These identified risks are separated into two categories: risks that could still be solved during the SAFESPOT project (project management risks) and deployment risks, for which an overall strategy has been defined to address the risks and to take the required contingency actions (see D6.4.5 – Preliminary Recommendations [4]). The project management risks, during the SAFESPOT project period, have been handed over to both SP4 & SP5 test leaders in order to verify if the relevant risks for the cooperative applications are properly treated. For this reason such risks are made explicit and directly included in the pilot plan test case template document. In this way, it is highlighted that among the purposes of the test sessions, one of the most relevant consists in the verification that all major risks for the application under test are still present or successfully excluded in the development.

1.3. Link with other SPs and WPs

This deliverable is closely linked to a number of SPs and WPs. Starting with SP4 (SCOVA), WP4.6 is the completion of the development process of applications, evaluating the final applications in terms of technical function (WT4.6.4), human factors (WT4.6.5) and traffic impact (WT4.6.6) and it is expecting verified applications and technical systems at the test sites, provided by WP4.5. The pilot plan deliverable is based on D4.2.4 – User Needs and Requirements [3], plus some high level objectives presented in the Technical Annex and on the SAFESPOT Web Site. Risks have been taken from D6.4.5 – Preliminary Recommendations [4].

WP4.5 has the task of verifying the technical functionality of the system parts (WT4.5.1) and of the complete system integrated into the SAFESPOT prototype vehicles (WT4.5.2). The test plans for systems parts and integration tests have been compiled using the specification deliverables: D4.3.1 – Safety Margin Application

Parameters: Analysis and Characterization [5], D4.3.3 – Application communication for co-operative vehicles and infrastructure [6], D4.2.4 – User Needs and Requirements [3].

SP4 (SCOVA) applications are mainly based on subsystems developed in SP1 (SAFEPROBE) and SP3 (SINTECH) (see Figure 1).

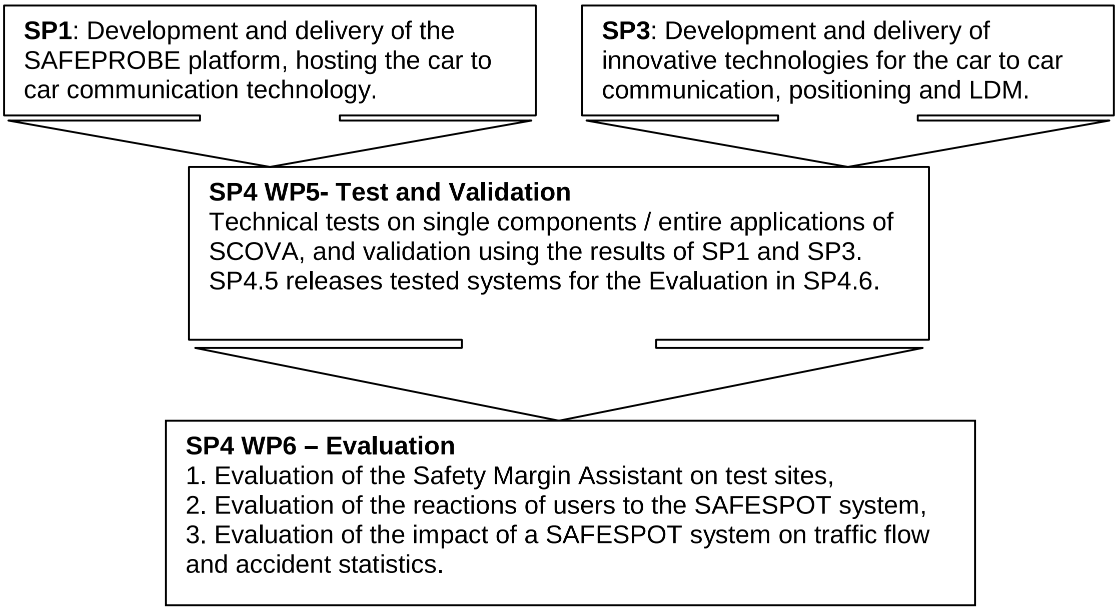


Figure 1: Interdependency of SP1 and SP3 towards WP4.6.

SP5 (COSSIB) is dedicated to applications based on the communication between vehicles and infrastructure. These applications will be evaluated in a similar approach by SP5 (COSSIB) and thus a close contact was established to the counterpart WP5.6.

SP6 (BLADE) contributes to the success criteria by defining the risks that might jeopardize the success of the applications.

The results obtained from the pilots will also be the basis for the SP6 (BLADE) efforts to promote SAFESPOT applications and platforms to a wider group of stakeholders through cost/benefit analysis. Finally the results obtained from the pilots will be key arguments for further research and developments on the SAFESPOT basis.

The strongest link within SAFESPOT can be seen to the 5 test sites and simulators. All WP4.6 pilots will be conducted at the sites. The pilots are planned in a cooperative work with test site leaders and SP4 application leaders.

1.4. Deliverable structure

The deliverable content starts in chapter 1.2 giving an introduction to the pre defined SAFESPOT SP4 objectives which take the role of success criteria for the evaluation of SP4 applications. Direct links to other related parts of the SAFESPOT project are highlighted in chapter 1.3. Chapter 2 gives a description of the applications that will be evaluated on the different test sites. In chapter 3 the V-model is described, which is the base for the methodology for SAFESPOT SP4 evaluation. In chapter 4 the common testing methodology in all pilots is addressed by proposing a set of basic indicators that should be applied to the evaluation of the SMA. Chapter 5 introduces the pilot plan test case template developed in WP4.6 and explains the procedure of test case planning with this template as performed within WP4.6. It also summarizes all pilot test cases that are already planned by the responsible application and test site leaders. Chapter 6 aims to standardize the results reporting by defining chapters for the internal report that follows each carried out pilot. In this chapter also video and picture taking is encouraged and the non compliance form used in all SAFESPOT SPs is introduced, since it will play a vital role during the final testing of the applications. Finally, **an annex document** has been prepared, **collecting all the tables with the operative content of the Pilot Plan**, which consists of the description of the pilot tests planned so far for the SP4 SCOVA applications in the different TSs.

2. SAFESPOT SP4 applications and test sites

In this chapter the SP4 applications are shortly summarised. A more extended and detailed description of the applications is part of D4.2.2 – Safety Margin Concept [12] and of D4.2.3 – Use Case and Typical Accident Situation [13]. These applications are based on hardware and software components from SP1 and SP3, in addition to the ones of SP4; hence the technical function of every application is directly dependent on the function of each component and on the integration. The functional testing of the components and the integration is achieved in WP4.5. The applications are installed in vehicles on SAFESPOT test sites, which are described in the Chapter 2.3. Detailed information about the test sites will be available in D4.6.2 – Pilot Plan Assessment [12].

2.1. Description of applications

In the following list the SP4 applications included in the test plan are shortly presented. Most of the applications use primarily ego vehicle data and information derived from the vehicle to vehicle communication. For some other applications, namely: Speed Limitation and Safety Distance, Lane Change Manoeuvre, Road Condition Status and Curve Warning, some limited usage of information provided by the infrastructure is included. For this reason, and within this limited domain, the co-operation with infrastructure sensed data is also included in the SP4 pilot tests. The applications to be tested in SP4 are:

- **Speed Limitation and Safety Distance (SLSD)**. The vehicle is informed via an infrastructure about legal speed and distance to keep in a specific road area. The safety margin is calculated by the state of the vehicle in front (it can be increased for dangerous goods for instance).
- **Road Intersection Safety (RIS)**. General intersection scenarios which ensure in advance that the driver will cross the intersection in a safe way (for instance in case of disobeying of the right of way rules from other vehicle).
- **Lane Change Manoeuvre (LCM)**. To prevent, during the road merging situations and approaching to the intersections, the risk of lateral collisions; safe lane manoeuvre with blind spot for trucks.
- **Safe Overtaking (SO)**. Refers to a collision where one vehicle which is overtaking collides with another vehicle coming behind.
- **Head On Collision Warning (HOCW)**. Refers to a collision where the front end of one vehicle collides with the front end of another vehicle while the two vehicles are travelling in opposite directions.
- **Rear End Collision (REC)**. A collision in which one vehicle collides with the rear of another vehicle. Safety Margin Assistance for preventing fender bender situations and backward collisions.
- **Frontal Collision Warning (FCW)**. The prevention of frontal collisions is achieved by using the knowledge of some obstacle in front of the vehicle (for

instance after a curve): it can be an accident, such as static obstacles, or a congestion with slow moving traffic.

- **Road Condition Status (RCS)**. Possibility to share with other vehicles information of slippery road status or bad condition, due for instance to weather condition, ice, fog... Information can be provided by other vehicles or infrastructure.
- **Curve Warning Application (CUWA)**. Digital map data informs the driver of the road curvature and the system evaluates whether the vehicle speed is adequate. The road curvature ahead is calculated dynamically, by a cooperating infrastructure, or statically, deriving it from the map data.
- **Vulnerable Road User detection and Accident Avoidance (VRUAA)**. After detection of a VRU, information is sent to the vehicle incoming behind (scenario is interesting in a 2 ways road in urban situation). Avoid accident with VRU on the side of the vehicle when it decides to turn. VRU can be a pedestrian, a bicycle or a motorcycle. This is a frequent accident with trucks and blind area.

2.2. Description of vehicles, components and RSU

The SP4 applications will be integrated and tested in cars, trucks and motorcycles. A detailed presentation of the respective vehicles can be found in D4.4.2 - Equipped Cars Integrating the Safety Margin Applications [7], D4.4.3 - Equipped Trucks Integrating the Safety Margin Applications [8] and D4.4.4 - Equipped Motorcycles Integrating the Safety Margin Applications [9].

The vehicle-based platform, is a delivery of SP1, SP3 and SP4 subprojects. Its hardware components are: Positioning PC (Pos PC), Main PC, Application PC (Appl PC), VANET Router, Gateway (which depends on OEMs: CRF gateway VGX SBC, Volvo GW, APX-Piaggio GW, TNO GW, LIVIC GW, Renault GW), ESPOSYTOR PC (ESP PC) and Ethernet Switch, providing LAN networking between the different computers. In some applications additional HW components like IBEO LS (IBEO laser scanner), and Road eye (1st gen) are used. The acronyms in brackets are used in Table 1 below (chapter 2.3).

The software components running on the above mentioned hardware are: gateway software (GW SW), VANET router software (VANET router SW), positioning software (Pos SW), SP1 framework (running on the Main PC), ESPOSYTOR software and application software (Appl. SW). In some cases, for instance in the VTT camera, special software (infrastructure based pedestrian recognition) is used.

Some SP4 applications are based not only on vehicle components but also on RSU. RSU are explained in detail in D5.2.2 – Common Architecture and Communication [10].

2.3. Description of test sites

Test sites are aiming to demonstrate the applications and use cases developed in the different subprojects and to prove interoperability among different countries. Five test sites, located in six European countries, are available in SAFESPOT:

- TS West, located in Western Europe: France and Spain (WE)
- TS Italy (IT)
- TS Germany (DE)
- TS Netherlands (NL)
- TS Sweden, (SWE)

SP4 applications will be validated in West, Italy, Netherlands and Sweden test sites. Test site Germany is not used for SP4 testing. The location of test sites is shown in Figure 2, while a short description of SP4 relevant test sites follows.

Figure 2: Location of test sites.

TS West contains test locations in France and in Spain. Two locations used for integration, validation, controlled environment preliminary tests and full tests too dangerous to be carried out on an open road are the LIVIC closed test tracks in Satory (Versailles) and the Volvo closed test track in La Valbonne nearby Lyon. Four real-world locations representing different types of network are the A85 motorway sections, located at Vivy, near Saumur, the RN 786 Etables s/ Mer rural section in the Côtes d'Armor, the RD 8 Bourbriac rural section in the Côtes d'Armor and the Spanish pre-urban sections in Valladolid technology park.

TS Italy main test location is located on the Torino-Caselle Airport Expressway. Other two locations, close to the first, will be devoted to the evaluation of the application for urban and rural environments. The CRF closed test track will be used for tests potentially dangerous or not practical to perform on public roads. Finally a second highway segment on the Brescia – Padova motorway will be used for further evaluation activities.

TS Netherlands has three different locations at which different applications will be tested and evaluated. The main test location is an urban intersection in Helmond located in the south east of the Netherlands. Other two test locations are the rural road N629 located in the South of the Netherlands, and the highway A16 from Antwerp to Rotterdam located at the South-West of the Netherlands.

TS Sweden contains test locations mainly in Gothenburg. Test locations are highway E6 between the northern and southern entrances to the city and a city tunnel named Lundby tunnel. For testing of applications and situations that for safety reasons require closed areas the proving ground at Stora Holm will be used. There will also be a test area at Lindholmen Science Park, near the Control Centre of the test site, for easy demonstrations.

Further information about each test site is available in D8.4.3 - Common Validation Plan [11]. More information about how the test sites are used for evaluation of SP4 applications will be available in the deliverable D4.6.2 – Pilot Plan Assessment [12].

As a summary, Table 1 gives an overview of the applications, the SW and HW components used by each application and the vehicles in which the application is integrated. The last column assigns the test site where the tests for the specific applications are planned.

Table 1: Applications, HW components, SW components, vehicles and test sites.

Application	HW Components	SW Components	Vehicle(s), RSU	Test site/ Simulator
Speed limitation and safety distance	Vehicles: CRF gateway VGX SBC, Main PC, IBEO LS (red), Radar (both cars), Appl PC, VANET router, Pos PC, ESP PC RSU: Main PC, VTT camera, wireless sensors, VANET Router	Vehicles: ESPOSYTOR sw, IBEO LS SW, GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle RSU: VTT camera SW, Router SW, ESPOSYTOR sw (for RSU)	Fiat Bravo red, Fiat Bravo blue, RSU	IT, CRF test track
	Vehicles: CRF gateway VGX SBC, Main PC, IBEO LS (red), Radar (both cars), Appl PC, VANET router, Pos PC, ESP PC RSU: Main PC, VTT camera, wireless sensors, VANET Router Mobile RSU: Main PC, VANET router, Pos PC	Vehicles: ESPOSYTOR sw, IBEO LS SW, GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle RSU: VTT camera SW, Router SW, ESPOSYTOR sw (for RSU) Mobile RSU: SP1/SP2 bundle, Router SW	Fiat Bravo red, Fiat Bravo blue, RSU, mobile RSU	IT, Torino-Caselle Highway

Application	HW Components	SW Components	Vehicle(s), RSU	Test site/ Simulator
Road intersection safety	GW, Main PC, Pos PC, Router PC, VANET Router	GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle	(1a) Egovehicle (Renault), Accident Vehicle (Renault)	WE, Satory test track and Valladolid technology park
	GW, Main PC, Pos PC, Router PC, VANET Router	GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle	(1b) Egovehicle (Renault), Approaching Vehicle (Renault), Vehicle blocking view (Volvo truck?)	WE, Satory test track intersection
	GW, Main PC, Pos PC, Router PC, VANET Router	GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle	(1c) Egovehicle (Renault), Approaching Vehicle (Renault)	WE, Satory test track intersection
	GW, Main PC, Pos PC, Router PC, VANET Router	GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle RSU: SP2 bundle, Router SW	(1d) Egovehicle (Renault), Approaching Vehicle (Renault), RSU (traffic light)	WE, Satory test track intersection
	GW, Main PC, Pos PC, Router PC, VANET Router	GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle	(1f) Egovehicle (Renault), Emergency Vehicle (Renault)	WE, Satory test track and Valladolid technology park
Lane change manoeuvre	APX-Piaggio GW, Main PC, GPS module, VANET Router	GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle	Car, PTW	IT
	TNO GW, Main PC, Pos PC, Router PC	GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle	Citroen C4, VW Passat	NL, Helmond

Application	HW Components	SW Components	Vehicle(s), RSU	Test site/ Simulator
Safe overtaking	APX-Piaggio GW, Main PC, GPS module, VANET Router	GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle	Car, PTW, other car	IT
Head on collision warning	CRF gateway VGX SBC, Main PC, IBEO LS (red), Radar (both cars), Appl PC, VANET router, Pos PC, ESP PC	ESPOSYTOR sw, IBEO LS SW, GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle	Fiat Bravo red, Fiat Bravo blue	IT, CRF test track
	CRF gateway VGX SBC, Main PC, IBEO LS (red), Radar (both cars), Appl PC, VANET router, Pos PC, ESP PC	ESPOSYTOR sw, IBEO LS SW, GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle	Fiat Bravo red, Fiat Bravo blue, non-equipped vehicle	IT, CRF test track
Rear end collision	CRF gateway VGX SBC, Main PC, IBEO LS (red), Radar (both cars), Appl PC, VANET router, Pos PC, ESP PC	ESPOSYTOR sw, IBEO LS SW, GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle	Fiat Bravo red, Fiat Bravo blue	IT, CRF test track
Frontal collision warning	Volvo GW, VANET Router, Pos PC, Aglia camera, Main PC, Appl PC	GW SW, VANET SW, Pos SW, SP1 Bundle, Appl SW	Renault truck, Volvo car	WE, La Valbonne test track and driving simulator
Road condition status	Volvo GW, VANET Router, Pos PC, Road eye (1st gen), Main PC, Appl PC	GW SW, VANET SW, Pos SW, SP1 Bundle, Appl SW	Volvo Truck 1, Volvo Truck 2, RSU	SWE
Curve warning	CRF gateway VGX SBC, Main PC, Appl PC, VANET router, Pos PC, ESP PC, RSU Appl PC	ESPOSYTOR sw, GW SW, VANET SW, Pos SW, Appl SW, SP1 bundle	Fiat Bravo red, Fiat Bravo blue, RSU	IT, CRF test track
Vulnerable road user accident avoidance	IBEO LS, Volvo GW, VANET Router, Pos PC, Main PC, Appl PC	IBEO LS SW, GW SW, VANET SW, Pos SW, SP1 bundle, Appl SW	Volvo Truck 1, Volvo Truck 2	SWE, Driving Simulator

3. Evaluation methodology

The evaluation of SP4 components and applications within WP4.6 follows the ISO 9001 procedures [13] and the system engineering V-model development process [14] which is common in the automotive industry and software development (Figure 3), adopted also by other integrated projects like PREVENT [15], CONVERGE [16] and APROSYS [17]. For SP4 a dedicated V model has been designed within WT 4.5 and WT 4.6 which contains the Verification and Validation approach of SP4 applications.

3.1. V-model approach

The V-model is described in this deliverable as source of all evaluation activity plans within SP4.

The left-hand branch in Figure 3 indicates the increasingly detailed system specifications as the system approaches the design stage. The right-hand branch shows the evaluation stages which become increasingly general towards the top right. The evaluation stage is then divided hierarchically in verification and validation stages: the verification is based on evaluation of the applications components on the base of technical specifications, while the validation phase is an evaluation of entire applications based on functional aspects.

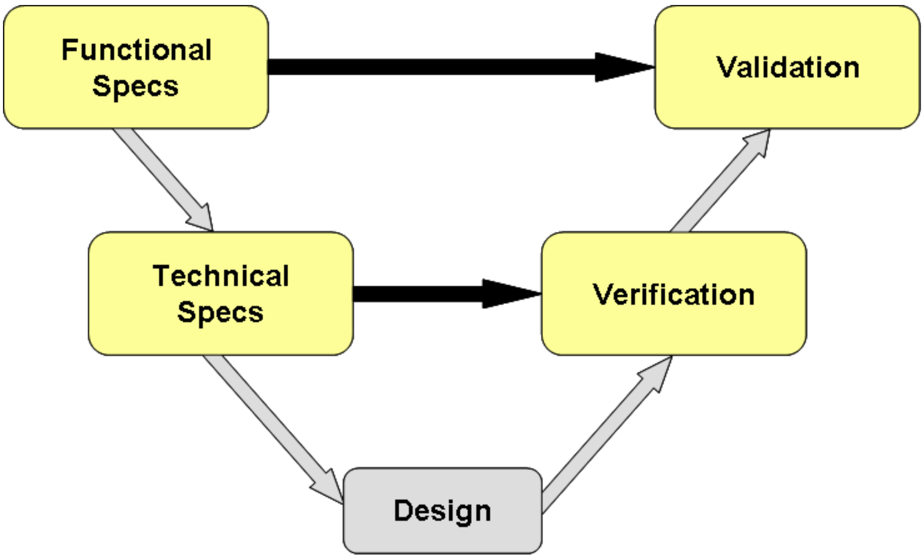


Figure 3: Simplified V-model design cycle

Within SP4 WP4.6, the evaluation focuses on the Safety Margin Assistant concept, which is a higher level of functional assessment. The lower level of technical assessment is the principal subject of the WP4.5, which is focused on the testing of application components singularly and integrated, and on the testing of integration of different applications. In general, the evaluation can be performed only after defining a set of performance criteria and their indicators for each element of the system and for the total system, and after defining a series of tests designed to assess whether the system meets these criteria.

This approach was followed to synthesize the SP4 pilot plan test case template where the HURR are stated as success criteria. In the template, appropriate indicative measures and success levels are specified to accomplish the HURR and hence the V-model methodology was applied to SP4 applications as specified more precisely in the next chapter.

3.2. V-model implementation in SCOVA

Figure 4 (next page) shows the V-model for SCOVA applications, including the two evaluation steps and the input specification documents that were used to derive the evaluation specifications.

Starting at the top left part of the V-Form, the initial definition of the Safety Margin Assistant (SMA) concept has contributed to the description of some specific use cases for the vehicle to vehicle applications. Both the SMA and the use cases were collected in two preliminary deliverables, D4.2.2 – Safety Margin Concept [18], and D4.2.3 – Use Case and Typical Accident Situations [19]. Refining even more, i.e. going further down on the left 'leg' of the V-Form, the requirements were defined from the use cases definition (see D4.2.4 - User Needs and Requirements [3]).

The specifications resulting from the requirements collection were then reported in several deliverables: D4.3.1 - Safety Margin Assistant Parameter Analysis and Characterization [5], D4.3.2 - Applications Functional Specifications [20], D4.3.3 - Application Communication for Co-operative Vehicles and Infrastructure [6], and D4.3.4 - Conceptualisation of On-board Information System and Extended HMI [21].

The implementation phase follows the specifications previously defined, and it is initially tested in the development environment. At the same time, the platforms for the vehicles are prepared by SP1, and the technologies developed by SP3, are released for the subsequent testing phase.

From that point on, the integration of the system could start (i.e. going up the right 'leg' of the V-Form), which was ended by the integration testing of WP4.5. Both the components and the integrated system thus have the same basic requirements and hence these requirements are tested both in WP4.5 and WP4.6 (as represented by means of the two bars on the right side of the figure). The most part of the requirements will be tested in WP4.5; some others – specifically the ones that are meaningful only in the TS environment because of a higher level character - will be carried out on test sites, and then fall in the responsibilities of the WP4.6.

The actual testing is explained in the horizontal branches of the diagram. All WP4.6 evaluations are on the same level, but of a different nature (safety impact, human factors, technical or functional). However, each one has to define a scenario, a method for evaluation (hypothesis, indicators, etc.) and a test plan. The elements of these branches are explained in more detail in the next chapter.

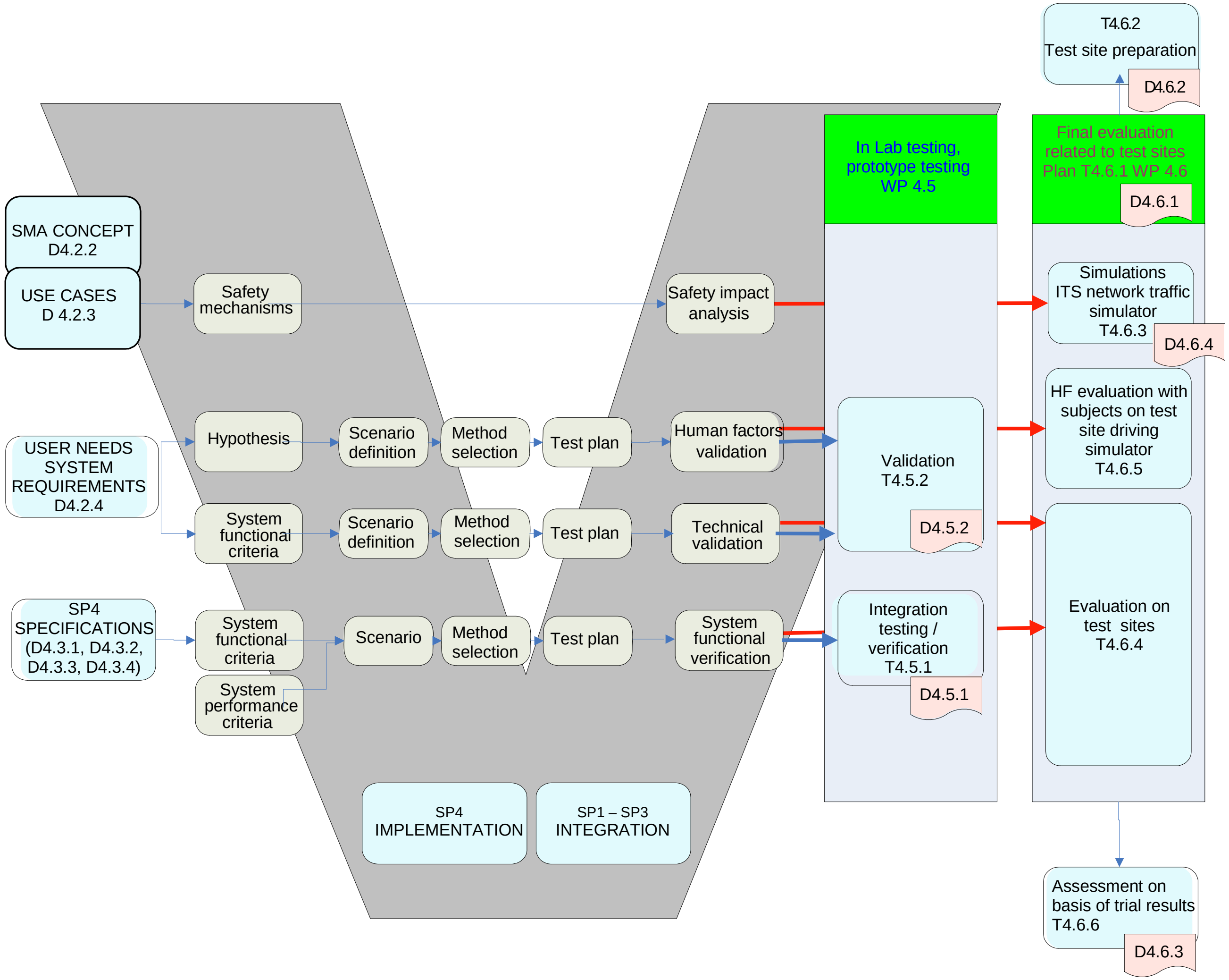


Figure 4: The V-model for SP4.

4. Evaluation

This chapter describes the single steps for evaluation of the SMA in detail (the horizontal branches in the V-model). According to the V-model, evaluation is started by collecting the SAFESPOT high level objectives, user needs, requirements and risks (referred to as HURR) which can be found in chapter 1.2. These represent the SAFESPOT SP4 success criteria. Consequently chapter 4 provides precise recommendations on how to measure such success criteria and indicators to be commonly used in all SP4 pilots. The evaluation chapter is divided into the following three categories which are addressed in SP4 WP6: technical evaluation, human factors evaluation and safety impact evaluation.

4.1. Technical evaluation

The technical evaluation according to the SP4 V-model consists of four elements: system functional criteria, method selection, scenario definition and test plan. While scenario definition and test plan are subject to chapter 5, the present chapter describes the system functional criteria in which indicators play an important role (chapter 4.1.1) and the different methods are described in chapter 4.1.2. It is recommended to all pilot test leaders to keep a strict compliance to the guidance given in the present chapter.

4.1.1. Technical indicators

The technical objectives of the system should be described in such a way that it allows the evaluation of the system. This is done by defining indicators (also called “criteria of response”, “measures of effectiveness”, “variables”, “assessment criteria”, etc.), which allow assessing if and to what extend the objectives are achieved.

The indicators should be defined in order to enable the measurement of the effectiveness of the system. To achieve this, these indicators should allow measuring **the nominal behaviour** of the system; in other terms these must be able to reflect clearly the related behaviour of the system (for instance what is the amount of speed decrease in a given test). In addition, the indicators should allow measuring **the reliability** of the system (for instance what is the reproducibility of the observed speed decrease).

Indicators for SAFESPOT SP4 evaluation are classified in 3 different types: Performance, Reliability and Correctness:

- **Performance:** The time the system needs to perform a task. Performance might be simply reported: i.e. “The system needs from A → B 1.3 seconds”, OR – whenever it is possible to define a maximum time – in terms of: “The system performs faster than X”. While Reliability and Correctness are basic indicators for the function of the system, Performance is especially important to measure the SMA – by following the underlying question: Do the warnings arrive in time for comfort, safety and critical situations?
- **Reliability:** A System is reliable if it performs always in the same way under the same conditions. It must not fail in unexpected situations or environments. A system is not reliable if the outcome of the same input varies.
- **Correctness:** A System performs correctly if the output is as expected. An example of correctness is the following: when headway becomes shorter than 10 meters at speed of 50km/h, then a **critical** Frontal Collision Warning is sent to the driver. Note that a system might be reliable but not correct; an example: the headway is shorter than 10 meters at 50km/h and a driver receives a **comfort** Frontal Collision Warning in 10 repetitions. In this case the system is “reliably wrong/not correct”.

a. Performance related indicators

Performance indicators measure if the system is able to process a timely warning in order to take intervention actions (by the driver or by other systems). Following indicators for performance are proposed:

- Levels of warning (comfort, safety and critical) for all applications and the (longitudinal) acceleration required to avoid the critical situation. So, does the comfort warning result in a comfortable avoidance manoeuvre (e.g. braking), the safety warning in a bit more demanding avoidance manoeuvre and does the critical warning indeed require immediate and strong intervention?
- Lateral acceleration for lane change assistant: in this case the lateral acceleration is not a technical indicator (it is more an indicator for impact). It can also be used as a comfort indicator.
- Longitudinal deceleration for the longitudinal applications.
- Impact speed reduction for the longitudinal applications.

For any system whose aim is preventive road safety, the time factor is fundamental i.e. the system must be able to achieve all necessary tasks within precise time boundaries. In the case of SAFESPOT, the warning messages must reach drivers in time to allow them to take avoiding actions. Hence, from the validation point of view, it must be ensured that the time constraints (which in principle are defined in the specifications) are met.

In SAFESPOT, the number of modules involved (belonging to the platforms and to the applications) makes it difficult to identify which components are responsible for the latency if the system performance does not meet the requirements.

There are two particular factors that make the estimation complex:

- On the one hand, tasks performed at component level run sometimes sequentially and sometimes parallel and asynchronous.
- On the other hand, the latency time of some components is non-deterministic [1].

To estimate global latency time it will therefore be necessary to draw for each event the full data chain from data source to receipt of a warning message.

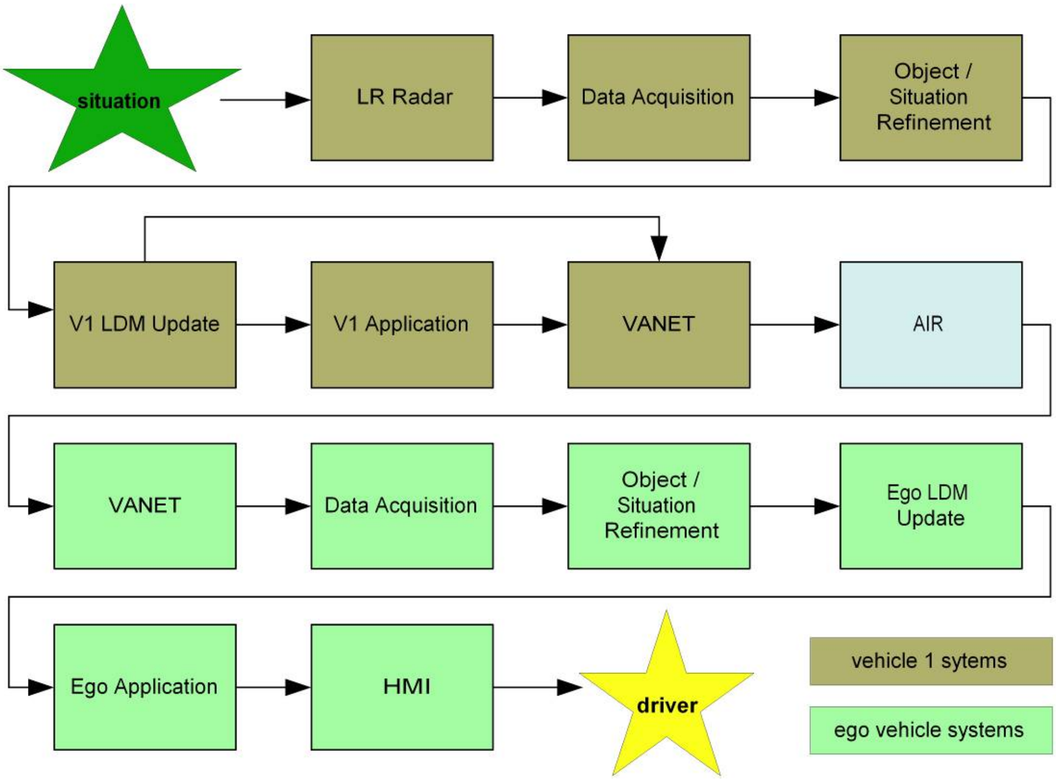


Figure 5: Sample data chain for a SP4 application

Figure 5 gives an example to illustrate the data chain. Each element in this chain requires a certain time to execute and/or transfer the aggregated data. For the SP4 evaluation the whole chain is of interest, from event occurrence till HMI activation or driver information, as this determines the speed at which a driver can react to the hazardous situation. Since the behaviour is of non-deterministic nature, multiple measurements of the latency of the whole system are advised and all measured results are requested to be reported.

[1] As an example, VANET introduces latency time very dependent on the situation especially if message relaying is active. Therefore, unless having perfectly repeatable scenarios, latency time will not be deterministic. Nevertheless, VANET is an entry component and then can not be overlooked since it provides all the information from other vehicles. However, it could be replaced with stimuli obtained by simulation.

b. Reliability related indicators

The reliability of a system can be defined as the probability of a component, sub-system of complete system to work correctly over a given period of time under a given set of operating conditions.

However, the reliability that the driver experiences in practice is associated with the capability of the system to provide reliable warnings. Thus system reliability is associated with a high number of true positives (TP, correct action when needed) and true negatives (TN, no action if none necessary) and a low number of false positives (FP, or false alarm) and false negatives (FN, or missed alarms).

TP, TN, FP and FN can be defined per km, per hour or per event. The relation between the actual situation and the prediction is shown in Table 2 and Table 3.

Table 2: Relation between actual and predicted data.

<table>
<tr><td></td><td></td><td colspan="2">Actual data</td></tr>
<tr><td></td><td></td><td>Negative (safe)</td><td>Positive (threat)</td></tr>
<tr><td rowspan="2">Prediction</td><td>Negative (safe)</td><td>a</td><td>c</td></tr>
<tr><td>Positive (threat)</td><td>b</td><td>d</td></tr>
</table>

Table 3: Definition of rates.

<table>
<tr><td colspan="2">Rate</td><td>Definition</td></tr>
<tr><td colspan="2">P (Precision)</td><td>$d/(b+d)$</td></tr>
<tr><td colspan="2">TP (True Positive rate)</td><td>$d(c+d)$</td></tr>
<tr><td colspan="2">FN (False Negative rate)</td><td>$c/(c+d)$</td></tr>
<tr><td colspan="2">TN (True negative rate)</td><td>$a/(a+b)$</td></tr>
<tr><td colspan="2">FP (False positive rate)</td><td>$b/(a+b)$</td></tr>
<tr><td rowspan="2">Reliability index</td><td>if positive and negative datasets have similar size</td><td>$(a+d)/(a+b+c+d)$</td></tr>
<tr><td>otherwise</td><td>$\sqrt{(d^2/(b+d)(c+d))}$</td></tr>
</table>

In some testing the status of what are true and false positives is judged by the operator of the vehicle. It is therefore important to distinguish between false positives and nuisance alarms. An example: a false positive for Frontal Collision Warning application (FCW) would be the alarm going on while being completely alone on the road. A nuisance alarm is getting an FCW when not wishing for one, i.e. the function works fine but the driver does not think s/he needs the information/ warning/ intervention at that point in time. Which of these is tested needs to be specified in the description.

The safety of the system as perceived by the driver can be defined in terms of the ability of the system to adequately respond to a hazardous situation. This is expressed by the Correct Alarm Rate (CAR), False Alarm Rate (FAR) and Missed Alarm Rate (MAR). FAR is related to FP, MAR to FN. CAR, MAR and FAR can be expressed with respect to all times an alarm should be given, or with respect to all situations, and with respect to driven km, time, scenarios,..

Examples:

- for FCW, the missed alarm rate is calculated as the number of missed critical scenarios over the total amount of critical scenarios (FN), FAR is the number of wrongly detected critical situations related to a time or distance of normal driving.
- for Road Intersection Safety application (RIS) the rates are defined with respect to the number of correct or false decisions at intersections, compared to all situations.

The sum of CAR, MAR and FAR can be more than 100%, depending on the definition of CAR, MAR and FAR.

Since no lengthy reliability testing is possible due to the required time for that, the evaluators are requested to perform a reliability testing per event, and not per km or over time, and perform at least 10 runs per setup. The results report should include the number of correct outputs with respect to the number of repetitions of the test, and any false warnings and missed warnings (i.e. the CAR, MAR and FAR) with respect to the number of repetitions.

c. Correctness related indicators.

These indicators are related to the parameters used inside the decision system for the application activation, i.e. basically in the “Application Manager” component of the SP4 applications (D4.3.2 - SP4 Applications Functional Specifications [22]). They measure if the decision system works correctly in terms of “correct” decisions and “in-time” decisions.

In SP4 the concept of the Safety Margin was introduced initially in D4.2.2 – Safety Margin Concept [23]. For this only few basic indicators are required to evaluate the correct decision to activate an application:

- position of the vehicles / RSUs
- speeds of the vehicles
- trajectories of the vehicles

All other indicators can be derived from these ones. It is left to the testers also to include intermediate results. For the lateral applications (LCM and CUWA) logging of the lateral acceleration, lateral speed (‘lateral motion’), heading and the indicator light may be useful.

These indicators are also needed to measure the quality of the different sub steps that add to the decision making. Some examples are the following:

- Detection rate: the number of detected objects related to the number of existing objects.
- Map data errors: Map object data are the basis for a warning. Data errors can occur due to coding errors (wrong coding or missed data) and the change of reality in comparison to stored data, e. g., by road construction.

Reference measurements may be needed in order to calculate and evaluate the accuracy of the different indicators. Two kinds of reference measurements are needed:

- Spatial
- Temporal

A precise spatial reference measurement is needed for localization accuracy (e.g. using light barriers of microwave sensors, high precision GPS or combinations of sensors).

A precise temporal reference measurement is needed in order to evaluate the accuracy of a perception system. For this, different sensors can be integrated in the host vehicle allowing comparing the measurements provided by the perception system with reference values (i.e. for the time to impact estimation, obstacle distance, etc.). A ‘trigger’ signal should be implemented to be able to identify the start of a critical situation.

The reference measurements can i.e. be used to calculate the accuracy of the comfort, safety and critical warnings emitted by the SAFESPOT Safety Margin Assistant.

4.1.2. Technical testing methods

A variety of facilities and methods is available for carrying out the tests. Each has advantages and disadvantages. None of them is perfect but some are complementary to others. To carry out exhaustive tests often several different approaches are used.

a Virtual approaches

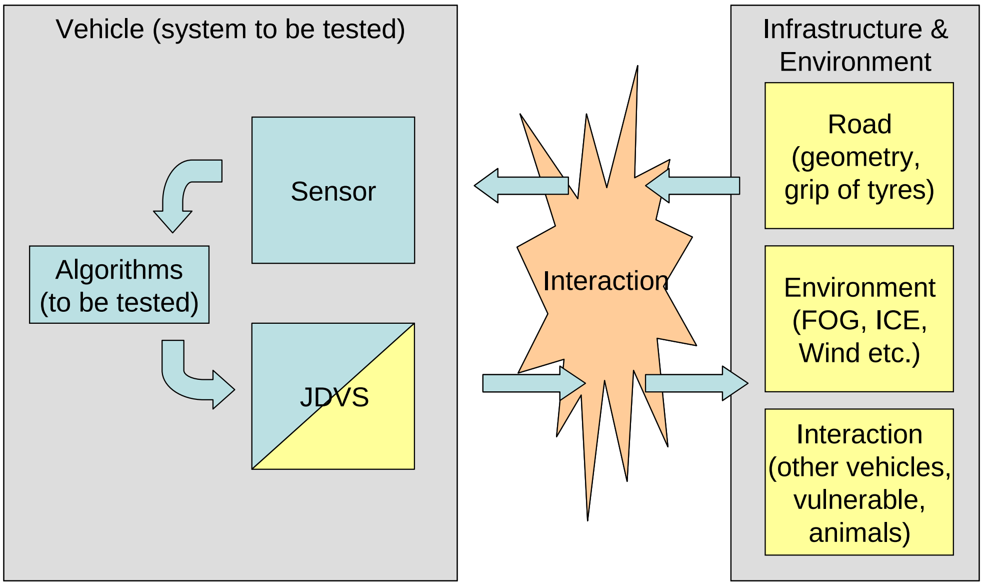
Virtual tests are employed to validate the software systems. They include modelling tools (typically, MATLAB, SIMULINK), driving simulators, and virtual R&D lab. All allow the execution of trials in absolute repeatable conditions. However, the link to reality may not be strong because the variety of situations is limited. To be representative, the tests require an accurate modelling of hardware components which they replace (sensors, actuators, vehicle dynamics etc) as well as environmental conditions (other actors, weather, etc.).

But an important advantage of the virtual tests is that it is not necessary to wait for the occurrence of a specific environmental condition to achieve the tests: then, validation of i.e. fog or ice sensing modules can be done at any time irrespective if is summer or winter time.

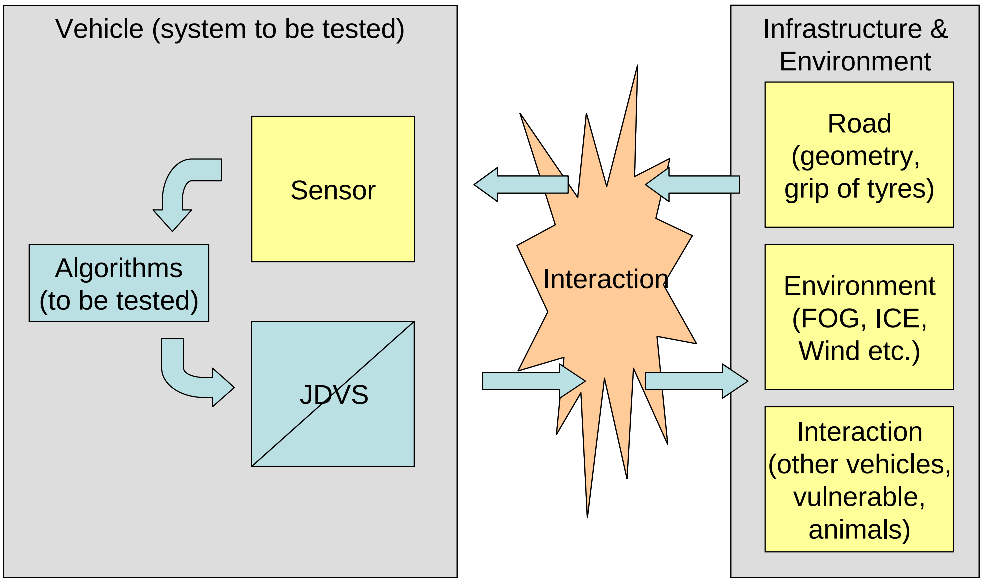
In the context of the SP4 evaluation, the MARS simulator, developed by TNO [24] will be used as a reference tool for the testing and validation, under simulated conditions, of the SP4 applications.

b Hardware in the loop

Midway between real and virtual testing is hardware in the loop (HIL) testing. Figure 6 shows the HIL architecture. HIL tests have meaning if there is a component chain including a feedback provided by actuators, the driver himself or a combination of both (Joint Driver-Vehicle System - JDVS).



6a) HIL with real sensors and virtual actuators



6b) HIL with virtual sensors and real actuators

Figure 6: HIL for lane change application.

HIL tests offer a good compromise between control of the test conditions and their representativeness: indeed, in HIL, environmental conditions are simulated therefore perfectly repetitive while sensors and actuators are real, thus avoiding errors due to bias in sensors and actuators models.

It is worth to note that the MARS simulator can be used in order to implement HIL tests. Real data collected during the tests or in some dedicated acquisition session can be fed in to the simulator, implementing so called “mixed environment” scenarios. In these scenarios, all “elementary parameters”, such as vehicle positions, VANET exchange of information among the actors, sensors and vehicle data can be based on real data, previously logged and imported in the simulator. All other higher level activities, including the functions of the LDM and the cooperative applications, should be based on real test-bench hardware, connected to the simulator, or completely replaced by the SW scenarios supported by MARS.

c Real testing

Real testing helps to validate both the hardware and software systems. This includes test beds, test tracks in a protected environment and in normal traffic on public roads. It provides the opportunity to experiment with a great variety of situations. However, environmental conditions are difficult to control, impairing repeatability of the tests.

In order to solve the most part of the issues related to the real testing, a possibility exists, and will be used by some SP4 partners, to replicate the real tests in a laboratory environment by using dedicated data loggers and data players.

These tools (special purpose SW) can be used – for instance – to record some real scenarios during its evolution in a real Test Site environment and to store all the relevant information collected in a suitable digital format. In the SAFESPOT architecture this can be easily achieved since the most part of the data exchanged locally inside each vehicle is standardised and based on a well defined UDP protocol [25].

In this way whatever information flowing on the Ethernet LAN can be replicated, by “playing back” the recorded UDP packets in a laboratory test bench. Of course, in the laboratory environment, the HW running the LDM and the SP4 applications (with the SP1 and SP4 frameworks) should be physically duplicated, since the functions of these components, including the database queries and transformations, the HMI activities, etc., are not based on UDP.

4.2. Human Factors evaluation

Working with human beings is totally different from working with a technical system. When conducting evaluation studies with test-participants special countermeasures have to be taken in order to avoid biasing the results by systematic influences, i.e. order effects, experimenter effects or non-standardized environments. It is also important to be aware of the fact that the result of tests with human beings can be influenced by the pre-defined hypothesis, an adequately sized and selected sample of participants, the test–design and selection of dependent and independent variables, the measuring tools and the results analysis and interpretation. Since instruction of how to avoid common mistakes and how to conduct methodologically correct experiments is available in abundance, the reader is referred to very good literature instead of repeating these rules once more. Especially two deliverables from European projects can be considered very useful for the SP4 evaluation of human factors: AWAKE_D7.1-Pilot Plan [26] and PReVENT Annex E to D16.4 [27]. Other interesting references on the subject are [28], [29], [30] and [31].

4.2.1. Human Factors indicators

The different SP4 applications will be evaluated with normal users in simulators and on real roads. Since applications, test sites, test objectives and tested HURR can require measuring different indicators with different testing methods, the present pilot plan deliverable gives some general recommendations and leaves much of the specific test design and testing methods to the single test responsible person. However in order to evaluate the applications with respect to human factors, it is proposed to consider at least the following common indicators in the experiments:

- Usability
- Acceptance
- Workload

For those indicators standardized measuring methods are proposed in Chapter 4.2.2.

Depending on the application to be evaluated and the selected HURR it is strongly recommended to further define evaluation indicators that contribute to meaningful results, such as:

- Reaction time of the drivers
- A behavioural description of the reactions shown
- Deceleration

Valuable hints for the definition of further indicators can be extracted from the HURR. The indicators should contribute to measure if the HURR are met with the SP4 applications. Many indicators can be easily measured with a structured interview as described in the next chapter.

The selection of participants and test drivers for the SP4 pilots is dependent on the selected indicators and the hypothesis underlying the experiment. Test participants might be of three different cohorts:

- Experts on driving assistance systems evaluation
- Experts on human factors evaluation
- Normal users and potential clients with characteristically little pre-knowledge about the kind of system tested

Also for on road tests different requirements might apply than for the simulator tests. To the pilot leader it is recommended to define (and report) a list of criteria for selecting the participants for the experiments, containing for example:

- Age
- Gender
- Minimum amount of driven km last year
- Experience with driving assistance systems evaluation
- Experience with human factors evaluation
- Experience with SAFESPOT SMA (similar) systems
- Interest in new driver assistance systems
- Interest of buying advanced drivers assistant systems

Further it is important to know the precise profile of each participant containing all information that might become interesting in the results interpretation. It is recommended to ask/record at least all the data mentioned in Table 4 for every participant in the SP4 pilot tests.

During the tests several not intended factors might influence the results. It is important to record those events in order to interpret the results correctly. Any kind of events that are mentioned in Table 5 should be noted.

Table 4: Questions about participants profile.

Question	Answer
Age:	
Gender:	
Km driven in the last year:	
Years of car driving experience:	
Brand, model, year of the car(s) driven in most of the time?"	Brand: Model: Year:
Experiences with the system of any kind before the test started (i.e. note if staff of the same company or same campus/location is participating):	
Driving style applied at most times during THIS test session:	sportive – spontaneous ☐ efficient – determined ☐ anticipatory – comfortable ☐
Time of day when the test was conducted:	
How was the person contacted:	
How was the person informed about the tests and the system:	
What was the precise instruction received before the test:	Refer to Annex for the written instruction.

Table 5: Not intended influences during the pilot.

Category	Note
Function performance of the system during the test (note what happened and when in terms of non-compliance or any other unplanned behaviour of the system).	
Any environmental influences, i.e. rain, noise, time pressure.	
Any personal influences, i.e. tiredness, sickness, negative attitude, etc.	

4.2.2. Human Factors testing methods

For SAFESPOT only two kinds of testing methods for human factors tests are applicable: tests in driving simulators and on road tests with real vehicles. Generally speaking simulator tests allow a much better control of influencing factors; on road tests are much closer to reality. In SAFESPOT also safety considerations will influence the decision which tests are carried out in the simulator and which are realized in a real environment.

In both cases it is possible to use questionnaires for the measurement of indicators. The recommended questionnaires for the main human factors indicators, usability, acceptance and workload are presented in this chapter. The tools are presented in English, hence all participants need good English skills; otherwise a translation is needed for the language of the current test site. Additional questions are recommended for specific aspects of the SMA evaluation. The questionnaires can be found also in the annex for easy printout. They might also be updated, translated or extended as necessary for the single experiments requirements.

The following section contains a set of recommended questionnaires and measuring methods for a standardized measurement of the main human factors indicators. It is recommended to use the selected questionnaires and methods and to add further questions and measuring methods where applicable in order to give answers to the testing hypothesis.

a. Testing usability

In order to test the usability of the SAFESPOT applications and HMI it is recommended to use a standardized questionnaire. Usability tests should be seen also as a valid source for optimization of the product in following development cycles.

Definition of concept

Usability is defined as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO 9241-11, 1998) [32].

The terms effectiveness, efficiency and satisfaction are defined as follows:

- Effectiveness: Accuracy and completeness with which users achieve specified goals
- Efficiency: resources expended in relation to the accuracy and completeness with which users achieve goals.
- Satisfaction: Freedom from discomfort and positive attitudes towards the use of the product.

Proposed tool

It is proposed to assess usability with the System Usability Scale (SUS) [33], which provides a reliable, low-cost tool that can be used for global assessments of systems’ usability. For further reading [33], [34], [35] and [36] is recommended. The SUS is applied after a user has used a system, but before any discussion and debriefing. Subjects are asked to respond immediately, rather than thinking for long. Table 6 presents the System Usability Scale.

Table 6: System Usability Scale (SUS).

	Strongly disagree 1	2	3	4	Strongly agree 5
1. I think that I would like to use this system frequently					
2. I found the system unnecessarily complex					
3. I thought the system was easy to use					
4. I think that I would need the support of a technical person to be able to use this system					
5. I found the various functions in this system were well integrated					
6. I thought there was too much inconsistency in this system					
7. I would imagine that most people would learn to use this system very quickly					
8. I found the system very cumbersome to use					
9. I felt very confident using the system					
10. I needed to learn a lot of things before I could get going with this system					

Reporting results

The results are valid for the functions of a system that could be tested, so the exact interaction of the user should be specified in the reporting chapter.

In order to calculate the SUS score, following steps are necessary:

- Odd numbers in scale position (1, 3, 5, 7, 9): Score value minus 1 (e.g. Item 3 gets score 2; 2-1=1).
- Even number in scale position (2, 4, 6, 8, 10): 5 minus score value (e.g. Item 10 gets score 4: 5-4=1.
- The sum of the scores is multiplied by 2.5.
- The final figure derived represents the usability score between 0 (very low usability) and 100 (very high usability).

The SUS score provides a standardized value for comparison of different products or of the same product tested by different user groups or in different development stages. The SUS score will enable the SAFESPOT consortium to assess the general usability of each application, detect applications with low or high usability, enhance usability aspects if necessary and measure the effect by testing again with the SUS and finally compare the usability results of SAFESPOT applications with the results of applications developed in other or future projects. Heuristically it can be assumed that SUS scores above 70 are ok, while values below 60 should lead to a further optimization circle of the system.

b. Testing acceptance

The user acceptance is a crucial requirement for any new system in a vehicle. Since the SMA will interact with the user it is important to evaluate the user acceptance towards this system.

Definition of concept

Acceptance as defined for the User Acceptance Scale [37] is a concept based on the perception on usefulness and satisfaction.

Proposed tool

Acceptance can be assessed very well by a simple question that should be asked to all participants after experiencing a system.

- “Would you keep the system on/active all the time or inactivate it in certain situations?”

If the answer is “I would inactivate it…” the following question should be asked additionally:

- “When?” and “Why?”

Additionally the User Acceptance Scale [37] is proposed. Subjects are instructed to tick a box on each of the nine scales of the following questionnaire indicating the extent to which the stated attributes are applicable with respect to the system under evaluation (Table 7).

Table 7: User Acceptance Scale.

My judgements of the system are ... (tick one box in every line)

1	Useful	☐	☐	☐	☐	☐	Useless
2	Pleasant	☐	☐	☐	☐	☐	Unpleasant
3	Bad	☐	☐	☐	☐	☐	Good
4	Nice	☐	☐	☐	☐	☐	Annoying
5	Effective	☐	☐	☐	☐	☐	Superfluous
6	Irritating	☐	☐	☐	☐	☐	Likeable
7	Assisting	☐	☐	☐	☐	☐	Worthless
8	Undesirable	☐	☐	☐	☐	☐	Desirable
9	Raising alertness	☐	☐	☐	☐	☐	Sleep-inducing

Procedural guidance for user acceptance scale

Using the User Acceptance scale is easy:

- The test leader should describe the system to be evaluated in terms of 'what is your judgement about a system that would… (short & clear explanation of the system functioning) and present the nine items (before-measurement).
- After experiences with the system under evaluation the nine items are presented again: What is your judgement about the system …(name), you just finished driving with (after-measurement).
- The results of those two judgements will be compared.

Reporting results

The results are reported with a specific description of the interaction of user and system within the study. The User Acceptance Scale results are calculated as follows:

- Individual items should be coded from -2 to +2 from left to right, scores on items 3, 6, and 8 should be coded ranging from +2 to -2 (these items are mirrored).
- Reliability analysis should be performed between the before-measurement and after-measurement per item and per subject (use of Cronbach's “alpha”[2] is suggested). If reliability is sufficiently high (above 0.65), c the end scores are computed per subject for the two scales by averaging the scores on the uneven items 1, 3, 5, 7, and 9 for the usefulness score, and averaging scores on the even items 2, 4, 6, and 8 for the satisfying score.
- The usefulness scores can now be averaged over subjects to obtain an overall system practical evaluation. The same can be done with the satisfying scores.
- Difference-scores should be calculated per subject by subtracting the before-measurement score from the after-measurement score per scale. The difference scores show whether and in which direction subjects' opinion was altered as a result of experience with the system.

c. Testing workload

A safety system should not increase the mental workload for drivers. It is proposed that for this reason the SMA is also evaluated regarding the mental workload.

Definition of concept

Definition of workload is difficult; however mental workload remains an important and practically relevant concept [35]. Workload can be assessed by using the NASA TLX [38] and should be seen, according to this scale, as a composition of 6 sub concepts:

- Mental Demand
- Physical Demand
- Temporal Demand
- Performance
- Effort
- Frustration Level

[2] *Cronbach's α is a reliability indicator and defined as*

$$\alpha = \frac{N}{N-1}\left(1 - \frac{\sum_{i=1}^{N}\sigma^2_{Y_i}}{\sigma^2_X}\right)$$

where N is the number of items, σ^2_X is the variance of the observed total test scores, and $\sigma^2_{Y_i}$ is the variance of item i.

Proposed tool

The NASA TLX is proposed to measure the workload of drivers. It is a widely spread technique and can be applied to many domains (Table 8).

Table 8: NASA TLX.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Procedural guidance for NASA TLX

The NASA TLX is an easy to use tool. A manual is also available for free [38]. An "overview in few dots" is reported hereafter:

- The participants should get a short introduction on the questionnaire in order to understand the sub-scales meaning and to be able to answer quickly.
- The NASA TLX can be used during or after the test trial. It is however recommended for SAFESPOT to use it after the test since it can be intrusive to the task.
- It is proposed to use the NASA TLX in test trials with the SAFESPOT system activated and also in trials with the system deactivated in an equal situation. The results of the two trials for workload can be compared.
- After the test trial participants are asked first to weight the sub scales regarding most effect on their mental workload. This is done by presenting pair wise comparisons between the sub scales using the paper/pencil version for this comparison in Annex 9.2. The NASA TLX manual [38] gives also specific instruction on how to perform this task.
- Then participants are asked to give a subjective rating for each sub scale between 0 (low) and 20 (high).

Reporting results

The calculation of the NASA TLX mental workload scale is done as follows:

- First the score of every sub scale is multiplied with the weighting (e.g. 0x11=0 or 6x15=90).
- Second all weighted results of all sub scales are added.
- Third the results are divided by 15 – a mental workload score between 0 and 100 is the result.

The results of the NASA TLX scores should be compared for the trials with the system versus the trials without the system. The mental workload should be approximately the same or lower when the SAFESPOT system is activated.

d. Questions for SMA evaluation

Additionally to the standardized scales it is recommended to conduct an open structured interview after the scales. An open structured interview is a situation in which the test leader and the participant talk about the experiences of the test participant during the experiment. The test leader asks a number of pre formulated questions and the participant can answer verbally. In case of doubts or especially interesting topics a discussion between test leader and test participants might develop and lead to further interesting results and statements. The test leader however should take care that he/she gets answers in the dedicated time to all of his/her questions. It is recommended to categorize answers for the reporting. Single statements might be reported as representative examples or interesting opinions. In the Interview, questions to the categories pointed out in Table 9 are recommended:

Table 9: Question categories for interview after pilot.

Question categories:
Did the comfort/safety/critical warning arrive in time for you to react comfortably/safely/critically?
Did the warning distract from the event? Why? / How?
Did the warning help to manage the event? Why? / How?
Did you understand the warning? What did you understand? To what event the warning was in your opinion related?
What did you appreciate most about the system?
What did you appreciate the least of the system?
How can the SMA (warnings) be optimized in your opinion?
How is your behaviour influenced by having the SMA system activated?
Do you have further comments about the presentation of information (HMI)?

In order to develop specific or further questions, the HURR should be considered in order to address those success criteria specifically.

4.3. Safety impact evaluation

Safety impact evaluation does not concentrate on the new application itself, but on the environment it is acting in, which is the traffic. This is different from both technical and human factors evaluation and gives an impression on the usefulness of the system from another point of view, addressing different HURR. The goal of the safety impact evaluation is to identify and predict the social benefits of the system in terms of reduced number of fatalities, injuries and accidents. It is necessary to identify these benefits to assist policy makers in making a proper consideration between the costs of investment against the potential benefits of the system.

4.3.1. Safety impact indicators

Unlike for technical testing, safety measures such as the number of crashes, crash rates, etc. are not directly measurable for the new applications. There are, however, so called surrogate safety measures that are measurable or can be estimated: the safe headway distance (and predicted crashes) [39], time to collision, speed variance and the efficiency coefficient [40]. The generation of these indicators is discussed in next chapter for example by means of simulation modelling.

4.3.2. Safety impact testing methods

Safety impact analysis can be done in different ways. One is to describe an approach using the so called behavioural effect approach [40]. This does not only cover the direct intended effect of applications, but also the indirect and unintended effects of the systems, including behavioural adaptation in long term use.

The focus of the (safety) impact analysis is not as much on the application and its intended use, but on higher level like how many accidents can be avoided and how the throughput of a road is influenced. Many aspects influence the outcome of an impact analysis, like “how does the user react to the application?”, “what does the application do normally?”, “what can be (unwanted) side effects?”, and so on. Real life testing is not possible for safety impact analysis, since many equipped vehicles are required and also long term evaluation is needed. This requires both a large investment and a long period of time. This undermines the goal of the safety impact analysis, which is to identify the safety impacts to make a proper consideration for investment costs. Besides all that, accidents cannot be invoked, if only for legal reasons.

Hence other ways to predict the impact are developed. One of these methods [40] is based on identifying which impact mechanisms are relevant for the new application (i.e. in/direct modification of driver behaviour, modification of interaction between users and non-users, etc.). The expected effects have to be motivated by references to literature or other evidence, like i.e. human factors evaluation. Then a percentage is estimated based on knowledge of the relevant accident type the application influences using accident databases and taking background variables, e.g. weather conditions, road type, vehicle type, etc., into account. Then, with the total number of (all) accidents and the proportion of accidents in each variable category based on the accident data set in question, the formulas give the impact of the application on safety.

Another approach to estimate the impact is via traffic simulation of the relevant situation (for which the application was designed). Dedicated traffic simulation programs exist (like TNO’s ITS-Modeller [41]) and the functionality of the application can be programmed into these. Many different situations can be simulated to encompass the effect of different parameters like penetration rate, driver behaviour to the application, traffic density, etc. Hence, estimates of different scenarios can be made. Besides that also different parameters are gained like traffic throughput, time to collision (indirect measure for safety), speed variance, fuel consumption, and number of messages of the V2V system.

Each of the mentioned methods has its pro's and cons. Depending on the application and how much the researcher is familiar with the approach a method is chosen.

5. Pilot plans

In order to decide which pilots shall demonstrate the technical function, the human factors and the impact of the SP4 applications, a test case template for pilot planning was developed within WP4.6. The SP4 application leaders and the test site leaders were asked to define the right use cases and to plan the single pilot test cases by using this template with the aim of demonstrating the SMA for each application. The usage of the pilot plan test case template ensures also the connection of the pilot tests to the SAFESPOT SP4 objectives (HURR).

The following sub-chapters explain the purpose of the developed test case template and contain a filled example of this template in order to facilitate the understanding of the procedure for pilot test case planning that has been followed in SP4.

Furthermore, the presentation of a table summarizing the tests that have been planned within SP4 for a given reference application enables the reader to get a clear overview of the relevant aspects to consider in the design and planning of the pilot tests. The current filled in templates, detailing the pilot tests planned so far for each single test case, for all SP4 applications and TSs, can be found in the Annex to the present deliverable.

5.1. Pilot plan test case template

A common pilot test planning has been defined within the work of WP4.6. All pilots are planned in a unified way by using a SP4 specific template. The template has been developed for SAFESPOT SP4 by following the ISO 9001 [13] procedures and the system engineering V-model approach as described in Chapter 3. Consequently it focuses on testing the HURR that have been defined early in the project. This approach was selected in order to achieve on the one hand a unified test planning assuring the quality of the test plan for each experiment, and on the other hand this approach aims and assures that each experiment is related to the SAFESPOT SP4 success criteria (HURR).

The compilation of the pilot plan test case template was the first outcome of WT4.6. The template was then used to compile and collect all planned tests for each SAFESPOT application. In a first step the application leaders were then asked to report their planned tests in the template – or to plan the tests they have been considering relevant - for proving the HURR for their specific application. After collecting a significant number of filled in templates (planned pilot test cases) the WT4.6 and D4.6.1 responsible partners revised the level of detail in the templates and small updates were carried out.

The achieved quality and level of detail enables the test sites and application leaders now to foresee the planned pilots and to prepare test tracks, vehicles and RSU accordingly. Currently the templates are revised for this reason by test site leaders. Test site leaders are also asked to add some relevant pilots if necessary from their point of view. Because of this running process, it is expected that the current versions of templates might be updated or some more tests might be planned during the period until the beginning of tests.

In order to introduce the template in this deliverable a filled in example is shown in Table 10.

The pilot test template can be divided into two main sections: the header section contains general information by which the different tests can easily be grouped or distinguished. Among this general information there are the applications to be tested, the test purpose regarding the employed indicators, the test site or simulator used and the use case the experiment refers to.

The second section of the template, test setup and scenario, aims at relating the pilot tests to the HURR. The pilot responsible is first asked to describe the experiment precisely, by filling in the specific section of the template.

The example showed in Table 10 refers to a filled-in template for application #5 (Rear End Collision) to be tested. The precise test setup and scenario description refers to the manoeuvres to be carried out and the variables to be manipulated which are the time to collision, the distance between the two vehicles, the relative speed and the lateral position of both vehicles (in relation to each other).

Following to this test description the pilot responsible states in four columns how the experiment contributes to the evaluation of each HURR. The template refers explicitly to the SP4 requirements (limited to those requirements that are explicitly higher objectives and that can only be tested with a fully integrated and operating vehicle on a test track – all other requirements are subject to SP4 WP5) and user needs. Since each application has its own requirements and user needs, the test leader is asked to check in the relevant deliverable which requirements and which user needs can be addressed with the pilot and to pick only those that can be evaluated in the conducted test. In the example template no requirements are evaluated but three user needs are extracted from the list of application #5 user needs.

By continuing with the columns to the right, it is then necessary to define how these HURR can be tested – this process is called “operationalization” and it is achieved by filling the columns to the right of each HURR. The columns help the test leader to define how he / she will test the specific HURR in the given test. The first of the four columns states the specific HURR, in the second column it is defined how the HURR is considered in the test and in the third column the test leader is asked to define the measurement proposed in order to measure the HURR. Here it needs to be defined what is the best indicator / what are the best indicators in order to measure if i.e. the distance to the preceding vehicle is known. As an example we might look at the first user need (‘SP4_UN_05_04) which specifies an interest of the driver for information about vehicles approaching from behind. In the second column the test leader states that this user need is taken into account because the scenario simulates a typical situation for a rear end collision. In order to deliver this information, it is mandatory to know several indicators which are stated in the third column like the vehicles ID, trajectories, distance and time to collision between vehicles. It is also stated in column three what are the thresholds to be achieved for the main indicator time to collision. In column four the test leader states the tool needed for the measurement. In the example this is data logging and a comparison of the input data (time) with the moment of warning. Hence the example aims for evaluating if emitted warnings arrive in a specified time at the vehicles HMI as an output.

The four columns are not applied for the high level objectives since due to the high level of success description a single experiment cannot prove a high level objective. Still the test leader states how the experiment can be related to the high level objectives; this will help to evaluate the application with respect to the high level objectives after receiving results from all pilots. The example experiment refers to the first high level objective by testing a situation that is only achievable with cooperative systems and hence contributes to the objective by demonstrating that warnings can be delivered at a moment totally impossible to achieve with nowadays ADAS.

The Risks formulated in SP6 are also topic of testing and have therefore a part in the template. Test leaders are asked to check the list of risks (D6.4.5 – Preliminary Recommendations [4]) and to state which risks can be dealt with in the planned pilot test.

Summarizing, the most important function of the pilot plan test case template is to achieve a common test planning and to ensure that the pilots contribute to the pre defined HURR. Deliverable D4.6.3 – Results Evaluation [42] will report how the SP4 pilot tests comply with the SAFESPOT Safety Margin Assistant concept by proving the pre defined HURR.

The empty template of the test cases for the pilot plan [43], together with the tutorial [44] explaining how to fill it in are available on the common BSCW area, at the following link:

http://bscw.safespot-eu.org/bscw/bscw.cgi/193084

As already stated, following example shows a filled in template (Table 10), selected in order to present and discuss the structure and the usage of the SP4 pilot plan test case template.

Table 10: The pilot plan test case template with example for Rear End Collision application.

SAFESPOT COOPERATIVE SYSTEMS FOR ROAD SAFETY	**SAFESPOT SP4 WP6 TEST CASE Definition for Evaluation**	

Task [1]	Task 4.6.4	Compiled by (Author) / Company:	CRF / G. Vivo	Date:	21/11/2008	Sheet No.: [2]	RECO_IT_01
Application(s) tested:	#5 - Rear End Collision	Vehicles / RSU: [3]	1. Fiat Bravo Red 2. Fiat Bravo Blue	HW - components to be used in test [3]	1. SP1 platform including: CRF gateway VGX SBC, Main PC, Laser Scanner and its own PC on Fiat Bravo Red, Radar on both Fiat Bravo Cars; 2. Application PC 3. Router PC 4. Positioning PC 5. ESPOSYTOR PC.	SW-components / modules to be used in the test [3]	1. Positioning SW 2. Laser scanner SW 3. RECO application 4. Router SW 5. ESPOSYTOR tool.

Test Type (multiple possible) [4]		Test Purpose (multiple possible) [5]		Test Environment	
Human Factors	☐	**Usability**	☐	**Traffic Simulation**	☐
Technical Evaluation	☒	**Acceptance**	☐	**Driving Simulation**	☐
Safety and traffic evaluation	☐	**Performance**	☐	**Test Site West**	☐
Other:		**Reliability**	☒	**Test Site Italy**	☒
		Correctness	☒	**Test Site Germany**	☐
		Other:		**Test Site Sweden**	☐
				Test Site NL	☐
				Specific area at test site:	**CRF Test Track**

Which SF use case(s) do you refer to (derive from D4.2.3 for your application)? Try to fuse different use cases to one test case scenario if possible. [6]	Application #5 - Rear End Collision: Use Case 5, General Case
Are multiple applications evaluated simultaneously?	☒ No, ☐ Yes

Test setup and scenario [7]
Vehicle A and B drive on the CRF Test Track. The test track will be prepared in order to test the manoeuvre in safe conditions; since the vehicle drivers will be professionals and the dynamics character is not as demanding as it is in other applications, these conditions should be easily achievable. The vehicles will drive in the same lane and in the same directions, starting from an initial distance of more than 200 meters. The speeds of each vehicle, in the 10-20 runs of the test, will be mostly constant for both vehicles. The leading vehicle - ego vehicle, and primary actor of the use case - will drive at a low speed (for instance ~40 Km/h); the following vehicle - secondary actor - will approach at a higher speed (for instance ~40 Km/h). The trajectories of the two vehicles will be kept in the head-to-tail collision path until the time to collision value drops to a minimum around 6 or 7 sec. (beginning of the Safety Area). Then the secondary actor will decrease its speed, in order to perform a queuing manoeuvre, which will reach a steady condition with the two vehicles aligned at 10 m of distance each other, both running at ~40 Km/h. The vehicles will be driven by expert technical drivers that know the system. No subjects (=non expert drivers unfamiliar with the system) will be used. The following variations will be applied in a total of 10-20 runs: -The timing when the second vehicle starts its queuing manoeuvre can be progressively reduced, up to 4 sec., in order to log the time when the warning is issued. The behaviour of the system will be assessed in the no-warning area, in the comfort warning area and in the safety warning area (the critical warning area will be tested with the test case #2); -The relative speed between the vehicle will be varied between 20 to 60 km/h - Acting on the direction indicators and driving more or less close to the lane borders (but still keeping the head-to-tail collision trajectory) to check for false alarms All data will be logged and evaluated afterwards.

Success Criteria [8] :			
Requirements (D4.2.4) for this application that are vital for WP4.6 (note: state only REQ that cannot be tested sufficiently in WP4.5):	How considered in the test case?	State what you are going to measure (units) and if possible define a threshold to prove that this success criteria is met.	How are you going to measure this? Define your measuring tools.
Note: for the Rear End Collision application the success criteria are defined based on the full coverage of the stated Users Needs and the fulfilment of the HLO. All vital requirements not reported here for the application are addressing the components and the basic technologies adopted for the application deployment; these aspects are fully covered within WP 4.5.			

User Needs from D4.2.4:	How considered in the test case?	State what you are going to measure (units) and if possible define a threshold to prove that this success criteria is met.	How are you going to measure this? Define your measuring tools.
#SP4_UN_05_01 - Driver of vehicle 1 want to be informed about the dynamic information (including relative speed, acceleration, direction indicators, lateral position) of a second vehicle, approaching from behind at a speed significantly higher speed respect to own vehicle.	The presence of a second vehicle which is approaching from behind (becoming a danger source in case its speed is significantly higher than the one of the EV) is considered in the computation of the warning signal going to the HMI.	The presence of a second vehicle will be assessed by means of: Other Vehicle ID, Analysis of the EV and OV trajectories (probability of intersection of the trajectories >= 80%), distance between EV and OV along the road (m.), time to collision between the two vehicles (in sec.), acting on the direction indicator performed by the OV. When the evaluated probability is higher than the stated threshold, TTC values below 12 sec. should lead to comfort warnings, TTC values below 8 sec should lead to safety warnings.	All of the (time stamped) data for the tested manoeuvres will be logged. The correctness of the measures will be provided by means of comparison tables where input values (TTC values) are put in relationship with the time the expected warning is provided.
#SP4_UN_05_04 - Vehicle drivers want to receive some safety driving recommendations regarding the presence of black spots or road scenarios potentially leading to the rear end collisions.	The hazardous situation consisting in the presence of a second vehicle on a head-to-tail collision trajectory in the same lane is considered in the computation of the warning signal going to the HMI.	The hazardous situation will be assessed by considering: Other Vehicle ID, Analysis of the EV and OV trajectories (probability of intersection of the trajectories >= 80%), distance between EV and OV along the road (m.), time to collision between the two vehicles (in sec.). When the above probability is higher than the stated threshold, TTC values below 12 sec. should lead to comfort warnings, TTC values below 8 sec should lead to safety warnings.	All of the (time stamped) data for the tested manoeuvres will be logged. The correctness of the measures will be provided by means of comparison tables where input values (TTC values) are put in relationship with the time the expected warning is provided.
#SP4_UN_05_09 - Driver interface should timely reflect the persistence or the changes related to the dangerous situation.	The timely intervention of the driver HMI is considered by precisely measuring and logging the absolute time (millisec. precision) of HMI activation.	Both the absolute times of HMI activation and deactivation for two Safety Margin areas (Comfort and Safety) will be measured (and logged) with the millisec. accuracy.	The analysis of the logged information will be performed off line and compared to predefined TTC thresholds. These TTC thresholds (12 s. and 8 s.) defines the transitions in the Comfort Area and in the Safety Area.

High level objectives (HLO) [9]	Is this HLO (partly = a,b,c) considered in the test case?	Describe how your test can be related to this HLO? How can your measurements be related to his HLO?
a) The main aim of the project will be to show the feasibility and benefits of Co-operative Systems solutions b) in improving road safety c) well beyond the level which can be achieved with autonomous solutions (vehicle or infrastructure based).	☐ No, ☒ Yes	The test is directly related with this HLO, especially concerning point c). For the transition in the comfort area, the foreseen activation time, at 12 sec. of TTC, places the application into a domain totally outside of the capabilities for a classical ADAS approach. Also in the other areas (Safety and Critical) the cooperative approach places the OV in the "detection area" - the LDM of the EV, by means of the VANET communication, leading to simpler, safer and more robust implementation and deployment of the tested application.
To demonstrate benefits for accident types with a calculated potential safety impact in terms of a) saving of lifes as well as other gains such as: b) c) d)	☐ No, ☒ Yes	The rear end collisions (head-to-tail between vehicles) are acknowledged among the most frequent accident situations, especially in urban areas and at intersections. Even thought the test case is not specific for these scenarios, the diffusion of the Rear End Collision Warning application will have direct (positive) consequences in terms of life saving.
To improve the a) range b) quality and c) reliability of the safety-related information.	☐ No, ☒ Yes	The reliability test is performed since any test condition is repeated a number of times that is sufficient for a prototype to demonstrate the right behaviour.
a) To support drivers preventively to the proper manoeuvres in the different contexts; b) To optimise the intervention of vehicle controls with respect to critical situations.	☐ No, ☒ Yes	The test is explicitly addressing the support to drivers by preventing the head-to-tail collisions.
To manage existing incidents to minimise further negative safety impact.	☒ No, ☐ Yes	
To increase safety for all road users in a specific situation.	☒ No, ☐ Yes	
To show that the safety impact can be achieved without affecting transport efficiency.	☒ No, ☐ Yes	

To increase the Safety Margin of vehicles using in-vehicle and infrastructure information.	☐ No, ☒ Yes	This HLO can be tested, for instance, by showing that the Safety Margin (in the transition to the comfort area, ~12 sec. before the time when TTC=0) is extended at a level where - possibly - the EV and OV are not even in the direct line of sight. This will demonstrate that a (proper) warning signal can be issued when there is no evidence of any danger for the EV driver, with an extension of the Safety Margin beyond his/her perception capabilities.
To create applications for extended cooperative awareness by means of real time reconstruction of the driving context and environment.	☐ No, ☒ Yes	This HLO is just partially covered. Even though the extended cooperative awareness in not the specific focus of the application, the Rear End Collision Warning (as it is for all other SP4-SCOVA applications) is based on the real time reconstruction of the driving context and environment around the EV, which is the straight concept of the LDM database.
To open the development of new safety applications based on a cooperative approach.	☐ No, ☒ Yes	The overall test is designed and devoted to show the direct benefits of the cooperative approach, by means of a novel class of applications (V2V communication).

Any risks of SP6 that are covered? Refer to the WP4.6 risks compilation with 25 risks. [11]	How considered in the test case?	State what you are going to measure (units) and if possible define a threshold to prove that this success criteria is met.	How are you going to measure this? Define your measuring tools.
Risk3 -The SMA supplies too much information and also too many warnings to the driver.	The HMI of the application is provided with an interface that decides the warning to be sent, based on a priority of applications, and a period of changing warning, that should avoid the continuing flikering among multiple warning.	The output of the applications tests is logged, together with the output going to the driver, to show how many warning are produced and how many are actually sent to the drivers.	The measure is a comparison between inputs of the HMI from application side, and outputs to the driver.

Extra success criteria that are worth testing? [10]	How considered in the test case?	State what you are going to measure (units) and if possible define a threshold to prove that this success criteria is met.	How are you going to measure this? Define your measuring tools.
Note: for the Rear End Collision application no extra success criteria are defined, since the compliance with the Users Needs and with all the stated HLO is judged as largely sufficient for the proper assessment of these criteria.			
Add further lines here:			

Non-Compliance Reporting
To fill in after performing test
Obtained values / results
To fill in after performing test

5.2. Pilot plan test cases planned in SP4

This section summarises the test cases planned in SP4 until submission of this deliverable. The full version of pilot test cases that are planned by using the pilot test case template is available in a dedicated Annex document. It can be also downloaded at any time and in the up to date version from the BSCW server at the following link:

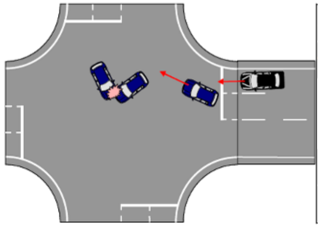
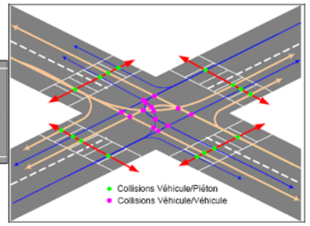
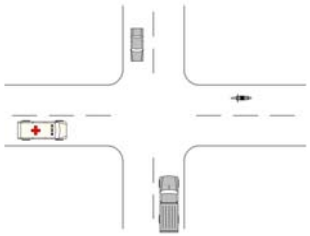
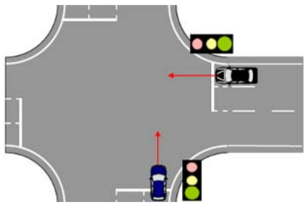
http://bscw.safespot-eu.org/bscw/bscw.cgi/188030

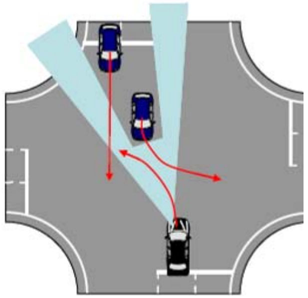
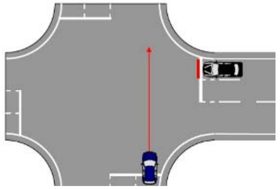
24 experiments have been planned so far and hence 24 templates are available in the Annex and for download. The pilot test cases have been defined by the respective application leaders in cooperation with the test site leaders and under guidance by WP4.6 members. The content of the templates is a first version of a detailed test case plan and might be further refined by application and test site leaders.

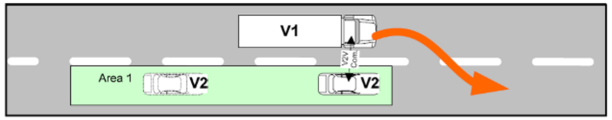
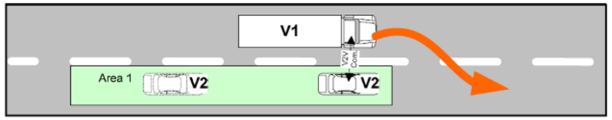
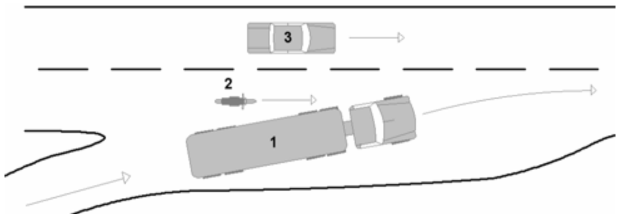
Having the pilot plan test case templates available on the BSCW area enables all partners to keep the work up to date ad facilitates the on-line consultation. This updating procedure might become necessary since the process of pilots planning can not be considered as complete in all details until having test sites and vehicles totally prepared and verified the function of all components and systems. Experience shows also that it is almost unrealistic to plan experiments in all details a long time on beforehand in such a complex environment like SAFESPOT. However an already very good level of detail for SP4 pilot test cases is already achieved at the current time.

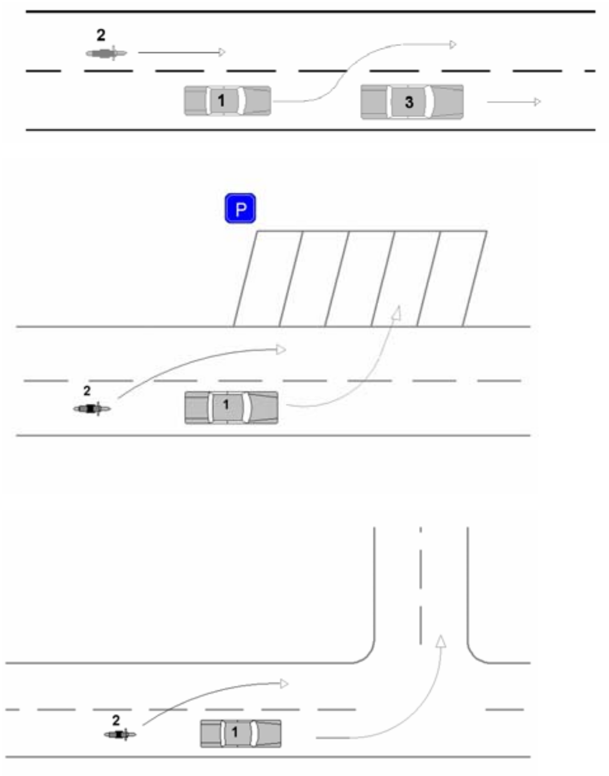
Table 11 shows a summary of the planned 24 tests. The table contains the relevant ground information about each test case planned and it is sorted by applications and test site. The use case that is referred to in the pilot was selected by the pilot leader for best representing his / her intent of demonstrating that the SAFESPOT SP4 HURR are met. An overview on test types and planned test purposes is already present in the column on right hand of the table.

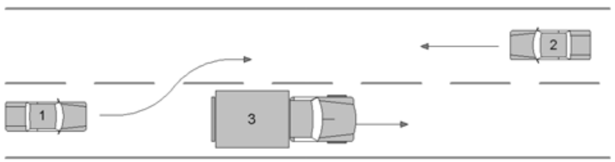
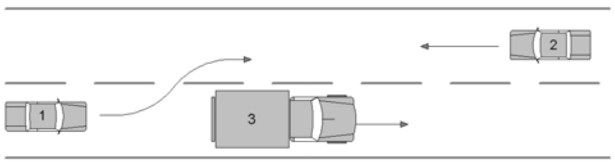
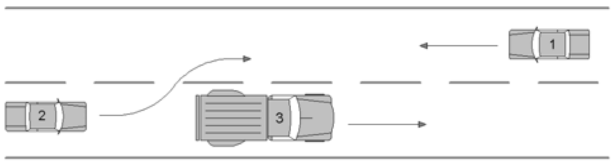
Table 11: Pilot test cases planned for SP4.

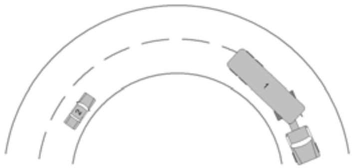
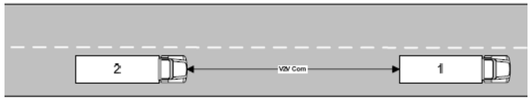
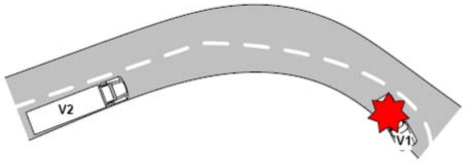
Application	**Test Site**	**Use Case**	**Reference Template / Test Type / Test Purpose**
1 Road Intersection Safety - Accident at intersections	Test Site West Satory tracks Valladolid technology park	Road Intersection Safety - Accident at intersections - UC1a  	SP4_WP6_TestCase_RIS_1A_CAS Technical Evaluation • Performance • Reliability • Correctness
1 Road Intersection Safety - Approaching emergency vehicle warning	Test Site West Satory tracks Valladolid technology park	Road Intersection Safety - Approaching emergency vehicle warning - UC 1f 	SP4_WP6_TestCase_RIS_1F_CAS Technical Evaluation • Performance • Reliability • Correctness
1 Road Intersection Safety - Defect traffic signs	Test Site West Satory tracks	Road Intersection Safety - Defect traffic signs - UC 1d 	SP4_WP6_TestCase_RIS_1D_CAS Technical Evaluation • Performance • Reliability • Correctness

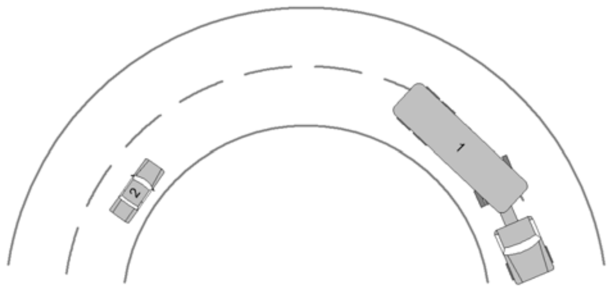
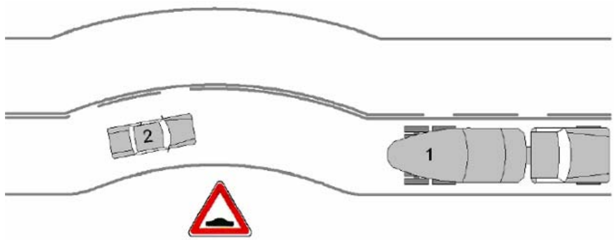
Application	Test Site	Use Case	Reference Template / Test Type / Test Purpose
1 Road Intersection Safety - Obstructed view at intersections	Test Site West Satory tracks	Road Intersection Safety - Obstructed view at intersections - UC 1b 	SP4_WP6_TestCase_RIS_1B_CAS Technical Evaluation • Performance • Reliability • Correctness
1 Road Intersection Safety - Permission denial to go-ahead	Test Site West Satory tracks	Road Intersection Safety - Permission denial to go-ahead - UC 1C 	SP4_WP6_TestCase_RIS_1C_CAS Technical Evaluation • Performance • Reliability • Correctness

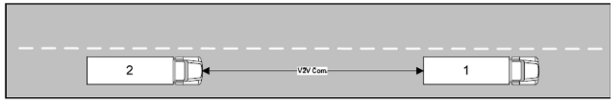
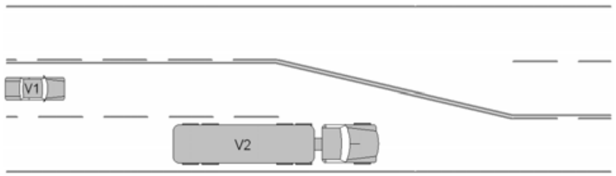
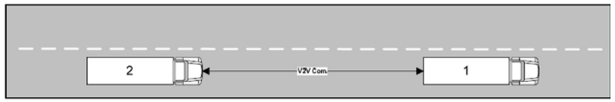
Application	Test Site	Use Case	Reference Template / Test Type / Test Purpose
2 Lane Change Manoeuvre	Test Site NL Helmond	Lane Change manoeuvre for a car with blind spots - 2a 	SP4_WP6_TestCase_LCM_2A_NL_1 Human Factors • Usability • Acceptance • Workload
2 Lane Change Manoeuvre	Test Site NL Helmond, public road	lane change manoeuvre for trucks with blind spots - 2a 	SP4_WP6_TestCase_LCM_2A_NL_2 Technical Evaluation • Reliability • Correctness
2 Lane Change Manoeuvre	Test Site Italy	Lane Change Manoeuvre General Case UC2GC 	SP4_WP6_TestCase_LCM_IT_01 Technical Evaluation • Correctness

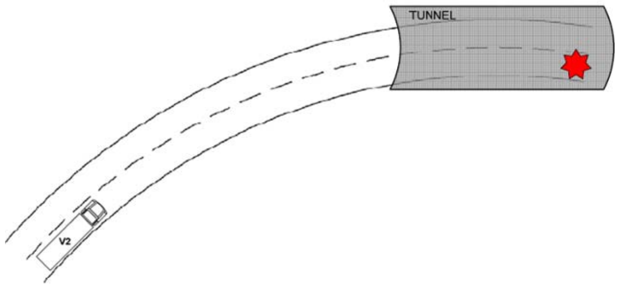
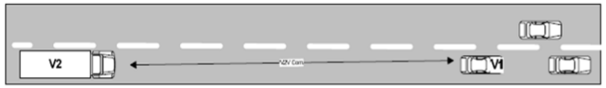
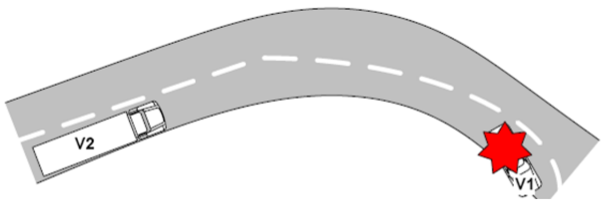
Application	**Test Site**	**Use Case**	**Reference Template / Test Type / Test Purpose**
3 Safe Overtaking	Test Site Italy	Safe Overtaking General Case UC3GC 	<u>SP4_WP6_TestCase_SO_IT_01</u> Technical Evaluation • Correctness

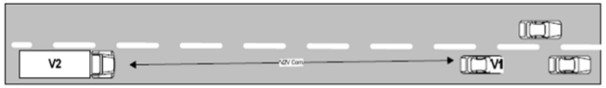
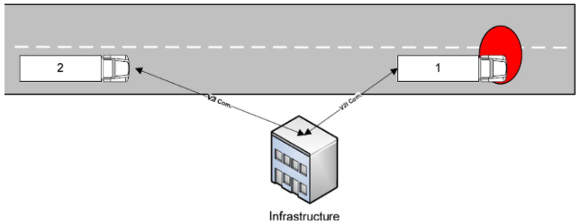
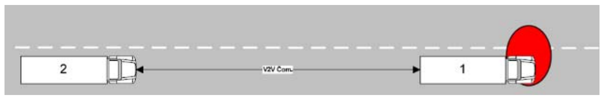
Application	Test Site	Use Case	Reference Template / Test Type / Test Purpose
4 Head On Collision Warning	Test Site Italy CRF Test Track	Head On Collision Warning: Use Case 4, General Case 	SP4_WP6_TestCase_HOCW_IT_01 Technical Evaluation • Reliability • Correctness
4 Head On Collision Warning	Test Site Italy CRF Test Track	Head On Collision Warning: Use Cases 4A  Head On Collision Warning: Use Cases 4B 	SP4_WP6_TestCase_HOCW_IT_02 Technical Evaluation • Reliability • Correctness

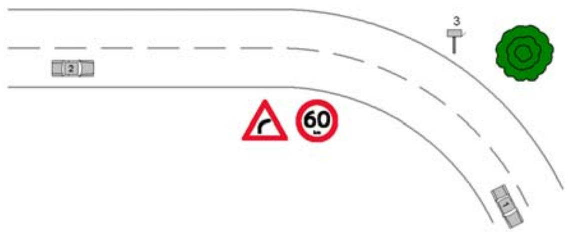
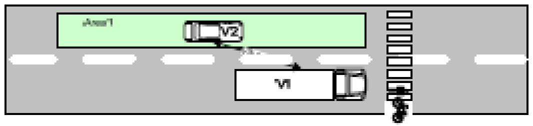
Application	Test Site	Use Case	Reference Template / Test Type / Test Purpose
4 HOCW 6 SLSD 8 RCS 5 RECO	Traffic Simulation	Rear End Collision – 5a  SP4_UC_SpeedAndDistance – 6a  SP4_UC_FrontalCollisionWarning – 7a  SP4_UC_FrontalCollisionWarning – 7c  SP4_UC_RoadConditionStatusV2V – 8b 	SP4_WP6_TestCase_RECO-SLSD-FCW-RCS_TS_TNO Safety and traffic Evaluation • Performance

Application	Test Site	Use Case	Reference Template / Test Type / Test Purpose
5 Rear End Collision	Test Site Italy CRF Test Track	Rear End Collision: Use Case 5, General Case 	SP4_WP6_TestCase_RECO_IT_01_v1 Technical Evaluation • Reliability • Correctness
5 Rear End Collision	Test Site Italy CRF Test Track	Rear End Collision: Use Case 5, General Case 	SP4_WP6_TestCase_RECO_IT_02 Technical Evaluation • Reliability • Correctness

Application	**Test Site**	**Use Case**	**Reference Template / Test Type / Test Purpose**
6 Speed Limitation Safety Distance	Test Site Italy CRF Test Track	Speed Limitation and Safety Distance between vehicles: Use Case 6, General Case 	SP4_WP6_TestCase_SLSD_IT_01 Technical Evaluation • Reliability • Correctness
6 Speed Limitation Safety Distance	Test Site Italy Torino - Caselle highway	Speed Limitation and Safety Distance between vehicles: Use Case 6c : lane restriction 	SP4_WP6_TestCase_SLSD_IT_02 Technical Evaluation • Reliability • Correctness
6 Speed Limitation Safety Distance	Test Site WEST La Valbonne	Speed Limitation and Safety Distance between vehicles: Use Case 6a, Speed limitation and safety distance 	SP4_WP6_TestCase_SLSD_WE_VOLVO Technical Evaluation • Reliability • Correctness

Application	Test Site	Use Case	Reference Template / Test Type / Test Purpose
7 Frontal Collision Warning	Test Site Sweden	Frontal collision warning due to static obstacle in a tunnel – 7b 	SP4_WP6_TestCase_FCW_SW_01 Human Factors and Technical Evaluation • Acceptance • Performance • Reliability • Correctness
7 Frontal Collision Warning	Test Site West La Valbonne	Frontal collision warning due to abnormal vehicle behaviour in front – 7c 	SP4_WP6_TestCase_FCW_WE_01 Human Factors and Technical Evaluation • Acceptance • Performance • Reliability • Correctness
7 Frontal Collision Warning	Test Site West La Valbonne	Frontal collision warning due to static obstacle in front – 7a 	SP4_WP6_TestCase_FCW_WE_02 Human Factors and Technical Evaluation • Acceptance • Performance • Reliability • Correctness

Application	Test Site	Use Case	Reference Template / Test Type / Test Purpose
7 Frontal Collision Warning	Driving Simulator Stuttgart (not Test Site)	Frontal collision warning due to abnormal vehicle behaviour in front – 7c  	SP4_WP6_TestCase_FCW_DS_USTUTT Human Factors • Acceptance • Usability • Workload
8 Road Condition Status	Test Site Sweden	Road Condition Status warning due to slippery area ahead with information from infrastructure and detection of slippery area from sensor equipped truck -UC8a  	SP4_WP6_TestCase_RCS_SWE_01 Human Factors and Technical Evaluation • Acceptance • Performance • Reliability • Correctness
8 Road Condition Status	Test Site Sweden	Road Condition Status warning due to slippery area ahead with information from vehicle to vehicle communication and detection of slippery area from sensor equipped truck -UC8b.  	SP4_WP6_TestCase_RCS_SWE_02 Human Factors and Technical Evaluation • Acceptance • Performance • Reliability

Application	Test Site	Use Case	Reference Template / Test Type / Test Purpose
9 Curve Warning	Test Site Italy CRF Test Track	Curve Warning: Use Case 9, General Case 	SP4_WP6_TestCase_CUWA_IT_01 Technical Evaluation • Reliability • Correctness
10 Vulnerable Road User Accident Avoidance	Test Site Sweden	Vulnerable Road User Accident Avoidance: Use Case 10a 	SP4_WP6_TestCase_VRUAA_SWE_01 Human Factors and Technical Evaluation • Acceptance • Usability • Workload • Correctness • Performance • Reliability

6. Standardized results reporting

After conducting the SP4 pilots the results will be reported in D4.6.3 – Results Evaluation [42]. Since many pilots are planned, it is important to standardize the results reporting. In this chapter some recommendations are given on how to report each pilot in order to gain a clear overview on the results of all pilots and to facilitate the reporting and interpretation. Furthermore it is suggested to document the pilots with photos and videos, as described in chapter 6.2. A SAFESPOT non compliance management tracking procedure is especially important for the SP4 pilot tests and is hence explained in sub-chapter 6.3.

6.1. Results reporting

The reporting of each pilot will be executed in an internal report that should consist of the following headings and information. It is recommended to have not more than 10 pages for each pilot and to follow strictly the standard reporting, since the internal reports will be clustered and annexed to the results reporting document D4.6.3 - Results Evaluation [42]. In reporting also the advice given in the respective chapters for technical, human factors and impact evaluation should be considered.

1. **Summary / Abstract**
2. **Introduction**
    - Give a short description of the application, test site and reference scenario (Scenario from D4.2.3 - Use case and typical accident situation [19]).
    - Refer to the information given in the pilot plan test case template header.
    - Give a short overview which HURR are referred to.
3. **Method**
    - Hypothesis / Aim of the experiment / Measured indicators.
    - Test site description.
    - Test scenario description (including photos or sketches and reference to SAFESPOT scenarios).
    - Technical Setup (SW/HW/Cars/RSU/…).
    - Setting: Weather, Temperature, Driver, ….
    - Description of the procedure followed during the test (e.g. report exemplary test trail).
    - Participants / Drivers profile.
    - Applied tools to obtain the results.
    - Data processing or statistical analysis methods applied.

4. Results

- Report the results in tables and bar-diagrams.
- Take the recommendations given in the respective sub-chapter of Chapter 4 – Evaluation into account.
- Report only relevant findings, but report all outcomes that will influence your conclusions.

5. Interpretation of results

- Interpretation of results (Reliability of results / How did [technical] problems [might have] influenced the results).
- Summarize the results that are in your opinion relevant and reliable for SAFESPOT SP4.

6. Discussion / Conclusion

- Which HURR are addressed in the test and what are the impact/conclusions that can be drawn from the results with respect to this/these HURR?

7. References

Give the reference to literature or SAFESPOT documents used for your report.

8. Annexes

- Questionnaires.
- Written Instruction given to participants.
- Photos + Videos.
- Time plots of most important signals / indicators.
- Tables.
- Statistical calculations.
- Test checklists.

6.2. Video Recording and Photos

This chapter describes the video recording procedure to be used in the SAFESPOT pilots. All pilots in SAFESPOT should be documented in order to provide material for the dissemination activities and for reporting special or illustrative scenes afterwards with a short video of the scene. A continuous analysis of behaviour for test purposes by videos is not recommended, this has been proven to be extremely time consuming.

Both video recording and pictures describing the experimental setting should be collected. For pictures it is important to have high visual quality and also some photos which are nice (take light, background, other objects, perspective and composition into account). Furthermore, the actual experiment should be video recorded so that interesting and illustrative results can later be shown. For reasons of general coherence and in order to integrate videos from different pilots to the final SAFESPOT video, it is needed that all partners use Hdv format 720p, Mini dv or beta, whenever possible.

The scenes should describe the scenario tested by combining scenes from the inside of the vehicle (drivers face and HMI usage) and of the outside while driving showing the cars and also showing scenes of the street and infrastructure and surrounding traffic. For SAFESPOT it will be especially important to visualize the collaboration between the vehicles and the infrastructure.

Each partner should at the end of the pilot produce a 5 minutes video clip in Hdv format 720p, Mini dv or beta standard, presenting the cooperative approach and the most significant methodological points, scenarios, set up and results at the pilot with optional English comments. These videos will be sent to ICCS to synthesize the final SAFESPOT DVD. A written documentation of the video recording, with information as partner name, pilot id, how the video was made, should be attached in order to support the DVD production.

It is very important to ask the participants and all people that appear in the videos and photos in a written document for approval to use the video and photos both for research and dissemination purposes.

The video and photo material should be either uploaded to the BSCW server or to be sent to the Dissemination Manager. In the SAFESPOT IP this role is in charge to ICCS.

6.3. Non compliance reporting

Problems during the testing activities shall be reported in a common Non Compliance Management Tracking Form [45] in order to have the possibility to manage the collection/submission/tracking and fixing of the problems.

The Non Compliance Management Tracking Form is transversal to all technical Sub Projects from SP1 to SP5. It is an excel document, composed by a sheet for each technical Sub Project.

It must be used any time that there is non compliance behaviour in a testing activity involving components provided by different sources, whose integration caused a fault.

Who individuated the fault (the owner) must recognize which Sub Project is involved in the fault, and fill in a new row in the correspondent sheet of the Non Compliance Management Tracking Form. Then he must set the status to OPEN and upload the document on the BSCW area, by sending at the same time the Non Compliance Management Tracking Form to the Sub Project involved responsible (the addressee).

If more than one Sub Project is involved in the detected fault, the owner must send an email to the reference persons of all the involved Sub Projects. The more involved Sub Project must coordinate the non compliance fixing phase and set the status of the row to ACKNOWLEDGE.

When the addressee solves the problem (or coordinates the solutions of more involved Sub Projects), he sets the status of the row to FIXED. He uploads the document on the BSCW, and sends back an email to the owner.

The owner must verify the solution of the problem and set the status to CLOSED.

The SAFESPOT Core Group will be in charge of the supervisory of the application of the Non Compliance process.

The entire process of downloading and uploading the document, the filling of each field of the new row for the non compliance detected, is explained in a dedicated tutorial [46].

7. Conclusions

This deliverable describes the methodology to adopt for implementing all SAFESPOT SP4 pilot test cases. Its aim is to standardize the conduction of the pilot tests and to define common measuring indicators which shall be used during their deployment.

The compliance with the described methodology will also provide, in the following steps of the project, the testing and the assessment of the SAFESPOT SP4 success criteria (HURR). For this reason a specific SP4 pilot plan test case template has been developed within WP4.6 with the explicit aim of relating the pilot tests to the HURR.

The deliverable contains an example table, for presentation purposes, of the pilots planned in SP4. Although changes might be necessary to the current pilot test cases planned, this table is almost complete and gives already a very good idea about the number, method and content of the planned tests. The full list of all the pilot tests currently planned for all SP4 applications is included as an annex document.

The deliverable will be followed by D4.6.2 – Pilot Plan Assessment which describes the test sites where all pilots will be conducted and also will provide more details about the actual environment and setting for the pilots.

After the conduction of the pilots, results will be reported in a common and consistent way. Indications for reporting are already given in the present document. Consequently the document D4.6.3 – Results Evaluation will contain the detailed report of the tests for all pilots. Additionally it will asses to what extent the SAFESPOT high level objectives, user needs, requirements and risks have been addressed in the final pilots.

The impact of a SAFESPOT Safety Margin Assistant on traffic is under investigation in the impact pilots and results will be reported in a special deliverable D4.6.4 – Impact Evaluation, underlining the importance of this kind of assessment to further investigation in the field.

8. References

No	Ref
[1]	SAFESPOT Technical Annex, SAFESPOT consortium, 2006.
[2]	SAFESPOT SP4 public page, http://www.safespot-eu.org/sp4.html
[3]	D4.2.4 - User Needs and Requirements.
[4]	D6.4.5 – Preliminary Recommendations
[5]	D4.3.1 - Safety Margin Application Parameters Analysis and Characterization
[6]	D4.3.3 - Application Communication for Co-operative Vehicles and Infrastructure
[7]	D4.4.2 – Equipped Cars Integrating the Safety Margin Application
[8]	D4.4.3 – Equipped Trucks Integrating the Safety Margin Application
[9]	D4.4.4 – Equipped Motorcycles Integrating the Safety Margin Application
[10]	D5.2.2 – Common Architecture and Communication
[11]	D8.4.3 – Common Validation Plan.
[12]	D4.6.2 – Pilot Plan Assessment
[13]	ISO 9001 - Section 7.3.1 Design and Development Planning, Section 7.3.5 Design and Development Verification, Section 7.3.6 Design and Development, Section 4.2.4 Control of Records.
[14]	V – Model - Systems Engineering – System life cycle processes, ISO/IEC 15288.
[15]	PREVENT - http://www.prevent-ip.org/
[16]	CONVERGE - http://cordis.europa.eu/telematics/tap_transport/research/projects/converge.html
[17]	APROSYS - http://www.aprosys.com
[18]	D4.2.2 – Safety Margin Concept
[19]	D4.2.3 – Use Case and Typical Accident Situations
[20]	D4.3.2 - Applications Functional Specifications
[21]	D4.3.4 - Conceptualisation of On-board Information System and Extended HMI
[22]	D4.3.2 - SP4 Applications Functional Specifications
[23]	D.4.2.2 – Safety Margin Concept.
[24]	Papp, Z.; den Ouden, F.; Netten, B.; Zoutendijk, A. (2006). MARS: Scalable HIL Simulator for Multi-Agent Systems Interacting in Physical Environments
[25]	User Datagram Protocol – RFC Standard 768
[26]	AWAKE_D.7.1-Pilot Plan (http://www.awake-eu.org/pdf/d7_1.pdf)
[27]	PReVENT Annex E to D16.4 (http://www.prevent-ip.org/en/prevent_subprojects/horizontal_activities/preval/preval_deliverables.htm)
[28]	Hays, W.L. (1981). Statistics (third edition). New York, USA: Holt, Rinehart and Winston.
[29]	Keppel, G. (1991). Design and Analysis – A Researcher's Handbook (3rd ed.). Englewood Cliffs, New Jersey, USA: Prentice Hall.
[30]	Siegel, S. (1956). Nonparametric Statistics for the behavioral sciences. New York, USA: McGraw-Hill Book Company Inc.

[31] Suter, W.N., Lindgren, H.C. (1989). Experimentation in psychology. Massachusetts, USA: Allyn and Bacon

[32] ISO 9241-11 (1998). Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11 :Guidance on usability: http://www.idemployee.id.tue.nl/g.w.m.rauterberg/lecturenotes/ISO9241part11.pdf

[33] Brooke, J. (1996) *SUS: a "quick and dirty" usability scale*. In P. W. Jordan, B. Thomas, B. A. Weerdmeester & A. L. McClelland (eds.) Usability Evaluation in Industry. London: Taylor and Francis.

[34] Nielsen, Jakob: Usability Engineering. Academic Press, Boston 1993.

[35] Stanton, N., Salmon, P.M.,Walker. G-H., Baber, C., and Jenkins, D.P., “Human Factors Methods”, 2005

[36] Usability Net: http://www.usabilitynet.org/tools/subjective.htm

[37] VAN DER LAAN, J. D., HEINO, A. & DE WAARD, D. (1997) A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research Part C-Emerging Technologies, 5, 1- 10.

[38] NASA TLX is free for download at the following page: http://humansystems.arc.nasa.gov/groups/TLX/index.htmlNeville

[39] Son, H., Kweon, Y.-J., Park, B., Development of Crash Prediction Models using Real Time Safety Surrogate Measures, report VACTS-15-0-70, University of Virginia, 2008

[40] Scholliers, J., et al., Framework for the assessment of preventive and active safety functions, Annex E to IP_D12/D16.4 Project final report and recommendations for future assessment, 31.01.2008, available at http://www.prevent-ip.org/en/public_documents/deliverables/d164_preval_final_report.htm

[41] Production Information Brochure ‘ITS Modeller’: http://www.tno.nl/downloads/ITS%20modeller.pdf

[42] D4.6.3 – Results Evaluation

[43] “SP4 WP6 TEST CASE example”, at the BSCW link: http://bscw.safespot-eu.org/bscw/bscw.cgi/193084

[44] “SF_WP4.6_Template_Tutorial-v1.0”, at the BSCW link: http://bscw.safespot-eu.org/bscw/bscw.cgi/193084

[45] Mortara, P., De Gennaro, M., “Non Compliance Management Tracking Form”, at the BSCW link: http://bscw.safespot-eu.org/bscw/bscw.cgi/187954 .

[46] Mortara, P., De Gennaro, M., “Non Compliance Management Tracking Form – Tutorial” (internal report), at the BSCW link: http://bscw.SAFESPOT-eu.org/bscw/bscw.cgi/187949

9. Annex

The Annex to this document can be found in a separate document.

SF_D4.6.1_PilotPlan_Annex.doc

The Annex contains three sections:

- Test check list for standardized procedure during the pilots
- Questionnaires for printout for pilots with subjects
- Compilation of all filled in pilot plan test case templates