PANACEA

FITNESS TO DRIVE

PracticAl and Effective tools to moNitor and Assess CommErciAl drivers' fitness to drive

Grant Agreement Number: 953426

D6.1: Evaluation framework, plans and material

This report is part of a project that has received funding by the European Union's Horizon 2020 research and innovation programme under Grant Agreement number 953426.

Legal Disclaimer

Executive Summary

This deliverable presents the evaluation framework, plans and material for all data collections of the PANACEA project. It describes the objectives of the studies and how they will be realised. The purpose of the PANACEA evaluation framework is to create a common framework to be used in all studies to make sure the data are collected in a way that makes it possible to consolidate the results at the end and to provide what is needed for impact analysis (WP7). This version of the deliverable has its focus on setting the framework and the work process. An update of this deliverable will be done in M22 (D6.2: 'Evaluation framework, plans and material - an update'). The key content of D6.1 is structured as follows:

Chapter 1 is the introduction to the deliverable, specifying its purpose, the intended audience, and interrelations with other project activities. **Chapter 2** introduces the project objectives related to the WP6 data collections. **Chapter 3** provides a brief overview of each Use Case and **Chapter 4** presents the various studies within the project including descriptions of the main actors, environment, vehicles, PANACEA sensors/technologies, and countermeasures. **Chapter 5** describes the process of developing the evaluation framework for the project and presents the PANACEA evaluation framework. Chapters 6-18 then follow the steps defined in the evaluation framework. **Chapters 6-11** describe the planning phase and present the Use Case Scenarios, Research Questions, Key Performance Indicators, study designs, data gathering tools, and data analysis plan. **Chapters 12-14** describe the implementation phase, including pilot site preparations, data collection, and data delivery. **Chapters 15-18** describe the data analysis phase and includes chapters about data analysis, results reporting, results consolidation, and impact assessment. Lastly, **Chapter 19** provides the conclusions of the deliverable.

The deliverable presents both a horizontal perspective of the pilot sites as well as more detailed descriptions of what will be included in the different studies. The general data gathering tools (objective and subjective) are identified and will be further refined in the update of the deliverable. A set of guidelines on practicalities and ethical aspects to take into consideration before and during data collection are presented.

The update of the deliverable, planned for M22, will include the detailed evaluation protocols, with ready-made templates for pilot sites, questionnaires to use, performance criteria, indicators, log files to use, crucial timelines, etc. In addition, the final pilot and experimental plans will be defined and described per pilot site and type of evaluation activity.

Document Control Sheet

Start date of project:	01 May 2021
Duration:	36 months
Del. ID & Title:	Deliverable 6.1: Evaluation framework, plans and material
Dissemination level:	PU
Relevant Activities:	A6.1: Pilot tools and framework
Work package:	WP6
Lead authors:	Anna Sjörs Dahlman (VTI), Anna Anund (VTI)
Other authors involved:	Katerina Touliou (CERTH), Selpi Selpi (Chalmers), Beatriz Delgado (Datik), Anastasia Azarko, Davide Shingo Usami (CTLUP), Sandra Trösterer, Cyril Marx (VIF)
Internal Reviewers:	VIF, CERTH
Actual submission date:	DD/08/2022 (M16)
Status:	DRAFT
File Name:	PANACEA_D6.1_ Evaluation framework, plans and material _v1.0

Document Revision History

Version	Date	Reason	Editor
0.1	11/04/2022	Table of Contents for feedback.	Anna Sjörs Dahlman (VTI)
0.2	25/04/2022	Updated table of contents and first suggestion of framework. Description of UC and UCS added.	Anna Sjörs Dahlman (VTI), Katerina Touliou (CERTH)
0.3	23/05/2022	Updated framework based on feedback. Introduction and	Anna Sjörs Dahlman, Anna Anund (VTI),

Version	Date	Reason	Editor
		framework chapters completed.	feedback from WP6 partners.
0.4	13/07/2022	Chapters 4, 7, 9, and 12 added.	Anna Sjörs Dahlman, Anna Anund (VTI), Katerina Touliou (CERTH), Selpi Selpi (Chalmers), Beatriz Delgado (Datik), Sandra Trösterer, Cyril Marx (VIF)
0.5	22/07/2022	Chapters 8, 18 and appendices added. Conclusions and executive summary completed.	Anna Sjörs Dahlman (VTI), Davide Shingo Usami (CTLUP), Pilot site representatives.
1.0	28/07/2022	Version sent for internal peer review.	Anna Sjörs Dahlman (VTI)
2.0	29/08/2022	Peer reviewed version sent for submission.	Anna Sjörs Dahlman (VTI)

Table of Contents

List of Tables

List of Figures

Abbreviation List

Abbreviation	Definition
ADAS	Advanced Driver Assistance System
AIT	Austrian Institute of Technology GmbH
AUDIT	Alcohol Use Disorders Identification Test
AV	Automated Vehicle
BMM	Biomathematical Model
BrAc	Breath Alcohol content
CAN	Controller Area Network
CDC	Centers for Disease Control and Prevention
CHT	Commercial Health Toolkit
CONSORT	Consolidated Standards of Reporting Trials
DBL	Deep Blue S.r.l.
DDA	During Driving Assessment
DoA	Description of Action
DPIA	Data Protection Impact Assessment
DSS	Decision Support System
EB	Ethics Board
ECG	Electrocardiography
EEG	Electroencephalography
EOG	Electrooculography
EU	European Union
FESTA	Field opErational teSt supporT Action

Abbreviation	Definition
FOT	Field Operational Tests
GDPR	General Data Protection Regulation
GSR	Galvanic Skin Response
KPI	Key Performance Indicator
KSQ	Karolinska Sleep Questionnaire
KSS	Karolinska Sleepiness Scale
MS	Milestone
OBJ	Objective
ODA	Off Duty Assessment
OM	Outcome Mapping
ONPDA	On site & Pre-Driving Assessment
PTW	Powered Two-Wheeler
PWA	Pulse Wave Analysis
QoL	Quality of Life
RCT	Randomized Controlled Trials
RQ	Research Question
RSA	Roadside Assessment
SATI	SHAPE Automation Trust Index
SD	Standard Deviation
SE	Sensitivity
SP	Specificity
STROBE	STrengthening the Reporting of OBservational studies in Epidemiology

Abbreviation	Definition
SUS	System Usability Scale
TAC	Transdermal Alcohol Content
TAQ	Technology Acceptance Questionnaire
UC	Use Case
UCS	Use Case Scenario
UCscr	Use Case script
VIF	Virtual Vehicle research GmbH
VRU	Vulnerable Road User
VSS	VTI acute Stress Scale
VTI	Swedish National Road and Transport Research Institute
WP	Work Package

1 Introduction

The **Evaluation framework, plans and material** deliverable creates and presents a clear framework for all planned data collections needed for the evaluation work of the PANACEA project. The PANACEA project will create commercial driver-oriented, health-based and Use Case (UC)-driven health monitoring and assessment methodologies and technical solutions, i.e., 'Commercial Health Toolkits' (CHT) and develop an effective strategic, tactical, and operational cloud-based coaching & supporting solution for commercial drivers. The PANACEA solution, including the CHTs and countermeasures' solutions will be evaluated in an iterative process. The data collections needed for the evaluations include both simulator studies, real-world evaluations, and roadside assessments. All material needed to complete the data collections, such as templates to be filled in, questionnaires to use, performance criteria, indicators, log files to use, timelines, etc. will be defined here. Initially, a general evaluation framework will be established and the principles for the data gathering tools will be developed to be applicable to all the project's UC. Based on the general framework, individual evaluation strategies will be designed that fulfil the requirements of each individual data collection. The planned data collections have a variety of study designs and purposes but nonetheless the methodology is kept as similar as possible across pilots. To achieve a harmonized way of collecting data and ensure good quality of the data collected, a general evaluation framework has been developed. This framework adheres to existing transportation frameworks (e.g., FESTA (Barnard et al., 2016)), but additionally incorporates components from clinical and experimental protocols, necessary to address the elements and dimensions of the evaluation objectives and the relevant project objectives. The deliverable provides a common template for harmonising and coordinating all tests with drivers at an early stage, to optimise the consolidation that will be made in A6.5.

Three types of studies are included in the PANACEA project; 1) simulator and roadside studies aiming to refine the algorithms developed and offer the possibility of repetition of measures to reach the targeted sensitivity (SE) and specificity (SP) levels per CHT identified in the project, 2) validation and assessment pilots for evaluation of the CHTs at three pilot sites, and 3) countermeasures' pilots where evaluation of both the content and the actual online coaching system will be performed in parallel with the CHT pilots.

1.1 Purpose of the document

The purpose of the Evaluation framework, plans and material is to create a common framework to be used in all Work Package 6 (WP6) data collections, to make sure the data are collected in a way that makes it possible to consolidate the results of the pilots' evaluations and to provide what is needed for impact analysis. The deliverable will describe what kind of data that will be collected, what the purpose is, how the data will be used in the project, and by whom it will be collected.

The study designs will differ between pilots, depending on the specific aim of each data collection. However, the Evaluation framework, plans and material will ensure that a common process for planning and implementation of data collection, analysis of data, and results reporting will be followed at all pilot sites.

The first version of the Evaluation framework, plans and material defines the evaluation framework, its dimensions and the overall KPIs. Moreover, it will include the first version of the pilot plans and selections of data collection tools. The update of the deliverable, planned for M22, will include the detailed evaluation protocols, with ready-made templates for pilot sites, questionnaires to use, performance criteria, indicators, log files to use, crucial timelines,

etc. In addition, the final pilot and experimental plans will be defined and described per pilot site and type of evaluation activity.

1.2 Intended Audience

The intended audience of the document is both internal to the project and external. The deliverable serves as a manual for the pilot sites in their planning and conduction of data collections. It is also an informative document to describe to external stakeholders how the PANACEA solution will be evaluated in the project.

1.3 Interrelations

The data collections covered by the Evaluation framework, plans and material deliverable are highly interrelated to many other activities in the PANACEA project. Firstly, WP1 developed the UCs and Use Case Scenarios (UCS) to be evaluated by WP6. Secondly, the main purpose of the deliverable is to provide the framework for the evaluation work of WP3, WP4, and WP5.Therefore, there will be extensive collaboration between WP6 and WP3, WP4 and WP5. All WP6 data collections are dependent on verification and validation performed in WP2 and WP4 before final evaluation of the PANACEA solution can start. A6.2 will collect data to improve/create algorithms of WP3 and improve/define the thresholds for each impairing state addressed. A6.3 and A6.4 deals with the PANACEA solution validation and assessment and will thus depend on the development of various parts of the PANACEA solution performed in WP2, WP3, WP4, and WP5. The results of WP6 will then be fed to WP7 for the impact assessment.

2 Project objectives

PANACEA aims to create a holistic pre-, during and roadside driving ability monitoring and assessment system. The system will reliably and efficiently assess the physical, cognitive, and psychological Fitness-to-Drive of commercial drivers. In cases of impairment, a complementary cloud-based countermeasures and coaching tool will deploy appropriate solutions targeting drivers, operators, and enforcement. Below, the objectives that are directly and indirectly relevant to the WP6 are included.

The objectives directly relevant to WP6 are the following:

OBJ3: Evaluate the usefulness, ease-of-use, satisfaction, and acceptance of the CHTs across 3 UC-driven Pilots, considering gender specificities (WP6).

OBJ4: Evaluate an effective strategic, tactical, and operational cloud-based coaching & supporting solution for commercial drivers combating driver impairment (WP5 & WP6).

The objectives that are indirectly relevant either by being a prerequisite for the WP6 studies, by using data collected during the WP6 studies or by use of the inferences drawn are OBJ1, 2, 5, and 7.

OBJ1: Create commercial driver-oriented, health-based and Use Case (UC)-driven health monitoring and assessment methodologies and technical solutions (i.e., 'Commercial Health Toolkits'; CHTs). The platform will be developed in WP2, the content and the algorithms in WP3 and the actual systems and the Decision Support System (DSS) in WP4.

OBJ2: Estimate the sensitivity, specificity, effectiveness, and operability of CHTs for alcohol, licit (benzodiazepines), illicit (methadone) drugs, fatigue, stress and cognitive load. The CHTs will cover before/ after/ during shifts as well as on-site (for fleet operators) and roadside (for enforcers; WP5 & WP6).

OBJ5: Create a new paradigm in Fitness to Drive (Fitness to Drive 2.0), considering new technologies and commercial vehicles' varying automation levels (WP3, WP4, WP5 & WP6).

OBJ7: Assess the safety, socioeconomic and Quality of Life (QoL) impacts of CHTs and relevant monitoring, assessment and coaching solutions and policies Europe-wide (WP7).

3 Use cases

Use Cases in PANACEA comprise the technologies, the actors involved, the vehicles they drive, and the impairments addressed at each of the three pilot sites (Sweden in UCA, Greece in UCB, Spain in UCC). They were developed in WP1 and a more detailed description of the UCs can be found in D1.1: 'Use Cases'. The driver clusters addressed per UC are shown below.

Use Case (UC)	Target drivers
A	Bus/shuttle drivers
B	Powered Two-Wheeler (PTW) courier delivery riders
B	Taxi drivers
C	Coach driver
C	E-truck driver (refuse/rubbish/garbage collection)

3.1 UCA

The target population in UCA is bus drivers who are also safety operators for autonomous shuttles. The focus is on the safety during shuttle operation in Linköping, Sweden. Key considerations are the impact of shift work, task related fatigue, and the need to interact with Vulnerable Road Users (VRU). It is intended that the PANACEA system will detect fitness-to-drive prior to starting work as this is the priority to ensure people are fit to drive when starting work. In addition, it is necessary to take into consideration that the task is very monotonous, so fitness (particularly alertness) needs to be maintained throughout shift. There is also a need to prepare drivers ahead of their future shifts. To make this happen also the operator is important. They need knowledge on the driver's status and how to plan to support the drivers and avoid unnecessary demanding shifts.

Priority: off-duty (lifestyle, to ensure fitness prior to starting the work shift), on-duty (pre-driving, the driver is at work and should be assessed before they are allowed in the vehicle), on-road (in the vehicle while driving as a guidance/assistance system).

3.2 UCB

Taxi drivers and courier service riders who work in the prefecture of Thessaloniki, Macedonia, Greece are targeted. Key considerations are the impact of stress, fatigue, alcohol, and (il)licit drugs consumption. Fitness will be assessed across all work shift phases with emphasis pre- and during the shift. It is very important to accommodate for the conditions that both types of professionals work in. For example, taxi drivers often drive in unfamiliar and not pre-scheduled routes, whereas courier service riders often know the delivery routes at the beginning of their shift. However, they both experience dense urban traffic and the related risks. Taxi drivers are often self-employed and freelancers, whereas the courier service riders are employees, as is the case with the target populations in the other two Use Cases.

Priority: on-duty (pre-driving, the driver is at work and should be assessed before they are allowed in the vehicle), on-road (in the vehicle while driving as a guidance/assistance system).

3.3 UCC

Truck and coach drivers that work in the San Sebastian and Barcelona area in Spain are targeted. Key considerations are the impact of shift work and different impairing states like stress and fatigue. It is intended that the PANACEA system will detect fitness prior to starting work as this is the priority to ensure people are fit to drive when starting work. In addition, for coach drivers it is necessary to understand that the task is very monotonous, so fitness

(particularly alertness) needs to be maintained throughout the shift. For truck drivers, the task is carried out in night shift which means extra effort to keep alert. There is also a need to prepare drivers ahead of their future shifts. Urban, inter-urban and rural road conditions are included.

Priority: off-duty (lifestyle, to ensure fitness prior to starting shift), on-site (pre-driving, the driver is at work and should be assessed before they are allowed in the vehicle), on-duty (in the vehicle while driving as a guidance/assistance system)

4 Pilot sites and studies

There are three main pilot sites in the project, related to the Use Cases A, B and C. In addition, a roadside pilot will be performed. Below is a description of each site including a description of the objectives, main actors, environment, vehicles, PANACEA sensors/technologies used to measure driver impairment, and countermeasures developed and tested. Simulator and roadside studies will be performed and serve the purpose to develop and test the PANACEA system. Real-world and semi-real-world studies will then be performed at the pilot sites to evaluate the system in operational settings.

4.1 UCA

The focus in PANACEA is to develop and evaluate a system that integrate sensors used to detect and avoid driving under impairment. Here alcohol/ drug use, fatigue and stress are of major interest, and the countermeasures that are relevant from strategical, tactical and operative level.

The A6.2 simulator study will be performed in a driving simulator at the Swedish National Road and Transport Research Institute (VTI) premises in Linköping, Sweden to enable safe testing of driving under the influence of alcohol. Real-life data collection in the A6.3 and A6.4 study will be conducted with autonomous shuttles in the nearby University campus and residential area.

4.1.1 Simulator study (UCA-S)

The objectives are to learn more about how moderate amounts of alcohol in the evening affects night sleep and next day driving performance and based on this develop a first version of a biomathematical model of fatigue (WP3) that takes next-day effects of alcohol into account.

The data collection needed will be done at VTI using two fixed based driving simulators in parallel. The simulators have three computer screens and a vehicle mock-up, see Figure 1. A total of 30 male drivers aged 25-50 years old will be included in the study and the data will be used to update the fatigue algorithms with data on alcohol sleep on fatigue development. The scenario will include both urban and rural road driving.

Figure 1. *Driving simulator environment.*

The PANACEA sensors to be included are: AIT smartPWA (Pulse Wave Analysis)., and Fitbit.

4.1.2 Real-world pilot (UCA-R)

The objective is to evaluate and assess the CHT-A and its countermeasures addressing both autonomous shuttle drivers and the managers.

The evaluation will be done in Linköping, Sweden at a site that consists of a 4.1 km long route including roads with both mixed traffic, meaning interaction with other motorized vehicles, but also a dedicated area with only pedestrians and cyclists allowed, see Figure 2. It covers the Linköping University campus and a residential area called Vallastaden. Two EasyMile autonomous shuttles using 13 bus stops will be included. The service is up and running 7 days a week according to a frequency-based timetable.

Figure 2. *An overview of the Linköping site (UCA). Left: route, Middle: EasyMile shuttle Right: EasyMile shuttle in Vallastaden.*

At the site there are 8 safety drivers working approximately 60 percent of their time as shuttle operators and the rest as city bus driver and/or tram driver. In addition, 2 managers will be involved. The impairments in focus are alcohol/ drug use, fatigue and stress, and the countermeasures that are relevant cover both strategical, tactical and operative level.

The sensors to be included are: DATIK FitDrive, AIT smartPWA, Senseair Wall, Leitat biosensor, Fitbit, BMM, and BACtrack Skyn.

For UCA safety drivers the selection of countermeasures defined in A5.2 are shown in Table 1 and countermeasures for managers are shown in Table 2.

Table 1. *Selected countermeasures for drivers.*

	Operational	Tactical	Strategic
UCA	**-Caffeine and napping advice for fatigue when sleepiness signs are detected** -Self-management of stress/cognitive load during shift	**-Raising awareness of fatigue for drivers, providing sleep/recovery advice before/after work** *-Advice about alcohol use before work (not during shift) e.g., evening before*	**-Lifestyle coaching relating to sleep and fatigue** (could inc. alcohol) -Lifestyle coaching for optimising rest (off duty) time in terms of reducing stress and related fatigue
UCB	**-Self-management of stress/cognitive load during shift (could inc. headway management)**	**-Advice about licit drugs prior to shift (taken the night before a morning shift or in the morning of**	**-Lifestyle coaching relating to stress and cognitive load**

	Operational	Tactical	Strategic
	-Guided breathing exercises	**a morning shift) focus on immediate and residual effects**	*-Lifestyle coaching relating to prescription drugs*
UCC	**Providing message, auditory, visual and/ or haptic warning/alert to a driver and operator that fatigue has been detected** -Self-management of stress/cognitive load during shift (could inc. headway management) -Caffeine and napping advice for fatigue when sleepiness signs are detected	Raising awareness of fatigue for drivers, providing sleep/recovery advice before/after work	**Lifestyle coaching for optimising rest (off duty) time in terms of reducing stress and related fatigue** Lifestyle coaching relating to sleep and fatigue

Table 2. *Selected countermeasures for operators/managers.*

	Operational	Tactical	Strategic
UCA	-Changing driver due to fatigue **-Changing driver due to alcohol** -Advice to operator on how to action results of DATIK pre-questionnaire (e.g., change driver/nap/caffeine) *-Providing facilities for rest breaks*	**-Advice/tools for Scheduling and how work is distributed within a shift** -Training on how to use and interpret PANACEA system -Training for managers in how to identify stress in drivers/when driving	**-Training and education on impact of alcohol and fatigue on driving** -Training and education on impact of licit/illicit drugs on driving *-Driver impairment risk management system* *-Establishing open culture to encourage reporting of PANACEA related impairment*
UCB	**-Advice to operator on how to action results of DATIK pre-questionnaire (e.g., change driver/nap/caffeine)**	-Training on how to use and interpret PANACEA system *-Medical assessment when drivers join company - licit drugs*	**-Training and education on impact of licit/illicit drugs on driving** **-Training and education on medication management**

	Operational	Tactical	Strategic
			-Training and education on impact of alcohol on driving
UCC	**-Changing driver due to fatigue** -Changing driver due to alcohol -Advice to operator on how to action results of DATIK pre-questionnaire (e.g., change driver/nap/caffeine) *-Providing facilities for rest breaks*	**-Training for managers in how to identify stress in drivers/when driving** -Training on how to use and interpret PANACEA system	**-Training and education on impact of fatigue on driving** -Training and education on impact of alcohol on driving *-Driver impairment risk management system* *-Establishing open culture to encourage reporting of PANACEA related impairment*

4.2 UCB

Two pilot sites will participate in the studies connected to UCB. The A6.2 studies will be performed at the site in Thessaloniki, Greece (CERTH) and at the site at Austria (ViF). The real-life pilots (A6.3 and A6.4 activities) will be conducted under controlled conditions in the area of CERTH premises due to ethical and legal restrictions (potential consumption of alcohol and drugs will be included). It includes the CERTH and ViF driving and the CERTH riding simulation laboratories (A6.2) and the CERTH premises (A6.3 and A6.4).

The infrastructure for the simulator pilots are the two passenger car simulators in CERTH and ViF premises, the motorcycle simulator at CERTH and an instrumented passenger car and motorcycle for the real-life tests inside the CERTH premises.

Fatigue, alcohol consumption and stress will be addressed in A6.2 pilots in Thessaloniki, Greece and distraction in Austria. Fatigue and stress will be addressed in semi-real-life conditions in A6.3/A6.4 pilots and alcohol and drugs will be addressed only in simulated environment due to legal and ethical restrictions.

4.2.1 Simulator study 1 (UCB-S1a UCB-S1b)

The objectives are to collect data for the refinement of the algorithms developed in WP3 and to ensure that the selected levels for the impairing and driver states are meaningful and measurable with targeted accuracy, sensitivity, and specificity. This will be done both for passenger car drivers (n=20) and for Powered Two-Wheeler (PTW) riders (n=20) and hence two different types of simulators will be used, see Figure 3. The car driving simulator is a dynamic car simulator with a complete car (SMART) on a rotating platform. The riding simulator is a dynamic motorcycle simulator. The simulator dynamics allow five degrees of freedom (roll, pitch, yaw, handlebar extension and shortening). The visual system of the

simulator employs three projection screens that cover the riders' field of view and an instrument panel with an LCD screen that presents information through the simulator Controller Area Network (CAN) bus and can be also used on a motorcycle.

Figure 3. *Driving simulator (left) and riding simulator (right).*

The environment will be peri-urban and urban and the impairments in focus will be fatigue, alcohol consumption and stress.

The PANACEA sensors to be included are: Datik FitDrive, AIT smartPWA, Senseair Wall and Go, BACtrack Skyn, Optalert, and GSR sensors.

4.2.2 Simulator study 2 (UCB-S2)

The objective with the simulator study at VIF is to evaluate different types of driver distraction (cognitive, visual) in different driving environments (urban vs. highway) to collect data for the development of a multisensory fusion algorithm for detecting a distracted driver state. Both steering / use of the steering wheel and visual behaviour will be included. The environment will be an urban road and a highway. Twenty experienced drivers will participate in the trials. The simulator can be seen in Figure 4.

Figure 4. *Driving simulator that will be used in the study at ViF.*

The PANACEA sensors to be included are AIT smartPWA and DBL index.

4.2.3 Semi-real-world pilot (UCB-R1 and UCB-R2)

The objective is to evaluate the performance and user experience of the holistic system in a semi-real life condition considering driver impairments caused by stress, alcohol and fatigue as well as the countermeasures use and compliance.

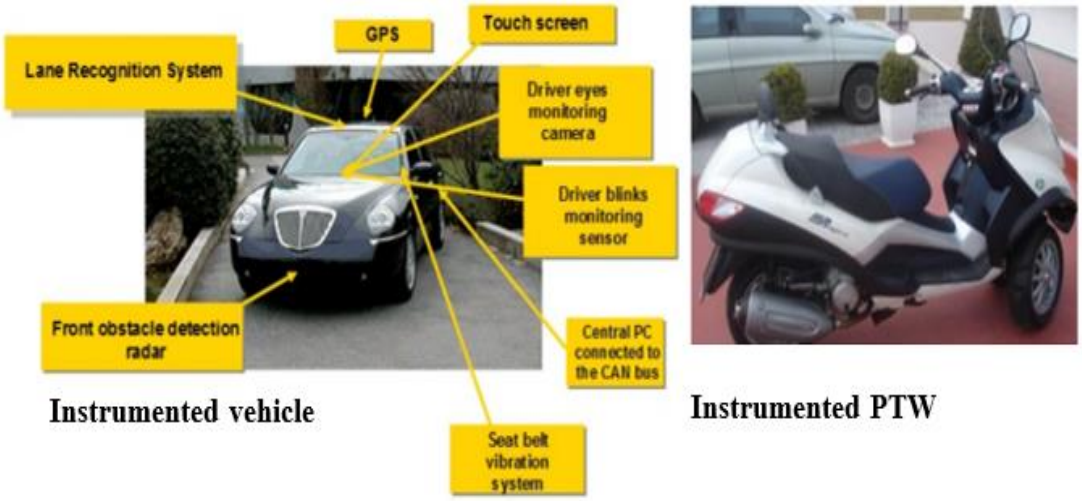


Figure 5. *Instrumented vehicle and instrumented PTW to be used at the test area in UCB.*

The environment will be real life testing in a controlled and closed traffic area with riders **(UCB-R1)** and taxi drivers **(UCB-R2)**. An instrumented vehicle will be used and a motorcycle, see Figure 5. The fatigue and stress tests will be conducted in the CERTH area, as shown in Figure 1. The alcohol and drug tests will be conducted in the CERTH riding and driving simulators (same as in UCB-S1) for ethical and legal reasons. There will be 20 taxi drivers and 20 delivery service riders participating in the study.

The PANACEA sensors to be included are: DATIK FitDrive, AIT smartPWA, ViF Driver Monitoring System, Senseair Wall and Go, BACtrack Skyn, Optalert, and GSR sensors.

The selected countermeasures for UCB drivers/riders and operators are presented in Table 1 and Table 2, respectively.

4.3 UCC

The UCC is focused on professional drivers and their managers running operations with garbage trucks and regular buses. This UC includes only real-world studies, and the data collection will be done in the Barcelona and San Sebastián areas of Spain. The focus on driver impairments in Spain site are alcohol/drug use, fatigue and stress detection. There will be three data collections at two locations for the use case.

- The R1 site is an urban scenario in Barcelona with two garbage trucks.
- The R2 site will be interurban coach travel between cities (start in San Sebastián).
- The R3 site will be a long-distance journey between two cities (start in San Sebastián).

4.3.1 Real-world pilot (UCC-R1, UCC-R2 and UCC-R3)

The objective is to evaluate the PANACEA a system with integrates sensors used to detect and avoid driving under impairment and the relevant countermeasures on strategic, tactical and operative level. In total 4 vehicles will be included in the evaluations (2 trucks and 2 coaches).

The type of professional drivers in focus are three different groups, see Table 3.

Table 3. *Driver profiles included in UCC evaluations***.**

SITE	VEHICLE	DRIVER	ITINERARY	SCHEDULE	Kms	OTHER
R1	ieTruck	Professional driver	From garage - urban - unloading point - urban garage	From 21 to 4	75/100 kms	
R2	Irizar i6s - MAN	2 - professional driver	Garage - Donosti -Bilbao (relief) garage	Morning shift 5:30/6/6:30 (depends) Afternoon shift 12:30/13/13:30 (depends)	450 kms	
R3	Irizar i6s	Professional driver	Garage - Donosti - París - garage	8 hours shift Morning shift starting at 5:30	420 kms	

UCC-R1: truck drivers drive an ieTruck (FCC) picking up garbage following a special service line in Barcelona. The drivers work night shifts only, see Figure 6. In total there will be 2 trucks equipped.

Figure 6. *Garbage Irizar truck in Bilbo Donosti*

UCC-R2: bus drivers drive a bus service in Bilbo Donosti with one departure every half hour from 6 am to 10 pm. Each service is 1 hour and 15 minutes. The drivers' shifts start and end depending on the first service assignment. There will be 2 drivers involved divided into morning or afternoon shift. There will be one coach equipped.

UCC-R3: two bus drivers drive a coach as a long journey's bus service with a starting point in San Sebastián going to Paris. They start 9:30 and arrive the destination at 20:10 the same day. They rest in Paris and start next morning at 08:30 for the return to San Sebastián. Here 8 drivers will be involved and they are grouped 2 in each group. In 2 of the groups, they work a fixed schedule with 4 days in a row and rest 2 days. Those days are driven by drivers in the remaining group. There will be one coach equipped.

The sensors to be used are: DATIK FitDrive, ViF Driver Monitoring System, AIT Smart PWA, Senseair Wall and Go, and LEITAT biosensor.

The selected countermeasures for UCC drivers and operators are presented in in Table 1 and Table 2, respectively.

4.4 Roadside study

Roadside assessment is an assessment normally conducted by an enforcer (i.e. police/authority) by asking a vehicle to stop to the side of the road, so the driver/ rider to be tested. The roadside study is related to the evaluation of the sensors developed for alcohol and drug testing at roadside. The objective is to evaluate the level of agreement between SENSEAIRs and LEITATs devices and the commercial devices currently in use by the Norwegian Police for roadside assessment (Dräger for alcohol and drug testing).

The alcohol roadside testing procedure in Norway is based on the regulation that a Breath Alcohol Control (BrAC) value >0.1 mg/L is seen as a positive sample and the driver needs to follow the police officer to the police station for additional breath or blood test. For drugs a similar procedure is followed, but here with different cut off values depending on the drug. In situations with positive tests the police also perform a "sign and symptom" test before bringing the driver to the police station for further blood testing.

In PANACEA the same procedure as normal will be followed, but with the PANACEA devices (SENSAIR & LEITAT) used in parallel with the normal devices as the police use today. Action taken due to positive answers will only be based on the devices the police normally for testing, not the PANACEA sensors. Countermeasures training of monitors and enforcement authorities are presented in Table 4.

Table 4. *Selected countermeasures for enforcers.*

	Operational	Tactical	Strategic
Alcohol	**-Roadside testing**	**-Training for enforcement offices for use of the PANACEA system** -Awareness campaign for roadside testing **-Provide guidance to operators/drivers**	-Influence on regulatory framework -Influence on policy documents
Licit / illicit drugs	**-Roadside testing**	**-Training for enforcement offices for use of the PANACEA system** -Awareness campaign for roadside testing **-Provide guidance to operators/drivers**	-Influence on regulatory framework -Influence on policy documents

5 Evaluation Framework

Evaluation frameworks facilitate a systematic approach to evaluation. They can enable multiple stakeholders to gain a shared understanding of the evaluation process and help to identify and agree upon appropriate objectives and methods. A range of evaluation frameworks have been published, both generic frameworks intended for use across a range of contexts, settings and sectors, and frameworks developed for use in a specific context or field (Fynn, Hardeman, Milton, & Jones, 2020). An evaluation framework sets out the plan for data collection, analysis, and reporting. The goal is to achieve effective and systematic data collection to provide a solid evidence base for assessment of progress and impact over time.

Reviews of evaluation frameworks have concluded that here is an abundance of frameworks available but no single framework that covers all aspects of evaluation (Fynn et al., 2020; Newman-Askins, Ferreira, & Bunker, 2003; Yusof, Kuljis, Papazafeiropoulou, & Stergioulas, 2008). Several transportation system evaluation frameworks exist but these mostly focus on evaluation of the societal impact and economic benefits (He, Zeng, & Li, 2010; Newman-Askins et al., 2003) or environmental impact (Jansuwan, Liu, Song, & Chen, 2021).

Many of the frameworks used in the automotive industry have their starting point in a specific technology that will be evaluated. This usually means that the evaluation is use case driven. In other fields, especially in basic research, the research questions are instead the starting point, and the focus is on creating knew knowledge. The study design and technologies used are then tailored to best answer the research questions. On the other hand, the PANACEA project has its starting point in the project objectives, which are a combination of technology development, technology evaluation, knowledge creation, and impact assessment objectives. The PANACEA evaluation framework thus needs to cover all these aspects. As there are several data collections planned, with very different purpose and settings, the framework needs to be flexible enough to allow for a range of different study designs. Moreover, the development of the PANACEA solution will follow an iterative development process, where the results of initial WP6 data collections will be fed back to WP3, WP4, and WP5 to refine the CHTs and countermeasures and coaching solutions. Lastly, the framework needs to cover the evaluation of the final PANACEA solution, including the technical performance, usefulness and operability, user experiences, safety, socioeconomic impact etc. Therefore, an evaluation framework was developed within A6.1 of the PANACEA project by reviewing and combining components from commonly used frameworks in both the automotive and clinical research field.

5.1 Review of evaluation frameworks

Six frameworks were selected for a review. This was not intended as a systematic review of frameworks, rather a comparison of a few commonly used frameworks from different fields. The purpose was not to select the best one and implement in PANACEA but to be inspired by previous work in setting up a tailored framework for PANACEA.

The frameworks chosen for review were the transportation related frameworks FESTA, Trilateral Impact Assessment Framework and System Dynamic modelling, the more general Rainbow framework, and the healthcare related initiatives STROBE and CONSORT and the Framework for Program Evaluation in public health.

5.1.1 FESTA

The methodology was developed for Field Operational Tests (FOTs) by the European union funded project FESTA (Field opErational teSt supporT Action). The FESTA project developed a

handbook on FOT methodology to improve comparability and significance of results at national and European levels (Barnard et al., 2016). A FOT is here defined as a study undertaken to evaluate a function, or functions, under normal operating conditions in road traffic environments typically encountered by the participants to identify real world effects and benefits. FOTs were introduced as an evaluation method for driver support systems and functions with the aim of proving that such systems can deliver real-world benefits. Although the FESTA methodology and handbook was originally developed for FOTs, its basic steps are applicable for a wide range of field and user tests.

Figure 7 shows the FESTA V-diagram, i.e. the steps that are followed during the evaluation. The blue boxes represent the sequential steps to follow and the grey arrows show how to work through these steps from preparing the study, to using the prepared material during data acquisition and finally analysing the collected data.

The starting point in the FESTA is the function or system that will be evaluated. This can be an Advanced Driver Assistance System (ADAS) that has already passed the basic verification but now will be evaluated in an operational environment. In FESTA it is recommended that the system is compared with a baseline condition (i.e., driving without the system).

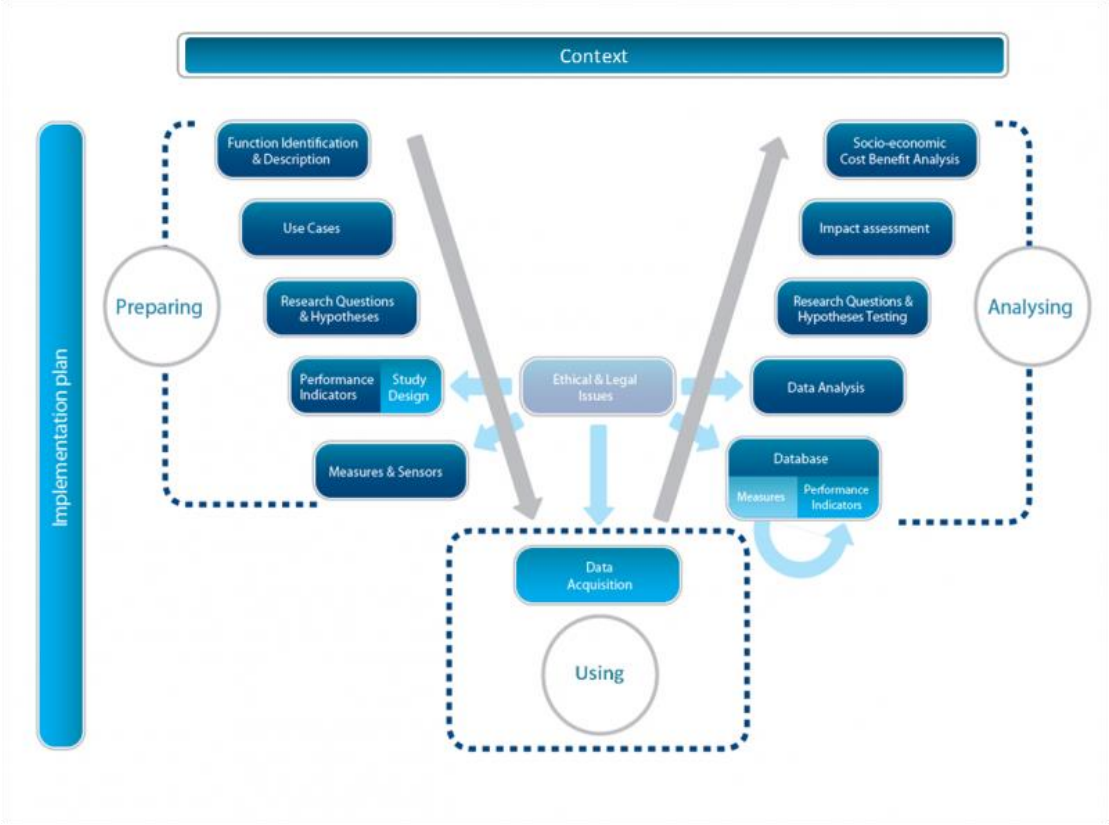


Figure 7. *FESTA V-diagram*

5.1.2 Trilateral Impact Assessment Framework for Automation in Road Transportation

The trilateral Impact Assessment Framework for Automation in Road Transportation was developed in cooperation between EU, US and Japan. The purpose was to harmonize the impact assessments performed in the field of automated driving, across the three regions (EC, US and Japan). The framework does not give detailed methodological recommendations, but it aims to facilitate meta-analysis across different studies. Therefore, the focus is on providing recommendations on how to describe the impact assessment study in a way that the user of

the results understands what was evaluated and under which conditions. The framework is partly based on the FESTA framework. It is a high-level framework and includes recommendations and advice on; classification of evaluated system/service, common vocabulary, direct and indirect impacts in 12 impact areas, impact mechanisms & paths, recommendations for experimental procedures, recommendations for data sharing, and KPI repository. Although, this framework was based on FESTA, it really focuses on impact assessment and automation, hence, only partially fitting the evaluation objectives of this project.

5.1.3 System Dynamic modelling

As a continuation and refinement of the trilateral framework, EU-US-Japan Trilateral Sub-Working Group for Impact Assessment, under the Trilateral Working Group for Automation in Road Transportation, has begun to use system dynamics to gain further insights into potential impacts. They developed a general framework from which detailed system dynamics models can be created for specific research questions (Rakoff et al., 2020). The work is ongoing and the goal is to develop a quantitative tool that can help planners and policy-makers understand how highly automated vehicles may fit within the transport system, and to begin to explore consequences of potential actions under various scenarios. The attention is on AVs, but in order to understand the wider context the framework identifies the major generic roles within the transportation system and considers how they interact within the context of both traditional and new modes. Figure 8 shows a high-level schematic of the causal loop structure identified by the working group.

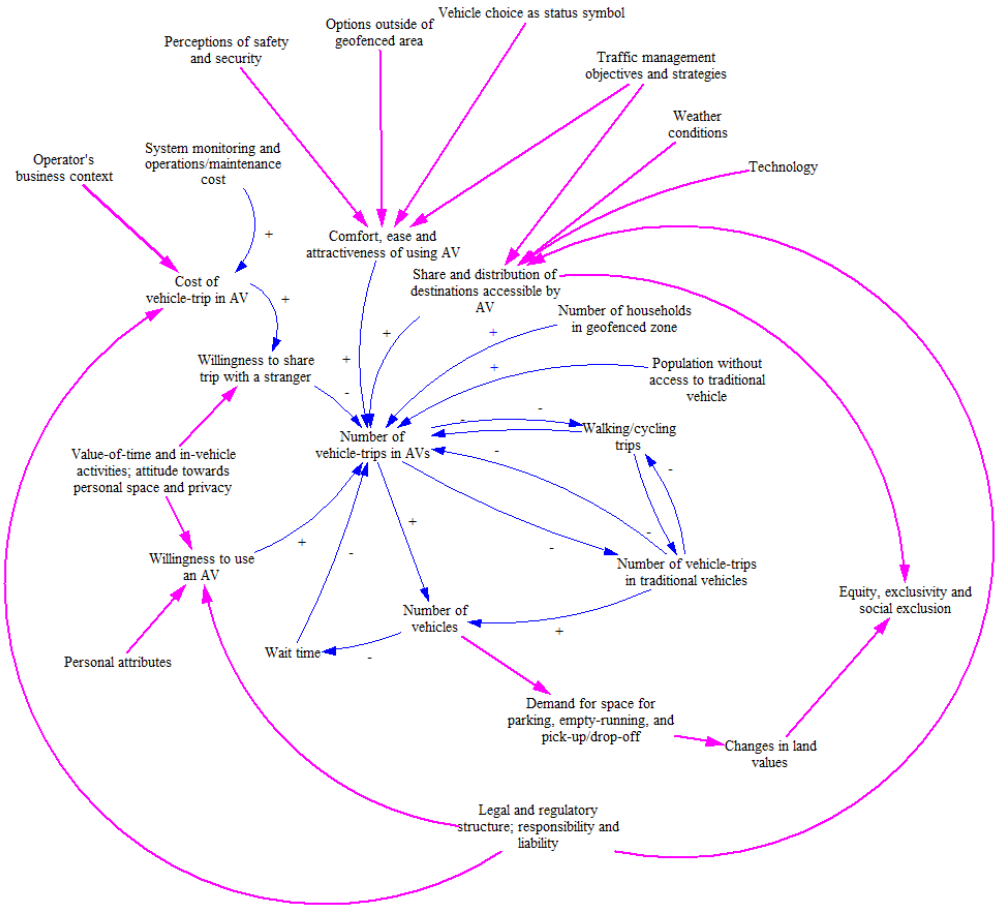


Figure 8. *Causal loop structure developed at a group model building workshop in Leeds UK, in April 2019 (Rakoff et al., 2020)*.

Arrows in Figure 8 indicate a link between a pair of items; links shown are not exhaustive but indicate the principal proposed impacts. Thin blue arrows have a clear proposed polarity whereas the pink arrows are the less-fully-defined groups of factors that can impact the system with a polarity that can be determined once the factor is better defined. A plus sign means that the link is positive (or "reinforcing") and a minus sign means that the link is negative (or "balancing"). This framework works well in identifying latent variables and complex interrelations, but it requires to be fed with considerable amounts of data, the types usually collected in large scale naturalistic and/ or field tests.

5.1.4 Rainbow framework

The Rainbow framework developed by BetterEvaluation describes the evaluation process in 34 different evaluation tasks, grouped by 7 colour-coded clusters. The purpose is to make it easy to choose and use appropriate methods, strategies or processes. It is a general framework that can be used for various types of studies, including Randomized Controlled Trials (RCTs) and Outcome Mapping (OM). The planning tool can be used to: commission and manage an evaluation; plan an evaluation; check the quality of an ongoing evaluation; embed participation thoughtfully in evaluation; develop evaluation capacity.

The clusters are named manage, define, frame, describe, understand causes, synthesize, and report and support use. Within each cluster, several tasks are listed and for each task a set of options are given (Figure 9). The framework provides many details around the planning, conduction, and reporting of data collections. It also covers general project management aspects that are out of the scope for the PANACEA evaluation framework. Moreover, the tasks in the define cluster and parts of the frame cluster were performed already in the application process and described in the grant agreement.

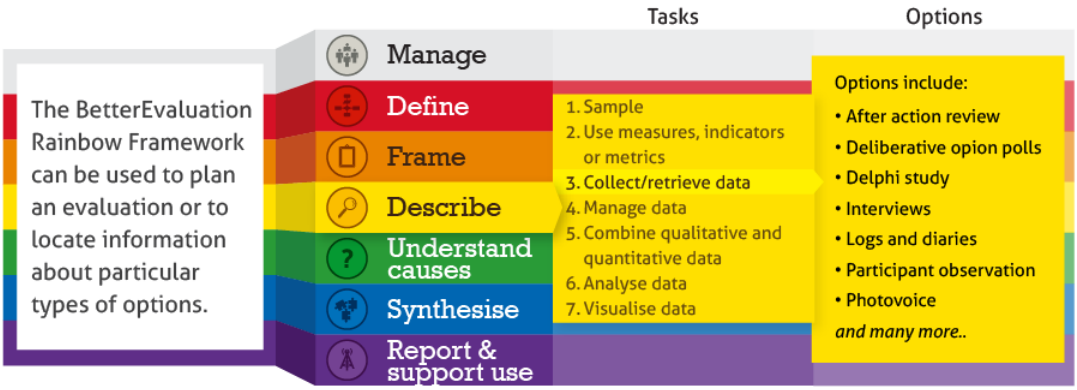


Figure 9. *Rainbow framework*

5.1.5 Framework for Program Evaluation in public health

The Framework for Program Evaluation in public health developed by the Centers for Disease Control and Prevention (CDC) is a practical, nonprescriptive tool, designed to summarize and organize essential elements of program evaluation (Milstein & Wetterhall, 1999). The general aim is to improve how program evaluations are conceived and conducted. The framework emphasizes six connected steps that together can be a starting point to tailor an evaluation for a particular effort, at a particular point in time (Figure 10). Because the steps are all interdependent, they might be encountered in a nonlinear sequence; however, an order exists for fulfilling each -- earlier steps provide the foundation for subsequent progress.

Figure 10. *Centers for Disease Control and Prevention Framework for program evaluation in public health MMWR 1999;48 (No. RR-11)*

The six steps in the evaluation process are:

- Engage stakeholders
  - Those persons involved in or affected by the program and primary users of the evaluation.
- Describe the program
  - Need, expected effects, activities, resources, stage, context, logic model.
- Focus the evaluation design
  - Purpose, users, uses, questions, methods, agreements.
- Gather credible evidence
  - Indicators, sources, quality, quantity, logistics.
- Justify conclusions
  - Standards, analysis/synthesis, interpretation, judgment, recommendations.
- Ensure use and share lessons learned
  - Design, preparation, feedback, follow-up, dissemination.

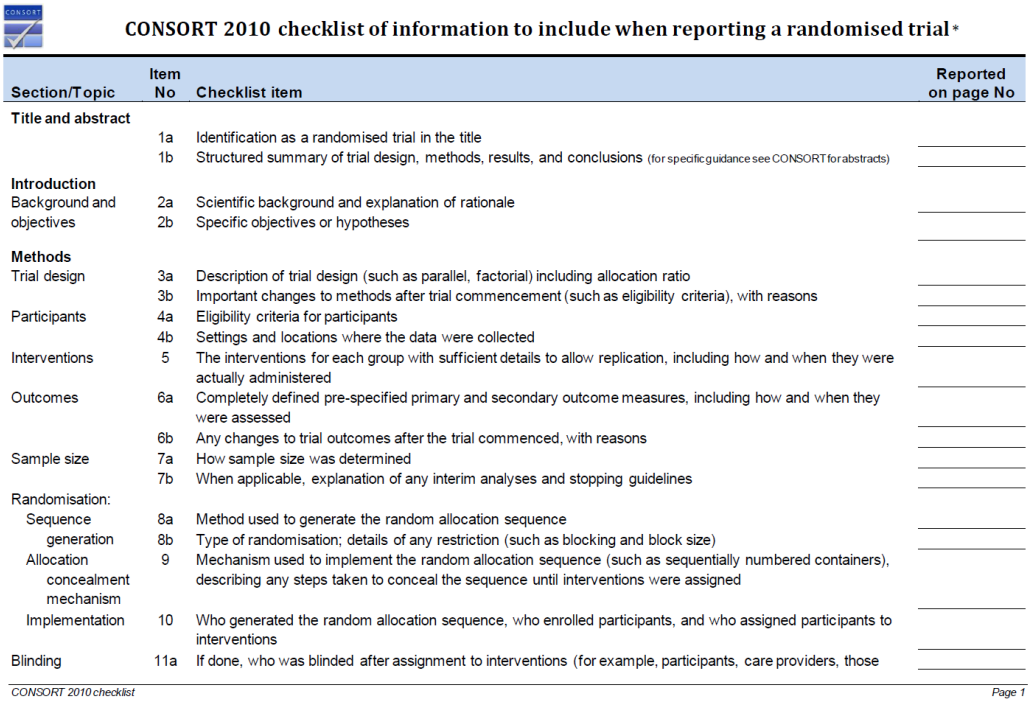
The second element of the framework is a set of standards for assessing the quality of evaluation activities, organized into four groups. The standards for Effective Evaluation are:

- Utility
  - Serve the information needs of intended users.
- Feasibility
  - Be realistic, prudent, diplomatic, and frugal.
- Propriety
  - Behave legally, ethically, and with regard for the welfare of those involved and those affected.
- Accuracy
  - Reveal and convey technically accurate information.

The framework is purposefully general and thus provides a guide for designing and conducting evaluation projects across many different program areas.

5.1.6 CONSORT and STROBE statements

In the field of clinical and epidemiological research, there have been several initiatives to standardize the conduction and reporting of studies. Two of them are the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) and the Consolidated Standards of Reporting Trials (CONSORT) initiatives (Altman et al., 2001; Von Elm et al., 2007). Both provide a set of recommendations for the reporting of studies. They offer a standard way for authors to prepare reports of study findings, facilitating their complete and transparent reporting, and aiding their critical appraisal and interpretation. Checklists that focus on reporting how the trial was designed, analysed, and interpreted are available for several types of study designs, e.g., randomized controlled trials, case-control studies, cohort studies, and cross-sectional studies (Figure 11). They emphasize the importance of transparency in the reporting to enable critical judgement of the generalizability and possible bias. These initiatives focus mainly on the reporting of research and do not provide guidelines for the planning and implementation of data collection. However, the checklists can also serve as guidelines of what to consider in the planning of a study.



CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Participants	4a	Eligibility criteria for participants	
	4b	Settings and locations where the data were collected	
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	7a	How sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	

CONSORT 2010 checklist Page 1

Figure 11. *Example page from the CONSORT checklist*

5.1.7 Conclusions

None of the reviewed frameworks provide a perfect fit for the type of evaluation planned within the PANACEA project. Some of the reviewed frameworks are more suitable for research projects, driven by research questions, and other frameworks are suitable for innovation projects focusing on evaluating a technical solution. As the PANACEA project is a research and innovation action, it has a combined need.

5.2 PANACEA evaluation framework

The framework developed within the PANACEA project incorporates components from several of the frameworks reviewed above. The FESTA methodology was used as the foundation and the various steps in the evaluation process were adapted to suit the purpose of the PANACEA project. The development of the PANACEA solution is an iterative process where results from WP6 data collections are fed back to WP3, WP4, and WP5 to refine the solution before the final evaluation (Figure 12). Technical validation of the systems used in the data collections will be performed before the start of each data collection. The results will be fed back to the relevant activity in responsible for the development or integration of the technology. Any issues discovered will be resolved before proceeding with the evaluation process. The results of the technical validations will be reported in milestones M15, M16, and M17. Results from the simulator and roadside data collections will be utilized to refine the algorithms developed in WP3. The conduction of A6.2 will happen in close collaboration with respective WP3 teams. The PANACEA solution validation and assessment pilots (A6.3) will conduct the validation tests to assess the readiness of the CHTs in collaboration with WP4 prior to the final evaluation at the pilot sites. In contrast to the technical validation, this validation will focus on the performance of the full PANACEA solution in operation, not the performance of individual sensors or parts. The collected data will be used to improve the technologies and their integration to CHTs and resolve any technology issues. Furthermore, the CHTs' assessment pilots will be also organised, monitored and executed in A6.3, to provide data for the final evaluation and impact assessment of the PANACEA solution. Activity A6.4 is about the realisation of the countermeasures' pilots. The evaluation of both the content and the actual online coaching system will be performed at the three pilot sites, in parallel with the A6.3 studies. The data collected will be fed back to WP5, to further improve the system.

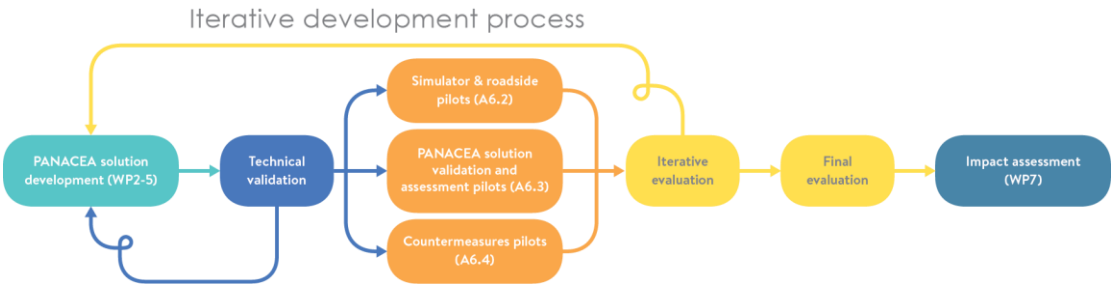


Figure 12. *PANACEA Iterative development process.*

The various data collections in WP6 used for the iterative development and for the final evaluation and impact assessment will follow the methodology of the PANACEA evaluation framework (Figure 13). The process is divided into three phases; planning, implementation, and analysis and reporting. Each box in Figure 13 represents a step to follow in the evaluation process. The steps are described as sequential steps in a linear way, where each step provides the necessary input for completion of the next step and the arrows show the dependencies between different steps. However, there might be a need to perform several steps in iteration during the process. As an example, there might be a need to revisit and adjust the study design after setting up the data analysis plan if it is discovered that different types of data are needed.

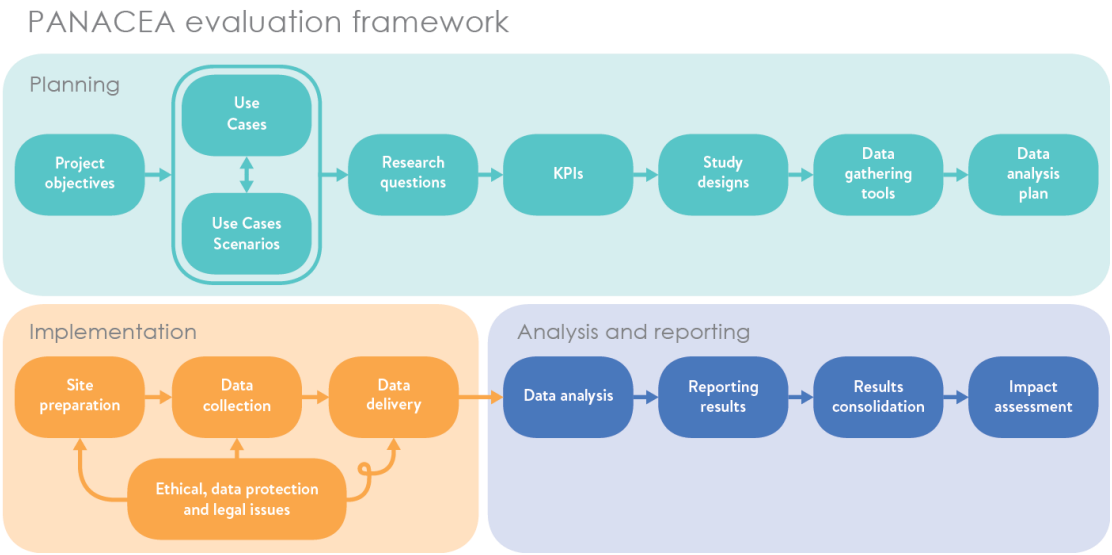


Figure 13. *PANACEA Evaluation Framework.*

The steps of the PANACEA framework are explained in the chapters below. Each step has its own chapter with a general description and an overview of how this will be implemented in the PANACEA project. In this version of the deliverable, the focus will be on setting the framework and describing the work process. More detailed evaluation protocols will be included in the update of the deliverable (D6.2). The detailed experimental plans with descriptions of how to carry out the data collections at the sites will be included in the annex of D6.2. In the current deliverable, an overview of the planned data collections at each site is presented.

Data collection in the first simulator study (UCA-S) started in month nine (M9) of the project (January 2022). The remaining A6.2 simulator studies, connected to UCB, will be conducted during the autumn 2022. Roadside assessments will be performed in two separate data collections, one during the autumn 2022 and one during the spring of 2023. Real-world and semi-real-world studies performed within A6.3 and A6.4 will follow thereafter. The main data collections used for the final evaluations will be performed between January and August 2023 (M21-M28). Preparations and baseline assessments will in some cases start earlier (UCA-R) and analyses and results consolidation will continue until M32 (December 2023). An overview of the timeline for all planned studies is presented in Figure 14.

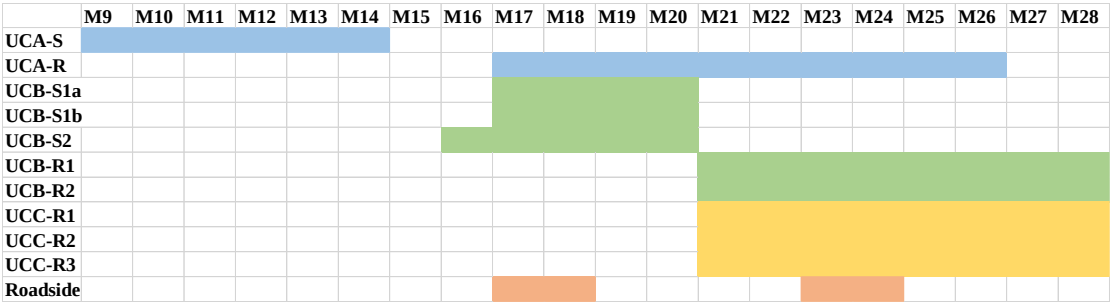


Figure 14. *Timeline of WP6 data collections.*

6 Use case scenarios

As defined in D1.1, the **Use Case Scenario** is a sequence of interactions happening under certain conditions, to achieve the primary actor's goal, and having a particular result with respect to that goal. The main purpose of use case scenario is to present in a detailed and clear and easy-to-learn way, the functional requirements of a system.

The following table presents the matching between the UC scenarios and scripts, as described in D1.1 and their connection with the Use Cases. Most of the UC scenarios apply to all UCs, because their implementation is horizonal. Those that target the technologies (CHTs; first column) do not apply to all UCs. Please refer to D1.1: 'Use Cases' for detailed descriptions of the Use Case scenarios.

Table 5. *Matching between Use Cases (UC) and Use Case Scenarios (UCS) or Use Case scripts (UCscr).*

CHTs and Technologies	UCs	Working shift flow	UCs	Administration, backend, and actors-oriented UC scripts	UCs
UCS01: FitDrive (Primary) – DATIK	**All**	UCS12: Baseline assessments	**All**	AII.1 UCscr17: Operators	**All**
UCS02: Alcohol sensor (Primary)– SENSEAIR	**All**	UCS13: Pre-Driving Assessment (incl. on-site) (ONPDA)	**All**	AII.2 UCscr18: Technology/ Service provider	**All**
UCS03: (Il)Licit drugs biosensor (Primary)- LEITAT	**All**	UCS14: During Driving Assessment (DDA)	**All**	AII.3 UCscr19: WP5 Development Team Countermeasures' specialist (responsible for the content of CCS)	**WP5/ outside UCs**
UCS04: - Smart Pulse Wave Analysis (PWA) device – AIT	**UCA/UCB**	UCS15: Roadside Assessment (RSA)	**All (but tested only in Norway)**	AII.4 UCscr20: Enforcer	**Norway / outside UCs**
UCS05: Steering wheel angle algorithm (SWA) and vehicle parameters (Primary)- ViF	**UCB**	UCS16: off duty Assessment (ODA)	**All**	AII.5 UCscr21: Administrator	**All**

CHTs and Technologies	UCs	Working shift flow	UCs	Administration, backend, and actors-oriented UC scripts	UCs
UCS06: DBL index (Secondary) - DBL	**UCB**			AII.6 UCscr22: Business rules	**All**
UCS07: BACtrack Skyn (Secondary) – VTI and CERTH	**UCB / UCA**			AII.7 UCscr23: General actor registration/ authentication/ login (with failures) and creation of profile	**All**
UCS08: Fitbit wrist band (Secondary) – VTI	**UCA**			AII.8 UCscr24: Feedback module	**All**
UCS09: Biomathematical model (BMM; Primary)– VTI	**UCA**			AII.9 UCscr25: Communication module among core actors (optional)	**All**
UCS10: Optalert and GSR system (Secondary) – CERTH	**UCB**			AII. 10 UCscr26: Errors (as exceptions) handling (closely related to UC20 and this a system and not a business UC scenario- Diagnosis procedures)	**All**
UCS11: Cloud based Countermeasures ' system (Primary) – CTLup	**All**				

7 Research questions

The research questions of the PANACEA project are related to the impact of the final PANACEA solution and to the development of specific technologies. The research questions were derived both from the Use Case Scenarios developed in WP1 (bottom-up approach) and by identifying the most relevant impact areas related to the overarching project objectives (top-down approach).

As part of activity in A2.5 in WP2, all PANACEA partners were asked to list question(s) that are of interest to them from their organisation's point-of-view, from their WP(s)' point-of-view, and from what they know would be important towards improving the health of professional transport workers.

To enable us to process the collected questions into research questions in WP6, a number of criteria of what make a good research question were defined. A good research question (RQ) must be clear, not too broad, and feasible to do within project time and budget. Further, a good RQ requires research and analysis to answer, and is of interest to partners and traffic safety community and useful for e.g., professional transport workers and community. Last but not least the RQ must be measurable.

With the general criteria of a good RQ in mind, we added several criteria related to the PANACEA project based on what was presented in the Grant Agreement, the sensors used in the project, etc.

With the general criteria and PANACEA specific criteria set, the selection and revision process began. The questions that were not clear, too broad, or not feasible do within the project time were not included for further process. The questions that were processed further, were checked by several people and reformulated (if necessary) to make them clear. They were grouped into four different categories related to the overall project objectives: validation of CHTs and technologies, evaluation of CHTs, evaluation of countermeasures, and impact. The RQs are also connected to the project KPIs. The short-listed RQs were then discussed and refined further in a workshop at the 4th plenary meeting in Greece.

The final set of RQs consist of 39 research questions. Only the high-level RQs that are relevant for all the UCs are presented in Table 6 below. Specific research questions for each study can be found in the complete list of RQs in Appendix II.

Table 6. *High-level research questions (RQ) and their connection to KPIs and data gathering tools.*

Project objective	RQ Category	High-level RQ	Tentative KPIs	Data gathering tool
OBJ2	Validation of CHT and technologies	Do the PANACEA sensors/systems detect targeted driver impairments effectively with high sensitivity and specificity?	KPI 3.1 Reliability of CHT, 3.2 Specificity of CHT, 3.3 Sensitivity of CHT, KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from sensors, subjective ratings of impairment
OBJ2	Validation of CHT and technologies	How is the performance of the PANACEA sensors compared to a reference measurement?	KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from PANACEA sensors & reference equipment

Project objective	RQ Category	High-level RQ	Tentative KPIs	Data gathering tool
OBJ2	Validation of CHT and technologies	Do the combined sensors improve driver state detection?	KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from sensors (individual and combined)
OBJ2	Validation of CHT and technologies	Does the PANACEA integrated solution work in a real-life setting to detect impairment and deliver countermeasures?	KPI 3 Technical performance of CHT, KPI 3.1 Reliability of CHT, 3.2 Specificity of CHT, 3.3 Sensitivity of CHT	Data from PANACEA solution and subjective ratings of impairment
OBJ3	Evaluation of CHTs	Are the PANACEA sensors/systems accepted by the users?	KPI 3.7 Acceptance of CHT	Questionnaires
OBJ3	Evaluation of CHTs	Are the CHTs perceived as useful, satisfying, trustworthy, and easy to use?	KPI 3.4 Ease to use CHT, 3.5 Usefulness of CHT, 3.6 Willingness to use CHT	Questionnaires
OBJ4	Evaluation of countermeasures	What are the immediate effects of implemented countermeasures?	KPI 4.2 Effectiveness of a countermeasure	Questionnaires and data from PANACEA solution
OBJ4	Evaluation of countermeasures	Is the PANACEA countermeasures system accepted by the users?	KPI 4.1 Acceptance of a countermeasure, 4.3 Satisfaction	Questionnaire
OBJ7	Impact	Does behaviour change/improve after the relevant countermeasure has been administered?	KPI 4.2 Effectiveness of a countermeasure, KPI10 CEA/CBA	Questionnaires and data from PANACEA solution
OBJ7	Impact	Will the PANACEA countermeasures reduce driver impairment and improve the driver performance?	KPI 4.2 Effectiveness of a countermeasure, KPI 8.1 N of saved lives (ON ROADS), KPI8.2 N of saved lives (OFF ROADS), KPI9 QoL	Questionnaires, data from PANACEA solution, and driving performance data from vehicles
OBJ7	Impact	Would it be possible to implement the PANACEA system in regular operation?	KPI 10 CEA/CBA	Focus group with different stakeholders
OBJ7	Impact	Does the PANACEA system increase perceived (drivers) and reported (operators) safety?	KPI 1.1 Perceived (drivers) safety, KPI 1.2, Reported (operators) safety	Questionnaire, focus group and data from PANACEA solution

8 Key Performance Indicators

In collaboration with the partners, the following table was defined including the preliminary list of KPIs extracted from Relevant Impact targets from the Description of Action with a preliminary KPI definition and Type associated.

Table 7. *KPIs extracted from the Description of Action (DoA).*

KPI_ID	Name of KPI	Relevant Impact targets (extracted from DoA)	KPI definition	Type
KPI 1.1	Perceived (drivers) safety	Perceived (drivers) and safety increases by 8%,	qualitative indicator based on questionnaire and measured through a Likert scale (e.g., 1 - 5). comportments connected to perceived safety	Impact
KPI 1.2	Reported (operators) safety	Reported (operators) safety increases by 8%,	KPI based on observed events related to safety in a vehicle fleet (e.g., number of severe braking events) before and after implementation of a tool	Impact
KPI 3	Technical performance of CHT	Increase capacity for monitoring Fitness to Drive assessment by 30%	Involves a technique of predicting the future value of a key technical performance parameter of the higher-level end product under development based on current assessments of products lower in the system structure.	Technology
KPI 3.1	Reliability of CHT	Increase capacity for monitoring Fitness to Drive assessment by 30%. Key Outcomes: reliability, of all sensors/modules/subsystems >25% against relevant SoA.		Technology
KPI 3.2	Specificity of CHT	Increase capacity for monitoring Fitness to Drive assessment by 30%. Key Outcomes: specificity of all sensors/modules/subsystems >25% against relevant SoA.	Refers to the probability of a negative test, conditioned on truly being negative	Technology

KPI_ID	Name of KPI	Relevant Impact targets (extracted from DoA)	KPI definition	Type
KPI 3.3	Sensitivity of CHT	Increase capacity for monitoring Fitness to Drive assessment by 30%. Key Outcomes: sensitivity of all sensors/modules/subsystems >25% against relevant SoA.	Refers to the probability of a positive test, conditioned on truly being positive	Technology
KPI 8.1	N of saved lives (ON ROADS)	Saves lives on and off road (8%)	Quantitative indicator based on Number of road fatalities in a given time period	Impact
KPI 8.2	N of saved lives (OFF ROADS)	Saves lives on and off road (8%)	Quantitative indicator based on Number of alcohol/stress/drug related off road fatalities in a given time period	Impact
KPI 9	QoL	QoL is estimated to increase by at least 2 points in Quality of Life in Years (QALY).	Quality Adjusted Life Years	Impact

Another table was constructed for those KPIs that are to be discussed with the partners further. The final KPIs related to WP6 data collections will be presented in the next version of the deliverable (D6.2)

Table 8. *Additional KPIs.*

KPI_ID	Name of KPI	Relevant project objective	Type
KPI 3.4	Ease to use the CHT	OBJ3	Impact
KPI 3.5	Usefulness of CHT	OBJ3	Impact
KPI 3.6	Willingness to use CHT	OBJ3	Impact
KPI 3.7	Acceptance of CHT	OBJ3	Impact
KPI 3.8	Trust in CHT	OBJ3	Impact
KPI 4.1	Acceptance of a countermeasure	OBJ4	Countermeasure
KPI 4.2	Effectiveness of a countermeasure	OBJ4	Countermeasure
KPI 4.3	Satisfaction	OBJ4	Countermeasure

KPI_ID	Name of KPI	Relevant project objective	Type
KPI 5	Adaptable countermeasure	OBJ4	Countermeasure
KPI 6	Sensitivity and specificity of a sensor or combination of technologies	OBJ2	Technology
KPI 10	CEA / CBA	OBJ7	Impact

9 Study design

There will be a variety of study designs in the PANACEA project, depending on the objectives of each data collection. Most of them will use a within-subjects design and common for all data collections is that they will have a control condition serving as a baseline for the validations and evaluation. The simulator and roadside studies are quite diverse with study designs tailored to fit the specific research questions connected to the study. The validation and assessment pilots and countermeasures' pilots are based on a repeated measures design, where the PANACEA system will be used repeatedly by the participating commercial drivers. In this chapter, an overview of the different study designs is presented. Detailed experimental plans for each study will be included in the next version of the deliverable (D6.2) and the template for the experimental plans can be found in appendix I.

9.1 Simulator studies (A6.2) and roadside assessment

The simulator and roadside studies will collect data to improve and/or create the WP3 algorithms and to improve and define the thresholds for each impairing state addressed. Templates will be included in the next version of this deliverable (D6.2) for sharing the main outcomes or needs for improvements with WP2 and WP3.

9.1.1 UCA-S

The specific aims of the UCA simulator study are to learn more about how moderate amounts of alcohol in the evening affects night sleep and next day driving performance and to develop a first version of a biomathematical model of fatigue that takes next-day effects of alcohol into account. The study is performed in a driving simulator and driver impairment is manipulated by experimenter-controlled administration of alcohol (target 0.05%). The study has a within-subject mixed-model design with a factor for next-day effects (driving with alcohol intake the day before versus driving without alcohol the day before) and a factor for time (in the morning and in the forenoon the day after). The experiment is carried out with 30 drivers who visit the lab three times, always in the same order.

1. Evening visit, 2 drives; one training drive and one drunk driving
2. Morning visit the day after the first visit, 2 drives
3. Morning visit without alcohol in the evening (baseline), 2 drives

Each drive in the car simulator includes 25 min rural road and 10 min urban road. Sleep is tracked off-site by diaries and wearables. Subjective sleepiness, objective fatigue indicators, and simulator data is collected during the drive. BrAc, attention and stress level are measured before and after each drive.

9.1.2 UCB-S1

Fatigue, alcohol consumption and stress will be addressed in the UCB simulator pilots in Thessaloniki, Greece. 20 taxi drivers and 20 delivery service riders will participate in simulator tests in a car and PTW simulator. A repeated within-participants design is applied with baseline measurements collected at the first session. The drivers will participate in three counterbalanced sessions, one before their shift starts, one after their shift ends and one where they arrive at the middle of their shift. Fatigue is assumed to increase from the start of the shift to the end of the shift. Stress is manipulated through events in the simulator scenarios. Alcohol will be manipulated through experimenter-controlled administration with four target levels in three sessions (0, 0.02%, 0.05%, >.05%).

Fatigue will be measured before the session, after the session and continuously using KSS. Stress will be measured before and after the session, and continuously during the drive through Galvanic Skin Response (GSR). BrAc will be measured before and after each drive. Fatigue and stress scales will be administered before and after the session and stress will also be measured after events. Each impairment state is measured by the PANACEA technologies and a reference technology.

Table 9. *UCB – S1 design and procedure*

Part of session	Time
Informed consent	-20 mins
Briefing and ethical rights	-5 mins
BASELINE & pre-shift (1st session)	**0 mins**
Pre-questionnaire completion on fatigue, stress and alcohol use.	10 mins
Driving/ Riding simulator familiarization	5 mins (only during their first session; sessions will be counterbalanced)
Fatigue, stress, alcohol baseline measurements (this includes 0% level alcohol) are taken.	30 mins (including 10 mins setting up and measurement collection) and collection with both reference and PANACEA technologies and 20 mins driving/ riding simulator.
Alcohol consumption (0.02%)	20 mins
Post questionnaire completion on fatigue, stress and alcohol state. Incl. some question items on the technologies (in the first session).	15 mins
Checking data collection status and quality	5 mins (in parallel with debriefing)
Debriefing	5 mins
During Driving/ Riding (2nd session)	**0 mins**
Pre-questionnaire completion on fatigue, stress.	10 mins
Driving/ Riding simulator familiarization	5 mins

Part of session	Time
Simulator fatigue driving/ riding scenario	20 mins
Post question completion on fatigue, stress	10 mins
Simulator stress driving/ riding scenario	20 mins
Post question completion on fatigue, stress	20 mins
Debriefing	5 mins
Post- shift (3rd session)	**0 mins**
Pre-questionnaire completion on fatigue, stress and alcohol	10 mins
Driving/ Riding simulator familiarization	5 mins
Simulator fatigue driving/ riding scenario	15 mins
Post question completion on fatigue, stress and alcohol	10 mins
Simulator stress driving/ riding scenario	15 mins
Post question completion on fatigue, stress	10 mins
Simulator alcohol (>0.05%) driving/ riding scenario	20 mins
Post questionnaire completion on fatigue, stress and alcohol	10 mins
Checking data collection status and quality	5 mins (in parallel with debriefing)
Debriefing	5 mins

9.1.3 UCB-S2

The study will be realized as permutated within-subjects design with two independent variables: (1) the *kind of driving environment*: city vs. highway, and (2) *kind of driver distraction*: no distraction vs. cognitive vs. visual/manual vs. cognitive/visual/manual. The different kinds of driver distraction will be induced by different secondary tasks that the driver needs to perform in permutated order during the drive.

As dependent variables, different parameters will be measured to capture the behaviour and state of a driver (see Figure 15 for an overview). Primarily, the focus will be on parameters capturing gazing behaviour (e.g., temporal gaze variance, gaze off road), driving behaviour (e.g., steering wheel angle, SD headway, whether the hand(s) are on/off the steering wheel, stress, and cognitive load. In addition, subjective measures such as perceived distraction or stress will be captured after each drive.

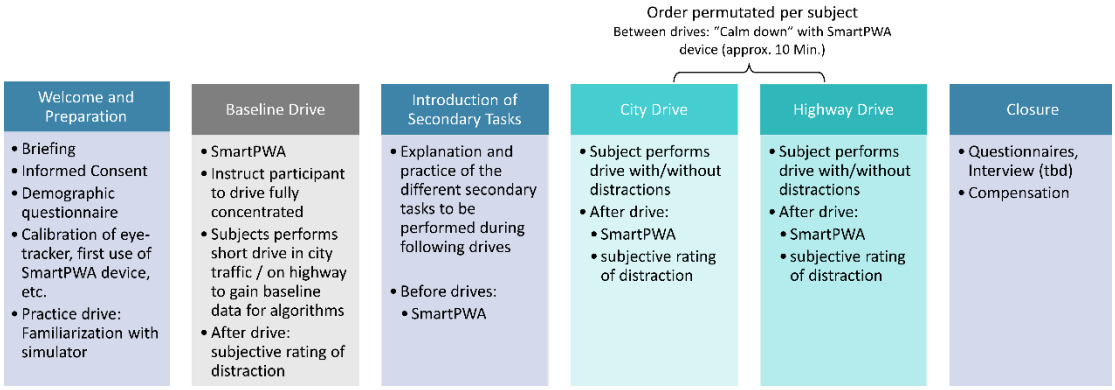


Figure 15. *Planned study procedure for the VIF simulator study*

9.1.4 Roadside

The roadside assessments for validation of the PANACEA roadside sensors will be performed in two separate data collections, one for the validation of the Senseair Go portable alcohol sensor and one for the validation of the Leitat biosensor. Testing will be done according to the regular operations of the traffic police in Norway, only adding the PANACEA sensors as an additional step in the testing procedure. The additional testing with PANACEA sensors will be optional for the drivers being stopped at the roadside. For each data collection, testing will be done for one month, with the target of reaching 20 positive and 20 negative samples for alcohol and drugs, respectively. The procedure for alcohol testing is shown in Figure 16.

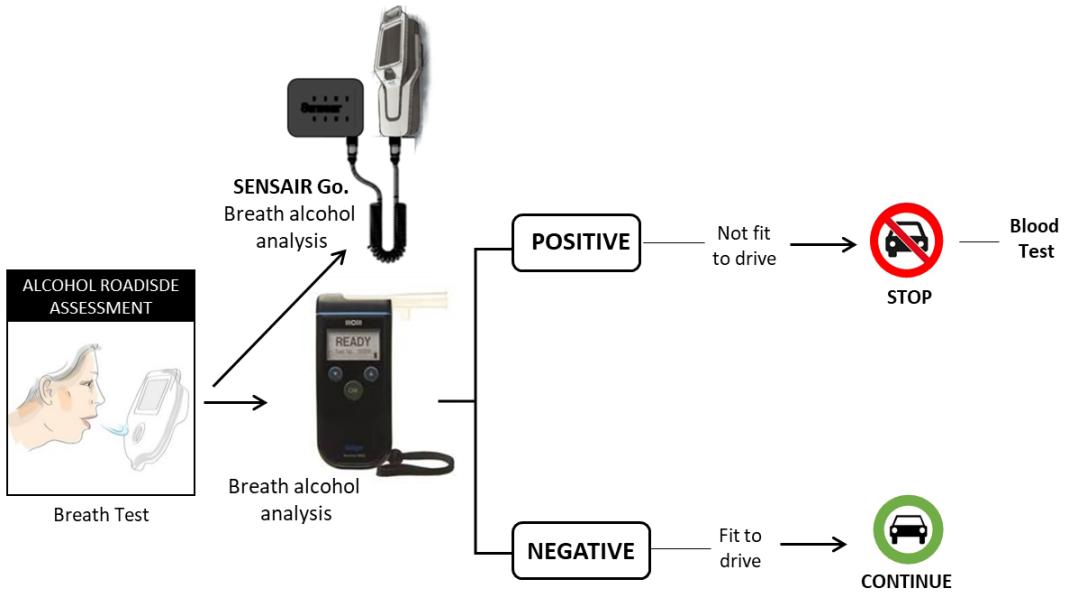


Figure 16. *Alcohol testing in the roadside study.*

Drug testing will follow a different approach, as described in Figure 17, since there is a greater need to check for false positive and false negative tests.

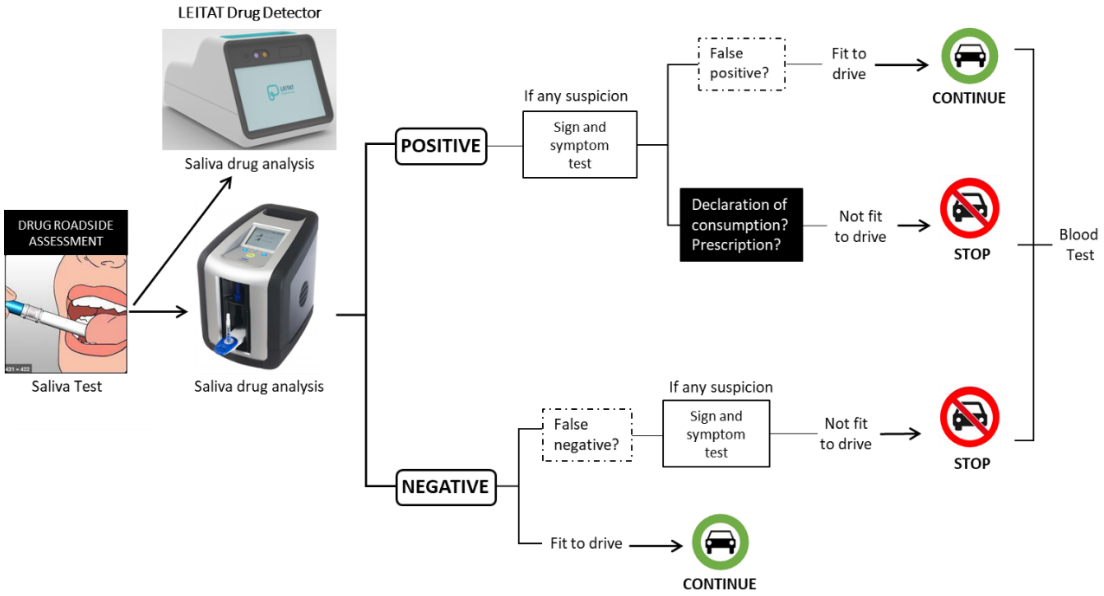


Figure 17. *Procedure for drug testing in Norway.*

9.2 Validation and assessment pilots (A6.3) and countermeasures' pilots (A6.4)

The real-world and semi-real-world studies UCA-R, UCB-R and UCC-R have the combined purpose of collecting data for validation and assessment of the CHTs (A6.3) and for evaluation of the countermeasures (A6.4). The CHTs' assessment pilots are based on a repeated measures design where the PANACEA solution will be evaluated on repeated occasions (at least 3 repetition per CHT). This is part of the iterative process, serving the feedback loop to WP4 and WP5. Templates will be included in the next version of this deliverable (D6.2) for sharing the main outcomes or needs for improvements with WP4 and WP5. The short-term and immediate countermeasures will be evaluated in the pilots running in parallel with the A6.3 studies. The evaluation of longer-term countermeasures and training content will be performed in dedicated focus groups (at least two per pilot site) with both drivers (or riders) and operators. The data collected will be fed back to WP5, to further improve the system.

9.2.1 UCA-R

In UCA, data collections will be done during the normal operation of the autonomous shuttles in Linköping with 8 safety drivers participating. A within-subjects design will be used with before and after measurements. Data collections will be done continuously for two 1-month periods, ensuring that all safety drivers will use the PANACEA solution during several work shifts. A baseline assessment will be done in October 2022 with "passive sensors", ideally collecting data with BACtrack Skyn, Fitbit and Datik but without countermeasures or other feedback to the drivers. During the baseline assessment, the sensors will not be connected to the PANACEA platform. The data collection for final evaluation with the full PANACEA solution activated, including the countermeasures system, will be performed for one month in March 2023.

9.2.2 UCB-R

40 drivers and riders will participate in the semi-real life evaluation phase. Fatigue, stress, and distraction will be evaluated with instrumented vehicles (shown in Figure 5), but for ethical and legal reasons, alcohol and drugs will be tested in the passenger car and motorcycle simulators (Figure 3). The design and procedure of the tests will be similar to the one for the simulator tests (see Table 9) with participants arriving to participate in three counterbalanced sessions.

Alcohol will be administered to the four levels, as in the simulator studies unless another distinction is requested by WP3 teams. Similarly, diazepam and methadone dosages will be based on WP3 final decisions (e.g., levels, thresholds, relation to countermeasures), and they will not be administered by an affiliated psychiatrist. A health care professional will always be present during testing when alcohol and drugs are administered. Ethical approval will be obtained prior any testing takes place. Data collection will be conducted in June 2023.

9.2.3 UCC-R

The study design will be the same for all three driver groups (R1-garbage truck drivers, R2-interurban bus drivers, and R3-long distance bus drivers). Approximately a total of 15-20 drivers will participate, counting the 3 demonstrators and sites. Data collections will be done during the normal operation of the garbage trucks and bus services. The period for testing will be 3 months, including 1-month baseline with “passive” sensors and 2 months with the full PANACEA solution, i.e., with all sensors and displays and countermeasures.

10 Data gathering tools

Several different types of data gathering tools will be used in the project. They include both subjective and objective tools to make sure the individual studies can answer their specific research questions and to provide good quality data for the impact assessment.

10.1 Objective data

The PANACEA sensors and technologies will be the main data gathering tools providing objective measurements of driver impairments in all data collections. Detailed descriptions of the technologies and their respective output parameters can be found in deliverable D3.1: 'Methodologies for a holistic fitness to drive assessment'. Instructions on how to carry out measurements off-duty, on-duty, on-site and roadside are available to the pilot sites in the internal deliverables ID 3.1: 'Off-duty assessment: Measures and Thresholds' ID3.2: 'On-duty assessment: Measures and Thresholds', ID3.3: 'On-site assessment: Measures and Thresholds', and ID3.4: 'Roadside assessment: Measures and Thresholds'. The terms off-duty, on-duty, and on-site describe the different work shift phases for professional drivers and these terms are relevant for the final evaluation of the PANACEA solution in the operational setting. For the simulator studies, these correspond to measurements taken off-site, during driving, and on-site. Below is an overview of PANACEA technologies used as data gathering tools per work shift phase and study.

Table 10. *Objective data collection tools used in the various work shift phases in the studies. DDA=during driving assessment (UCS14), ODA=off duty assessment (UCS16), ONPDA=on site & pre-driving assessment (UCS13), RSA=roadside assessment (UCS15).*

Sensor or technology	Output	ODA	ONPDA	DDA	RSA
DATIK FitDrive (UCS01)	Fatigue level, detected events			UCA-R, UCB-S1, UCB-R, UCC-R	
DATIK pre-questionnaire (UCS01)	Fatigue risk level		UCA-R, UCB-S1, UCB-R, UCC-R		
Senseair Go (UCS02)	Breath alcohol content (BrAc)			UCB-S1, UCB-R, UCC-R	
Senseair Go Portable (UCS02)					Roadside
Senseair Wall (UCS02)	Breath alcohol content (BrAc)		UCA-R, UCB-S1, UCB-R, UCC-R		

Sensor or technology	Output	ODA	ONPDA	DDA	RSA
Leitat biosensor (UCS03)	Benzodiazepines and methadone concentration in saliva		UCA-R, UCB-R, UCC-R		Roadside
AIT Smart PWA (UCS04)	Stress, fatigue, and cognitive load	UCB-R, UCC-R	UCA-S, UCA-R, UCB-S2	UCB-R, UCC-R	
ViF Driver Monitoring System (UCS05)	Cognitive distraction			UCB-R, UCC-R	
DBL index (UCS06)	Cognitive load and Stress		UCB-S2	UCB-S2	
BACtrack Skyn (UCS07)	Transdermal Alcohol Content (TAC)	UCA-R, UCB-R,	UCA-R, UCB-S1, UCB-R	UCA-R, UCB-S1, UCB-R	
Fitbit (UCS08)	Activity, sleep/wake patterns and sleep stages	UCA-S, UCA-R	UCA-R	UCA-R	
BMM (UCS09)	Fatigue level	UCA-R	UCA-R	UCA-R	
Optalert (UCS10)	Fatigue level			UCB-S1	
GSR (UCS10)	Arousal (Skin conductance)			UCB-S1	

10.1.1 Output from PANACEA solution/platform

In addition to the measurements obtained from the various PANACEA sensors, the integrated PANACEA solution will enable collection of data regarding usage, impairment levels, triggered warnings, delivered countermeasures, statistics/ analytics (through dedicated dashboard) and engagement with the countermeasures' system. A preliminary data clustering was enclosed in D9.4 'Data Management Plan' (M6). The complete list of data types and characteristics along with any restrictions, embargo periods and open sharing possibilities will be annexed in D9.5 'Data Management Plan – an update' (M34). Likewise, the data available from the technologies, along with the agreed upon thresholds will be available in D3.1 'Methodologies for a holistic fitness to drive assessment' (M16) and the final decisions based on the A6.2

outcomes and PANACEA solution prototype will be included in D3.2 ‘Methodologies for a holistic fitness to drive assessment - an update’ (M24).

10.1.2 Reference sensors

In the simulator and roadside studies, reference sensors will be used to enable validation of individual PANACEA technologies in relevant contexts.

Reference sensors in UCA-S are Smart Eye Pro which is a 4-camera remote eye tracking system and Vitaport 3 that measures Electrocardiography (ECG) and vertical Electrooculography (EOG) continuously during the drive. Both reference equipments enable measurement of fatigue/sleepiness indicators. A Dräger 6820 breathalyzer will be used to measure BrAc. In addition, a Psychomotor Vigilance Task will be used as a measure of alertness.

UCB-S1 will use reference technologies for fatigue via measurements of Electroencephalography (EEG) and ECG, for stress via measurement of ECG and for BrAc using a breathalyzer (standard equipment used by the police force).

In the UCB-S2 study, a SmartEye eye-tracking system will be used as a reference equipment for cognitive distraction using the parameters temporal gaze variance, gaze off road (AttenD), gaze variance on road, blink-rate, and fixation duration.

The roadside study will use a Dräger 6820 and 6810 breathalyzers as the reference equipment for BrAc and a Dräger DrugTest5000 for benzodiazepines and methadone or Securetec´s WipeAlyser in combination with DrugWipe® for benzodiazepines. Drug testing in blood samples will be done according to regular procedures used by the police force in Norway.

10.1.3 Vehicle data

To enable evaluation of the effectiveness of countermeasures and driver impairments on driving performance, vehicle and simulator data will be collected. In the UCA-S study, simulator data will be logged continuously during the drive including speed and speed variability, lane position and steering, surrounding traffic, including time headway and time to collision. In UCB-S1a data will be logged in the driving/ riding simulators about steering wheel angle, speed, lane position and headway variability along with braking activation and number of events. The UCB-S2 study will log Steering Wheel Angle, SD Headway, SD Lateral Position, and SD Speed from the driving simulator. A camera will also be installed to assess hand off wheel.

In UCA-R, shuttle data will be logged including: % automation activated, % hard brakings/ jerk, number of other road user interactions, number of passengers. For UCB-R both the simulators and instrumented vehicles will be used. Data logged will be the same as in the simulator study. In the instrumented vehicle, the data will be collected through the CANbus. UCC-R will log vehicle data including speed, acceleration, and lane position through the CANbus of the buses and garbage trucks. Parameters such as speeding, high RPM, harsh braking, excessive idle, and harsh acceleration are generated from the vehicle data.

10.2 Subjective data

Self-reported measures like questionnaires, rating scales, and focus groups will be used in the evaluation of the PANACEA system. A user profile will be included in the PANACEA solution with basic information about each driver or operator. The collected information will be the same for all final evaluation studies (A6.3 and A6.4). The user profile includes information about age, gender, profession, medical conditions etc. and a first version was described in Appendix IV of deliverable D1.1: ‘Use Cases’. Common before (background) and after

questionnaires will be used in the final evaluations. In addition, the countermeasure system has built-in evaluation questions as described in D5.1: 'Countermeasures for drivers, operators, and enforcement. Content of the cloud-based coaching and support system'. These are for example quick evaluation questions like *Was this useful?* that are completed by the user after receiving a countermeasure.

10.2.1 Questionnaires

Questionnaires will be used to capture both background data of the participants (e.g., demographics) in each data collection, to track subjective experiences of the various driver impairments, and to evaluate acceptance, trust, usability, quality of life and other measures needed for evaluation and impact assessment. When available, validated questionnaire instruments will be used.

In the simulator studies, study specific background questionnaires will be used, comprising questions of relevance for the data collection. These include demographics, questions about the impairment states targeted in the study and other questions of relevance for the data analysis. During the trials, questions about impairment level (acute stress, sleepiness, intoxication etc.) will be used to follow the development of driver state over time. Self-assessments of driving quality will also be included in UCA-S1 and UCB-S1 trials. The roadside study will have a questionnaire to the police officers asking about the efficiency and usefulness of the PANACEA sensors for roadside assessment.

The questionnaires used in the final real-world and semi-real-world evaluations are the same across studies to enable comparisons between sites and to provide harmonized data for the impact assessment. The full before and after questionnaires will be completed by the professional drivers participating in the trials whereas a subset of questions will be completed by operators/managers. Additional questions can be added by the sites depending on the specific research questions addressed in the UC. The common before questionnaire includes the EQ-5D instrument for assessment of Quality-of-Life (QoL), the Alcohol Use Disorders Identification Test (AUDIT), the Karolinska Sleep Questionnaire (KSQ), and questions about drug use and stress symptoms. In the after questionnaire, instruments needed for evaluation of acceptance, trust, usability, safety, and willingness to have the PANACEA solution are included. The after questionnaire comprises the same questions as the before questionnaire as well as the Technology Acceptance Questionnaire (TAQ), System Usability Scale (SUS), SHAPE Automation Trust Index (SATI), and questions about willingness to have/use/buy and perceived safety. The suggested questionnaire tools are tentative and the final before and after questionnaires will be included in the update of this deliverable (D6.2). An overview of questionnaire instruments per study is shown in Table 11. The suggested questionnaire tools are included in Appendix III. They were selected because they are well-established, validated across different EU countries and commonly used in transportation research.

Table 11. *Questionnaire instruments.*

Measure	Name of instrument	Output	Reference	Administration	Study
Sleepiness	Karolinska sleepiness scale (KSS)	Sleepiness score between 1 and 9	(Åkerstedt, Anund, Axelsson, & Kecklund, 2014)	Repeated measures DDA, ODA, ONPDA	UCA-S, UCA-R, UCB-S1, UCB-R, UCC-R
Sleep problems	Karolinska Sleep Questionnaire (KSQ)	Indices for: sleep quality, non-restorative sleep, sleep apnea, and sleepiness	(Nordin, Åkerstedt, & Nordin, 2013)	Before & after questionnaire	UCA-S, UCA-R, UCB-S1, UCB-R, UCC-R
Acute stress	VTI acute stress scale (VSS)	Stress score between 1 and 9	Not validated	Repeated measures DDA, ODA, ONPDA	UCA-S, UCA-R, UCB-S1, UCB-R, UCC-R
Stress symptoms	Perceived stress	Single item question on 5-point Likert scale	(Elo, Leppänen, & Jahkola, 2003)	Before & after questionnaire	UCA-R, UCB-R, UCC-R
Alcohol use	Alcohol Use Disorders Identification Test (AUDIT)	Score from 0 to 40	(Babor, Biddle-Higgins, Saunders, & Monteiro, 2001)	Before & after questionnaire	UCA-S, UCA-R, UCB-R, UCC-R
Acceptance	Technology Acceptance Questionnaire (TAQ)	Usefulness and satisfying scores ranging from -2 to +2	(Van Der Laan, Heino, & De Waard, 1997)	After questionnaire (focus on CHTs and countermeasures separately)	UCA-R, UCB-R, UCC-R, Roadside

Measure	Name of instrument	Output	Reference	Administration	Study
Usability	System Usability Scale (SUS)	Usability score from 0 to 100	(Brooke, 1996)	After questionnaire	UCA-R, UCB-R, UCC-R, Roadside
Trust	SHAPE Automation Trust Index (SATI)	Mean score from 0 to 6	(Dehn, 2008)	After questionnaire	UCA-R, UCB-R, UCC-R, Roadside
Quality of life	EQ-5D	EQ-5D index, EQ-5D VAS score from 0 to 100	(Balestroni & Bertolotti, 2015)	Before & after questionnaire	UCA-R, UCB-R, UCC-R

The drivers will also rate their level of impairment repeatedly during the test days to be able to follow the development of e.g., stress and sleepiness over time. Sleepiness will be measured with the Karolinska sleepiness scale (KSS), and stress with the VTI acute stress scale (VSS) as indicated in Table 11. Intoxication will be measured using the question *How intoxicated do you feel?* (0: completely sober; 10 very affected). In addition, to be able to track behavior related to the various driver impairments, drivers participating in the real-world studies will also complete diaries to track sleep, stress, alcohol consumption, and drug use.

Simulator studies UCA-S and UCB-S2 will also have self-assessments of driving quality before and after each drive in the simulators. The questions asked are *How well do you think you will drive?* (0: worst imaginable; 10 best imaginable), and *How well did you drive?* (0: worst imaginable; 10 best imaginable).

10.2.2 Focus groups

Focus groups with stakeholders will be performed in all final evaluation studies. The evaluation of longer-term countermeasures and training content will be performed in dedicated focus groups (at least two per pilot site) with both drivers (or riders) and operators. The data collected will be fed back to WP5, to further improve the system.

11 Data analysis plan

The UC teams are responsible for creating a data analysis plan for each data collection based on the study design and connected research questions. Repetitive data treatment will ensure collection of adequate volume and to reach the set KPIs and answer the research questions. After each repetition, data will be used to improve the technologies and their integration to CHTs (WP4) and resolve any technology issues. The general data analysis plan for simulator, roadside, and real-life studies has its starting point in the high-level research questions (Table 12).

Table 12. *General data analysis plan.*

High-level RQ	Analysis plan
Do the PANACEA sensors/systems detect targeted driver impairments effectively with high sensitivity and specificity?	Measure the number of correctly classified driver impairments according to the thresholds defined in WP3 when driver impairment level is known (by manipulation of driver state or via gold standard reference measurement of driver state).
How is the performance of the PANACEA sensors compared to a reference measurement?	Analysis of correlation between PANACEA sensor and reference sensor. Compare number of correctly classified driver impairments between PANACEA sensor and reference sensor.
Do the combined sensors improve driver state detection?	Compare number of correctly classified driver impairments between individual PANACEA sensors and combined sensors.
Does the PANACEA integrated solution work in a real-life setting to detect impairment and deliver countermeasures?	Analyze number of correctly classified driver impairments in real-life settings (compare with subjective rating of impairment) Analyze usage data from PANACEA solution.
Are the PANACEA sensors/systems accepted by the users?	Calculate scores for acceptance from questionnaires and compare with cut-offs or normal ranges for each instrument
Are the CHTs perceived as useful, satisfying, trustworthy, and easy to use?	Calculate scores for usability, satisfaction and ease-of-use from questionnaires and compare with cut-offs or normal ranges for each instrument
What are the immediate effects of implemented countermeasures?	Analyze difference in driver impairment lever before and after receiving a countermeasure.

High-level RQ	Analysis plan
Is the PANACEA countermeasures system accepted by the users?	Calculate scores for acceptance from questionnaires and compare with cut-offs or normal ranges for each instrument
Does behaviour change/improve after the relevant countermeasure has been administered?	Compare sleep habits, stress level, alcohol and drug use before and after receiving countermeasures.
Will the PANACEA countermeasures reduce driver impairment and improve the driver performance?	Analyze changes in driver impairment level and driving performance over time when the PANACEA solution is used.
Would it be possible to implement the PANACEA system in regular operation?	Analyze output from focus groups with stakeholders after they have experienced the PANACEA solution.
Does the PANACEA system increase perceived (drivers) and reported (operators) safety?	Analyze changes in driver impairment level and driving performance over time when the PANACEA solution is used. Analyze results from questionnaire about safety.

The UC teams should take potential risks of bias and threats to validity into consideration in the data analysis plan. This can be done by identifying potential confounding factors, risk of bias, and other interfering effects beforehand. Examples are carry-over effects, learning effects, drop-outs, timing of tests, incentives, and experimenter bias. These can be handled either by employing a study design that balances out potential risks of bias or by measuring these factors to be able to control for them in the statistical analyses.

12 Pilot site preparations

The teams located at the pilot sites will refine and operationalise the procedures as defined within A6.1. Each UC team is also responsible for obtaining Ethics approval, if needed, prior to any testing. For the studies to be conducted smoothly and without delays, preparations will go beyond what is described in this deliverable. Apart from the necessary technical equipment, the following aspects will be considered while preparing the data collections, if applicable to the study.

12.1 Ethics

PANACEA is a complex project with ethical issues related to security, privacy and interoperability. Each phase of the project will be addressed accordingly from the project concept development to the project closure.

Core ethical issues within PANACEA are related to:

- Data privacy protection, confidentiality, and transparency
- Informed consent
- Incidental findings
- Transparency of the collected data management by the PANACEA solution and during its WP6 pilots
- IT-Security and identity management
- Risk assessment (Insurance)
- Delegation of control
- Incentives (financial inducements, compensations, etc.)

Local Ethics Representatives will be the main contact point for any ethics related issues (e.g., submission of research/test protocols for approval by the Institutional/National Ethics Committees, GDPR issues, etc.) from the pilot site point of view. The Ethics Management Panel will tackle user involvement and ethical and data protection issues. In addition, one of the main tasks of the nominated persons will be to co-ordinate and be responsible for obtaining approval by the local/regional/institutional ethics committee before any pilot related activities take place (e.g., even before recruitment starts) - if needed. On the other hand, the Ethics Board (EB) will scrutinise the research, to guarantee that no undue risk for the user, whether technically or related to the breach of privacy, is possible.

As evaluations will take place in 3 countries across Europe, attention should be specifically paid to the (relevant) national/regional/institutional regulation of each country. To collect national regulation and local ethics practices, a questionnaire has been formulated and provided in Annex I and the results of which are reported in chapter 4 of D9.2.

An Ethics Site Responsible has been chosen for each Use Case (local ethics representative), who represents the country with respect to ethics issues in specific. In case the pilot site managers decide to place another person in charge of ethics, then the table below must be updated. EB will train and monitor the Local Ethics Representatives to abide to the European and national regulation, laws, and guidelines and PANACEA Ethics Policy. In turn, the ethics responsible person at each pilot site will train and appoint the person who will be managing and organising recruitment processes and safekeeping of participants contact details. The ethics responsible person will inform the EB of any recruitment issues and threats that may appear with regards to data protection and end-user involvement in pilots.

Training delivery (face to face, online remote, documentation sharing, etc.) to the local ethics representatives will be managed case-by-case.

PANACEA Ethics Board will also be closely collaborating with the WP6 pilot leader who will act as the moderator and communicator between the pilot sites and the project's EB team. All Ethics approval will reside on sharepoint and will annexed in this Deliverable or the next version of the Ethics deliverables' series (D9.3; M22).

12.2 Data protection

For PANACEA to achieve its mission and to meet its objectives, a series of data, including personal data, is required to be collected, processed, used, and managed. Data collection and processing in PANACEA adheres to the respective European regulations, encompassing General Data Privacy Regulation (GDPR) and the PANACEA Data Management Plan (D9.4; M6 and D9.7; M34). Pilot site leaders will complete a Data Protection Impact Assessment (DPIA; Annex VII in D9.4) necessity form to investigate if a DPIA needs to be initialised beforehand. This process will start in M15 and will be completed before any tests take place. In addition, pilot representatives will participate in the completion of the FAIR templates (section 5.3 and Annex VI in D9.4) to identify the data characteristics, restrictions, etc. If data exchange requires an agreement, this will be prepared accordingly.

12.3 Covid-19 measures

Considering the current Covid-19 pandemic, each UC team is also responsible for taking necessary measures to ensure minimal risk of spreading the SARS-CoV-2 virus. These could include the use of personal protective equipment, intensified cleaning of vehicles and facilities, measures to avoid crowding, or modified data collection procedures depending on the situation at the time of data collection in each study. The pilot sites are responsible for adhering to local Covid-19 restrictions during data collection activities. Adaptations should be clearly described in the internal reports from each study.

12.4 Technical validation

The aim of the technical validation is to check the technical functioning of the PANACEA data collection systems in the real operational (or simulator) setting. It will enable identification of potential problems with the sensors and should also permit to validate the data collection procedure from data acquisition and data transmission to data storage. The iterative process will ensure that any problems encountered during implementation can be fed back to relevant WPs and be resolved before starting the main data collections.

The technical validation must be prepared and conducted prior to the visit of the first participant. The technical validation can be performed with a member of the working group that is not directly involved in the preparation of the study. This will assure a higher independency of the feedback given regarding failures and improvements. The technical validation should be conducted exactly as if it was a session with a real participant (information sheets, technical protocol, experimenter guide and instructions should be used). This serves to verify if all equipment is working properly and if the procedure is efficient. During the technical validation, data must be recorded as this allows to confirm if the output dataset can be used to perform the planned analysis.

Protocols for technical validation will be developed in A6.2, A6.3 and A6.4 in collaboration with WP4. The results of the technical validations will be reported in MS15-MS17. The protocols will be included in the next version of this deliverable (D6.2).

13 Data collection, analysis, and reporting

13.1 Data collection

This section presents an overview of what the steps that will take place at the sites during data collection.

13.1.1 Participant recruitment

When recruiting participants to the studies, selection criteria will be considered such as gender, and age. Care should be taken to ensure a representative sample, and a sufficient sample size. The recruitment will be done before the data collection takes place in all studies except the roadside study and will be conducted by the respective team on site. All people that will be actively participating in a study, will take part in a thorough recruitment and informed consent procedure, that will be particularly stringent to ensure no coercion (not even soft or indirect) is exerted. In the Ethics manual of PANACEA, the recruitment process is described and information to be included in the recruitment material is listed. The study can be advertised in the media (e.g., website, local newspapers, email messages), locally (distribution of prospects and information sheets in the facilities), and via direct contact of potential participants. Some extra participants should also be recruited in case of drop-out. Appointments will be scheduled with the participants and, to assure that drivers do not forget an appointment, a member of the pilot team will call the driver/operator/passenger a day before reminding him/her about the scheduled session's time.

13.1.2 Information sheets, consent forms and questionnaires

The informed consent procedure is described in detail in the Ethics manual of PANACEA. Each UC team will edit the required templates of the informed consent and information sheets and will define the procedures regarding the collection, storage, and protection of personal data, in compliance with the European and national legislation. The Pilot sites are responsible for translating all the material that need to be read or filled out by participants if the participants do not have enough English skills. Consent forms need to be signed before the data collection starts and should follow the requirements specified in the Ethics manual of PANACEA. Questionnaires and scales will be implemented in web-based applications, which will ease storing information and reduce the amount of work prior to data analysis.

13.1.3 Protocols and instructions

It is recommended to create a study protocol consisting of a checklist for each data collection to ensure that all equipment is in place and working. It facilitates reviewing that all sensors and vehicles/simulators are working as intended before the data collection starts. Before starting the data collection, members of the staff should go through this protocol. A schedule of the study should be attached to the protocol. The schedule should contain a list of all participants with a time plan for when each participant is scheduled for data collections.

The protocol should also show, step by step, which actions the experimenter from the UC team should take to set up and run the study, including which materials are needed, where he/she should ask the participant to do, and instructions that must be given to the participant. Certain information, like goals of the study, test procedure and system description, must be read (verbally) in order to assure that all participants receive the same instructions.

13.1.4 Procedure

The procedure for collecting data using the PANACEA sensors is described in detail and available to the pilot sited in the internal deliverables ID 3.1: 'Off-duty assessment: Measures and Thresholds' ID3.2: 'On-duty assessment: Measures and Thresholds', ID3.3: 'On-site assessment: Measures and Thresholds', and ID3.4: 'Roadside assessment: Measures and Thresholds'. Baseline measurements in the final evaluations should be taken using the applicable PANACEA sensors "passively", i.e., without having the connected countermeasures' system activated. For reference measurements, the pilot sites are referred to the respective technology's user manual.

13.2 Data delivery

All datasets will be harmonised and collected in the agreed formats and types before they are shared with the consolidation and analyses teams in A6.5 and feed the impact assessment simulations to be performed in WP7. Data collected at the sites is delivered as raw data to the PANACEA platform. Data analyses will be performed both centrally and at the sites, depending on the purpose of the data analysis. Templates will be included for sharing the main outcomes or needs for improvements with WP3, WP4, and WP5. The data harmonisation templates will be developed in WP6 after the type and format of data are gathered in A2.5; the harmonisation template will be included in the update of this deliverable (D6.2).

13.3 Data analysis

Each UC team will be responsible for collecting and pre-processing and/ or processing datasets according to the data analysis plan. Most of the analyses are done by the UC teams at the pilot sites. For A6.2 analysis, WP3 is responsible for the final analyses of the datasets and the subsequent setting/ refining of thresholds, levels and algorithms. However, some analyses are performed in A6.5 with the purpose of consolidating findings from the different professional driver groups. Impact analysis and calculation of high-level KPIs is done in WP7.

Some steps of the data analysis are common for all studies. The first step is to perform a data quality check. This should preferably be performed at regular intervals also during the data collection to see if any problems arise over time. Thereafter, cleaning and pre-processing of data will be done by removing bad quality data and calculating output parameters. In this step, it is important to register how much of the data was removed due to bad quality. For the PANACEA technologies the output parameters will be calculated by the PANACEA system, but for the reference equipment used in the simulator studies the data processing will be done by the respective pilot site. Data from questionnaire instruments used in the evaluation will also be processed at the sites. This includes re-coding of individual ratings and calculation of scores and indices according to the description for each instrument. The final questionnaire tools and detailed descriptions about scoring will be included in D6.2.

13.4 Reporting results

Results from each study will be compiled by each site and they will write internal reports based on pre-defined templates (in Appendix IV). Each report will include a description of the research questions, methods, analysis, results and conclusions of each data collection. The structure of the internal report is similar to the experimental plan for the respective study but includes the analyses performed, results and conclusion sections. The consolidated results of all studies performed in WP6 will be reported in D6.3: 'Consolidation of Pilots' results' as described in the chapter below.

13.5 Results consolidation

Following the tests conducted in A6.3 and A6.4, results will be collected and processed in A6.5, analysed, discussed, and made publicly available in D6.3 and/or journal/conference publications. Results (both raw and metadata based/consolidated) will be provided to WP7 for performing the impact assessments, as well as to A7.4 to adapt to the relevant exploitation plans. The results of A6.4 will be also analysed to: a) answer the research questions and address the evaluation-oriented objectives as described in this deliverable, and b) assess acceptance, trust, willingness to use of drivers/ riders/operators and stakeholders of CHTs, countermeasures and of the PANACEA solution in general. The conclusions are expected to lead to recommendations for future system(s) optimisation, application guidelines and areas requiring further research and lessons learnt.

14 Impact assessment

Impact assessment will be performed in WP7, starting in M22 of the project. The main aim is to assess the project impacts enabling and verifying the release of the impacts/benefits of the project. The specific aspects investigated are: the project impact in relation to the EU safety targets; the impacts of the countermeasures proposed and developed by the project (related to A6.3); cross-modal transferability, ensuring that the outputs of the project are beneficial also in other transport modes (related to A6.4); the simulation of various scenarios to explore the impacts of the project solutions at different levels. As illustrated in Figure 18, the PANACEA impact assessment process is highly dependent on data from the WP6 data collections as input to the various WP7 activities.

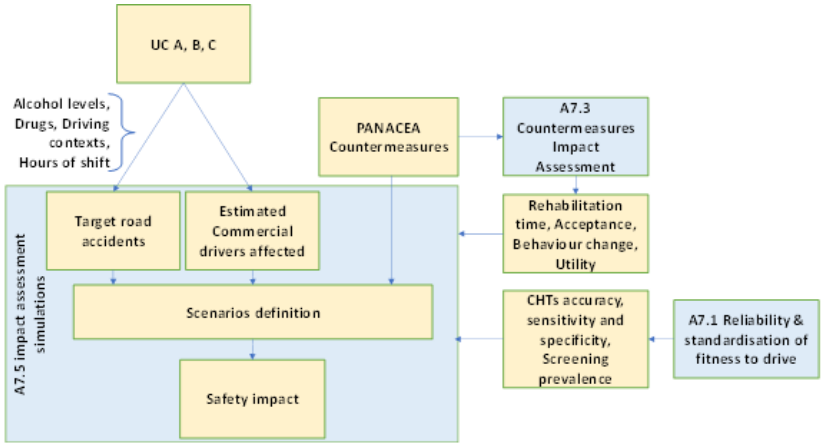


Figure 18. *The PANACEA impact assessment process*.

A7.3 aims to assess the impact of the countermeasures developed in WP5 and piloted in WP6 and to evaluate their potential impact in relation to the PANACEA impact targets beyond SoA. The results from the countermeasure pilots (simulator and on road/test track pilots and the cloud-based coaching and supporting system tests) will be used to assess the impact of these countermeasures. The impact of the pilots will be measured in terms of behaviour change, fit for purpose assessment and user acceptance. The potential impact of the countermeasures will be evaluated in terms of combating driving impaired by medicines or excess fatigue will be evaluated. A focus will be the extent to which they can accelerate rehabilitation (project target of 20%) and combat the appearance and perseverance of the addressed impairment types (project target 25%). The impact of the cloud-based coaching and supporting system on improving efficiency and effectiveness of roads policing/traffic police operations will also be assessed. This task is dependent on the work in WP5 to develop countermeasures and the design and running of the pilots in WP6.

Moreover, the EU Road Safety Policy Framework (2021-2030) has set a long-term, comprised by interim ones, goals to reach zero deaths and injuries by 2050 (addressed by A7.1). An analysis of the safety impact mechanisms of each UC will define the target road accidents and related road injuries addressed by each one of them. The AIT mobile unit, for example, can play an important role in the early detection of high and low arousal states Initially, the Fitness-to-Drive assessments are planned as on-the-spot measurements at pre-defined occasions (e.g., start of shift, in regulated breaks, etc.), which will return indicators for the arousal states to initiate preventive strategies. In the long term, an integration into the driving environment (e.g., steering wheel) is realistic to allow for a continuous assessment triggering the immediate initiation of needed preventive strategies. Commercial drivers are at high risk for crashes with severe impact on various social (e.g., injuries, deaths) and economic (e.g.,

consequential costs due to acute injuries, long-term health complications, environmental damages, traffic breakdown, delivery problems) levels. Thus, already a small reduction in crashes can lead to a significant reduction in fatal/non-fatal events and consequential costs. The AIT innovation might as well have an impact on the automotive industry by opening a new area of integrated and unobtrusive assessment of the driver's fitness even in the non-commercial driving business. Furthermore, the obtained findings can be translated to other domains (e.g., medical domain) and environments (e.g., sports) as well. Several scenarios will be built according to various input like the number of commercial drivers affected, the performance of CHTs and the countermeasures proposed in PANACEA. Each scenario will be compared to the reference scenario, which assumes no major improvements are implemented. The safety impact of the proposed solutions will be estimated based on results in terms of rehabilitation time, user acceptance, behaviour change and CHTs reliability and screening prevalence coming from activities **A7.1 and A7.3.**

15 Conclusions

This deliverable provides the framework for all WP6 data collections. This version of the deliverable has the main focus on setting the framework and the work process. The purpose of the PANACEA framework is to create a common framework to be used in all studies to make sure the data are collected in a way that makes it possible to consolidate the results in the end and to provide what is needed for impact analysis. Studies will be done to serve different purposes during the project. Simulator (A6.2) and roadside (A3.4) studies will be performed to validate PANACEA sensors and refine WP3 algorithms. Real-road and semi-real-road studies will be performed to validate and assess the final CHTs (A6.3) and countermeasure solution (A6.4).

The deliverable presents both a horizontal perspective of the pilot sites and what will be included in the different studies, but also the details for each site to be able to perform the data collections needed to for the generic evaluation and impact assessment. The general data gathering tools (objective and subjective) are identified and will be further refined in the update of the deliverable. A set of guidelines on practicalities and ethical aspects to take into consideration before and during data collection are presented.

An update of this deliverable will be done in M22. In the next version of the deliverable, the following will be added:

- Final RQs and KPIs.
- Detailed experimental plans for each study with descriptions of how to carry out the data collections at the sites.
- The final questionnaire tools and detailed descriptions about scoring.
- Protocols for technical validation of simulator studies, validation and assessment pilots, and countermeasures' pilots.
- Data harmonization templates for sharing the main outcomes or needs for improvements with WP3, WP4, and WP5 and for results consolidation in A6.5.

References

Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., . . . Lang, T. (2001). The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration. *Annals of Internal Medicine, 134*(8), 663-694. doi:10.7326/0003-4819-134-8-200104170-00012

Babor, T. F., Biddle-Higgins, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Health Care*. Retrieved from Geneva:

Balestroni, G., & Bertolotti, G. (2015). EuroQol-5D (EQ-5D): an instrument for measuring quality of life. *Monaldi Archives for Chest Disease, 78*(3). doi:10.4081/monaldi.2012.121

Barnard, Y., Innamaa, S., Koskinen, S., Gellerman, H., Svanberg, E., & Chen, H. (2016). Methodology for Field Operational Tests of Automated Vehicles. *Transportation Research Procedia, 14*, 2188-2196. doi:https://doi.org/10.1016/j.trpro.2016.05.234

Brooke, J. (1996). SUS: A 'Quick and Dirty' Usability Scale. In P. Jordan, B. Thomas, I. McClelland, & B. Weerdmeester (Eds.), *Usability Evaluation In Industry* (pp. 189-194). London: CRC Press.

Dahlgren, A., Kecklund, G., & Åkerstedt, T. (2005). Different levels of work-related stress and the effects on sleep, fatigue and cortisol. *Scandinavian journal of work, environment & health*, 277-285.

Dehn, D. M. (2008). Assessing the Impact of Automation on the Air Traffic Controller: The SHAPE Questionnaires. *Air Traffic Control Quarterly, 16*(2), 127-146. doi:10.2514/atcq.16.2.127

Elo, A. L., Leppänen, A., & Jahkola, A. (2003). Validity of a single-item measure of stress symptoms. *Scandinavian Journal of Work, Environment and Health, 29*(6), 444-451.

Fynn, J. F., Hardeman, W., Milton, K., & Jones, A. P. (2020). A scoping review of evaluation frameworks and their applicability to real-world physical activity and dietary change programme evaluation. *BMC Public Health, 20*(1), 1000. doi:10.1186/s12889-020-09062-0

He, J., Zeng, Z., & Li, Z. (2010). Benefit Evaluation Framework of Intelligent Transportation Systems. *Journal of Transportation Systems Engineering and Information Technology, 10*(1), 81-87. doi:https://doi.org/10.1016/S1570-6672(09)60025-8

Jansuwan, S., Liu, Z., Song, Z., & Chen, A. (2021). An evaluation framework of automated electric transportation system. *Transportation Research Part E: Logistics and Transportation Review, 148*, 102265. doi:https://doi.org/10.1016/j.tre.2021.102265

Milstein, B., & Wetterhall, S. F. (1999). Framework for program evaluation in public health.

Newman-Askins, R., Ferreira, L., & Bunker, J. M. (2003). *Intelligent transport systems evaluation: From theory to practice.* Paper presented at the 21st ARRB and 11th REAAA Conference.

Nordin, M., Åkerstedt, T., & Nordin, S. (2013). Psychometric evaluation and normative data for the Karolinska Sleep Questionnaire. *Sleep and Biological Rhythms, 11*(4), 216-226. doi:10.1111/sbr.12024

Rakoff, H., Smith, S., Innamaa, S., Barnard, Y., Harrison, G., & Shaw, J. (2020). Building Feedback into Modelling Impacts of Automated Vehicles: Developing a Consensus Model and Quantitative Tool.

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), 1-10. doi:https://doi.org/10.1016/S0968-090X(96)00025-3

Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Bulletin of the World Health Organization, 85*, 867-872.

Yusof, M. M., Kuljis, J., Papazafeiropoulou, A., & Stergioulas, L. K. (2008). An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit). *International Journal of Medical Informatics, 77*(6), 386-398. doi:https://doi.org/10.1016/j.ijmedinf.2007.08.011

Åkerstedt, T., Anund, A., Axelsson, J., & Kecklund, G. (2014). Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *J Sleep Res, 23*(3), 240-252. doi:10.1111/jsr.12158

Appendix I Template for Experimental plans

PANACEA

1 Study UCA-S, UCA-R, UCB-S1, UCB-S2, UCB-R, UCC-R, or Roadside

1.1 Research questions

1.2 Participants

1.3 Vehicles

1.4 Environment

1.5 Driver impairments

1.6 Countermeasures

1.7 Study design

1.8 Data collection tools

1.9 Data analysis plan

1.10 Time plan

Appendix II Research Questions

RQ-category	Specific RQ	KPI (tentative)	Data gathering tool	UC	Data collection
Technical validation	**Do the PANACEA sensors/systems detect targeted driver impairments effectively with high sensitivity and specificity?**	KPI 3.1 Reliability of CHT, 3.2 Specificity of CHT, 3.3 Sensitivity of CHT, KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from sensors, subjective ratings of impairment	All (not all sensors in all UC)	All (not all sensors in all data collections)
	How is the performance of the PANACEA sensors compared to a reference measurement?	KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from PANACEA sensors & reference equipment	All	Simulator and roadside studies
Technical validation	How is the performance of the LEITAT sensor compared to the commercial drug sensor used by the Police in Norway?	KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from Leitat sensor & reference equipment	All	Roadside
Technical validation	How is the performance of the SENSEAIR Go Portable compared to the commercial alcohol sensor used by the Police in Norway?	KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from Senseair sensor & reference equipment	All	Roadside
Technical validation	How is the performance of the LEITAT sensor compared to the blood tests used by the Police in Norway?	KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Leitat sensor and blood test	All	Roadside
Technical validation	Does the sleep/wake history (24h data) in combination with a BMM give the same information compared to the subjective before-driving rating used by Datik?	KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from fitbit, BMM, & Datik	UCA	UCA real-world data collection

RQ-category	Specific RQ	KPI (tentative)	Data gathering tool	UC	Data collection
Technical validation	How do the measurements of the DATIK system and Optalert match?	KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from Datik and Optalert	UCB	UCB CERTH simulator study
Technical validation	How do the measurement of SENSEAIR and BACtrack skyn match?	KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from Senseair and BACtrack sensors	UCB	UCB CERTH simulator study
Technical validation	Does the AIT Pulse Wave Analysis (PWA) device and Galvanic Skin Response (GSR) sensors' measurements match?	KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from PWA and GSR sensors	UCB	UCB CERTH simulator study
Technical validation	Will addressed levels of driver state and/ or impairment be captured?	KPI 3.1 Reliability of CHT, 3.2 Specificity of CHT, 3.3 Sensitivity of CHT	Data from sensors, subjective ratings of impairment	UCB	UCB CERTH simulator study
Technical validation	What sensor data are the best driver state behaviour impairment indicators?	KPI 3.1 Reliability of CHT, 3.2 Specificity of CHT, 3.3 Sensitivity of CHT	Data from sensors	UCB	UCB ViF simulator study
Sensor fusion	Which combination of algorithms can best capture impaired driving in the respective environment?	KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from sensors	UCB	UCB ViF simulator study
Technical validation	What are the critical differences in detecting impaired driving in city traffic versus motorway / country road traffic?	KPI 3.1 Reliability of CHT, 3.2 Specificity of CHT, 3.3 Sensitivity of CHT	Data from sensors	UCB	UCB ViF simulator study
Effectiveness and operability	Is the LEITAT/SENSEAIR Go Portable sensor reliable and easy to use in roadside assessments?	KPI 3.4 Ease to use CHT, 3.5 Usefulness of CHT, 3.6 Willingness to use CHT	Questionnaires	All	Roadside
Sensor fusion	**Do the combined sensors improve driver state detection?**	KPI 6 Sensitivity and specificity of a sensor or combination of technologies	Data from sensors (individual and combined)	All	All real-world data collections ?

RQ-category	Specific RQ	KPI (tentative)	Data gathering tool	UC	Data collection
Sensor fusion	Can sleep/wake history (24h data) in combination with a BMM be used to distinguish different types of fatigue (and thus give more accurate countermeasures)?	KPI 6 Sensitivity and specificity of a sensor or combination of technologies,	Data from fitbit, BMM and subjective ratings of fatigue	UCA	UCA real-world data collection
Validation of the integrated system in real life	**Does the PANACEA integrated solution work in a real-life setting to detect impairment and deliver counter measures?**	KP 3 Technical performance of CHT, KPI 3.1 Reliability of CHT, 3.2 Specificity of CHT, 3.3 Sensitivity of CHT	Data from PANACEA solution and subjective ratings of impairment	All	All real-world data collections
Validation of the integrated system in real life	Is it possible to get around using highly specific baseline/calibration recordings and still get accurate estimates of driver state?	KPI 3 Technical performance of CHT?	Data from sensors	UCB	UCB real-world data collection
Acceptance	**Are the PANACEA sensors/systems accepted by the users?**	KPI 3.7 Acceptance of CHT	Questionnaires	All	All real-world data collections
Usability	**Are the CHTs perceived as useful, satisfying, trustworthy, and easy to use?**	KPI 3.4 Ease to use CHT, 3.5 Usefulness of CHT, 3.6 Willingness to use CHT	Questionnaires	All	All real-world data collections
Acceptance	How willing are the participants to use wearable devices 24h a day? What is the data availability after an extended period (several months) of usage? Is it too intrusive?	KPI 3.4 Ease to use CHT, 3.5 Usefulness of CHT, 3.6 Willingness to use CHT, 3.8 Trust in CHT	Questionnaires and focus groups	All	All real-world data collections
Willingness to use	Why do drivers not engage with the CHT if they don't engage?	KPI 3.6 Willingness to use CHT	Focus group	All	All real-world data collections
Effectiveness of the countermeasure	**What are the immediate effects of implemented countermeasures?**	KPI 4.2 Effectiveness of a countermeasure	Questionnaires/Data from PANACEA system	All	All real-world data collections

RQ-category	Specific RQ	KPI (tentative)	Data gathering tool	UC	Data collection
Effectiveness of the countermeasure	Will the 24h data reveal poor sleep hygiene, and if so, is it possible to fix with the Panacea countermeasures?	KPI 4.2 Effectiveness of a countermeasure	Data from fitbit and PANACEA system	UCA	UCA real-world data collection
Effectiveness of the countermeasure	From iCloud System data is it possible to measure the effects (short-term and lifestyle) of an implemented countermeasure?	KPI 4.2 Effectiveness of a countermeasure	Data from PANACEA solution	All	All real-world data collections
Effectiveness of the countermeasure	Is the AIT system sensor effective as a countermeasure for stress?	KPI 4.2 Effectiveness of a countermeasure		UCB, UCC	UCB & UCC real-world data collection
Acceptance	Does the countermeasures for sleep related fatigue (while driving) work in a professional setting with tight schedules?	KPI 4.1 Acceptance of a countermeasure	Questionnaire	All	All real-world data collections
Acceptance	Are drivers willing to sacrifice their breaks to do scheduled measurements and relaxations tasks?	KPI 4.1 Acceptance of a countermeasure, 4.3 Satisfaction	Questionnaire	All	All real-world data collections
Acceptance	**Is the PANACEA countermeasures system accepted by the users?**	KPI 4.1 Acceptance of a countermeasure, 4.3 Satisfaction	Questionnaire	All	All real-world data collections
Acceptance	To what extent do drivers/operators engage with the countermeasures delivered by the cloud based system	KPI 4.1 Acceptance of a countermeasure	Usage data from PANACEA system	All	All real-world data collections
Willingness to use	Why do drivers not engage with the countermeasure if they don't engage?	KPI 4.X Willingness to use countermeasure ?	Focus group	All	All real-world data collections
Impact of countermeasures	**Does behaviour change/improve after the relevant countermeasure has been administered?**	KPI 4.2 Effectiveness of a countermeasure, KPI10 CEA/CBA	Questionnaires/Data from PANACEA system	All	All real-world data collections

RQ-category	Specific RQ	KPI (tentative)	Data gathering tool	UC	Data collection
Impact of countermeasures	**Will the PANACEA countermeasures reduce driver impairment and improve the driver performance?**	KPI 8.1 N of saved lives (ON ROADS), KPI8.2 N of saved lives (OFF ROADS), KPI9 QoL?	Questionnaires/Data from PANACEA system/driving performance data from vehicles	All	All real-world data collections
Long-term usage (business case)	**Would it be possible to implement the PANACEA system in regular operation?**	KPI 10 CEA/CBA	Focus group with different stakeholders	All	All real-world data collections
Safety	**Does the PANACEA system increase perceived (drivers) and reported (operators) safety?**	KPI 1.1 Perceived (drivers) safety, KPI 1.2, Reported (operators) safety	Questionnaire & focus group (& data from PANACEA solution?)	All	All real-world data collections
Study-specific RQ	Can fatigue prediction using BMM be improved by taking next-day effects of alcohol consumption into account?		Data from sensors, subjective ratings of impairment	UCA	UCA simulator study
Study-specific RQ	How does moderate alcohol intake in the evening affect night sleep and next day driving performance?		Data from sensors, simulator data, subjective ratings of impairment	UCA	UCA simulator study
Study-specific RQ	How do fatigue levels change across the working shift?		Data from sensors, subjective ratings of impairment	UCB	UCB CERTH simulator study
Study-specific RQ	How do stress levels change across the shift?		Data from sensors, subjective ratings of impairment	UCB	UCB CERTH simulator study

Appendix III Questionnaire instruments

Karolinska Sleep Questionnaire (KSQ)

Please indicate the degree to which the following have happened to you during the last 3 months.						
	Never	**Seldom** *(One or few times a year)*	**Sometimes** *(Several times a month)*	**Often** *(1-2 times a week)*	**Most of the time** *(3-4 times a week)*	**Always** *(5 times or more a week)*
a. Difficulties falling asleep	o	o	o	o	o	o
b. Difficulties waking up	o	o	o	o	o	o
c. Repeated awakenings with difficulties going back to sleep	o	o	o	o	o	o
d. Your own loud snoring	o	o	o	o	o	o
e. Gasping for breath, “snorting” during sleep	o	o	o	o	o	o
f. Interrupted breathing during sleep (sleep apnoea)	o	o	o	o	o	o
g. Nightmares	o	o	o	o	o	o
h. Not fully rested at awakening	o	o	o	o	o	o
i. Premature (final) awakening	o	o	o	o	o	o
j. Disturbed/restless sleep	o	o	o	o	o	o
k. Involuntary twitching of the legs that disturbs sleep	o	o	o	o	o	o
l. Too little sleep (at least 2 hours too little per main sleep period)	o	o	o	o	o	o
m. Being constantly tired throughout the day	o	o	o	o	o	o
n. Physical exhaustion	o	o	o	o	o	o
o. Mental exhaustion	o	o	o	o	o	o
p. A feeling of being exhausted at awakening	o	o	o	o	o	o
q. Sleepy during work	o	o	o	o	o	o
r. Sleepy during leisure time	o	o	o	o	o	o
s. Unintended periods of sleep (nodding off) during work	o	o	o	o	o	o

t. Unintended periods of sleep (nodding off) during leisure time	o	o	o	o	o	o
u. Having to fight against sleep in order to stay awake	o	o	o	o	o	o

Karolinska Sleepiness Scale (KSS)

Karolinska Sleepiness Scale (KSS) is a subjective rating scale where participants can rate their own sleepiness level (Åkerstedt & Gillberg, 1990). Respondents indicate on a nine-point scale how sleepy they have felt (1 = extremely alert to 9 = very sleepy, great effort to keep awake, fighting sleep) on average during the previous 5 minutes.

Table 13. *Karolinska Sleepiness Scale (KSS)*

Scale	Risk Level
1	Extremely alert
2	Very alert
3	Alert
4	Rather alert
5	Neither alert nor sleepy
6	Some signs of sleepiness
7	Sleepy, but no effort to keep awake
8	Sleepy, some effort to keep awake
9	Very sleepy, great effort to keep awake, fighting sleep

VTI Acute Stress Scale (VSS)

A modified version of the Stockholm University Stress scale (SUS; Dahlgren, Kecklund, & Åkerstedt, 2005) serves as a reference for quantifying stress. The 9 verbal anchors were here changed to match the KSS anchors: 1 completely relaxed (feeling entirely calm and relaxed), 2 very relaxed, 3 relaxed, 4 rather relaxed, 5 neither relaxed nor stressed, 6 slightly stressed, 7 stressed (feeling some tension and pressure), 8 very stressed, 9 extremely stressed (feeling very tense and under high pressure, on the verge of what I can handle).

Table 14. *VTI acute Stress Scale (VSS)*

Scale	Risk Level
1	Completely relaxed (feeling entirely calm and relaxed)
2	Very relaxed
3	Relaxed
4	Rather relaxed
5	Neither relaxed nor stressed
6	Slightly stressed
7	Stressed (feeling some tension and pressure)
8	Very stressed
9	Extremely stressed (feeling very tense and under high pressure, on the verge of what I can handle)

Stress symptoms

Perceived stress is assessed by the following single-item measure of stress symptoms (Elo et al., 2003); “Stress means a situation in which a person feels tense, restless, nervous or anxious or is unable to sleep at night because his/her mind is troubled all the time. Do you feel this kind of stress these days?” The response is recorded on a 5-point Likert scale where 1 = not at all, 2 = just a little, 3 = to some extent, 4 = rather much and 5 = very much.

Technology Acceptance Scale by van der Laan

The following User Acceptance Scale was developed by Van Der Laan et al. (1997). The participants 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.

My judgements of the (… name…) system are … (please tick one box in every line)

Useful	☐ ☐ ☐ ☐ ☐	Useless
Pleasant	☐ ☐ ☐ ☐ ☐	Unpleasant
Bad	☐ ☐ ☐ ☐ ☐	Good
Nice	☐ ☐ ☐ ☐ ☐	Annoying
Effective	☐ ☐ ☐ ☐ ☐	Superfluous
Irritating	☐ ☐ ☐ ☐ ☐	Likeable
Assisting	☐ ☐ ☐ ☐ ☐	Worthless
Undesirable	☐ ☐ ☐ ☐ ☐	Desirable
Raising Alertness	☐ ☐ ☐ ☐ ☐	Sleep-inducing

Procedural guidance for user acceptance scale

1. 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).
2. After experiences with the system under evaluation present the nine items again: 'what is your judgement about the system …(name), you just finished driving with' (after-measurement).
3. 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 (N.B. these items are mirrored).
4. Perform reliability analysis on the before-measurement (use of Cronbach's □ is strongly suggested). If reliability is sufficiently high (above 0.65), compute per subject the end-scores for the two scales by averaging the scores on items 1, 3, 5, 7, and 9 for the ***usefulness*** score, and averaging scores on items 2, 4, 6, and 8 for the ***satisfying*** score.
5. 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.
6. Compute difference-scores 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.

System Usability Scale (SUS)

The questionnaire below was designed to evaluate the Human Machine Interface of an in-vehicle system (Brooke, 1996).

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

The participant's scores for each question are converted to a new number, added together and then multiplied by 2.5 to convert the original scores of 0-40 to 0-100.

- For each of the odd numbered questions, subtract 1 from the score.
- For each of the even numbered questions, subtract their value from 5.
- Add up the converted numbers to a total score. Then multiply this by 2.5.

Though the scores are 0-100, these are not percentages and should be considered only in terms of their percentile ranking. Based on research, a SUS score above a 68 would be considered above average.

SHAPE Automation Trust Index (SATI)

The SATI provides a measure of human trust in automated systems (Dehn, 2008). The respondent answers six questions on a seven-point Likert scale ranging from “never” to “always”.

	never						always
1)…the system was useful.	0	1	2	3	4	5	6
	never						always
2)…the system was reliable.	0	1	2	3	4	5	6
	never						always
3)…the system worked accurately.	0	1	2	3	4	5	6
	never						always
4)…the system was understandable.	0	1	2	3	4	5	6
	never						always
5)…the system worked robustly (in difficult situations, with invalid inputs, etc.).	0	1	2	3	4	5	6
	never						always
6)…I was confident when working with the system.	0	1	2	3	4	5	6

Subtotal scores on each item can be obtained and a mean overall trust score can be derived for the system in question.

Appendix IV Internal Study Report Template

PANACEA

1 Study UCA-S, UCA-R, UCB-S1, UCB-S2, UCB-R, UCC-R, or Roadside

1.1 Research questions

1.2 Participants

1.3 Simulator/Vehicles

1.4 Environment

1.5 Driver impairments

1.6 Countermeasures

1.7 Study design

1.8 Data collection tools

1.9 Data analysis

1.10 Results

1.11 Conclusions

Internal Del ID: Internal Del name 1