Final report on POD test scenarios & performance requirements. Deliverable 4.1.4 of the Capri project

PUBLISHER

CapriMobility

VERSION

VoR (Version of Record)

Final Report on POD Test Scenarios & Performance Requirements

Capri Deliverable 4.1.4

Loughborough University

30 September 2020

Quality Information

Prepared by	Checked by	Approved by
Laurie Brown Steve Reed	Ruth Welsh	
Loughborough University	Loughborough University	

Revision History

Revision	Revision date	Details	Authorized	Name	Position
1.1 Draft	06/08/19				
2.1 Final	30/09/20				

Table of Contents

List of Acronyms

ACS	Autonomous Control System
ADS	Autonomous Driving System
AEB	Autonomous Emergency Braking
ASIL	Automotive Safety Integrity Level
AV	Autonomous Vehicle
CAV	Connected and Autonomous Vehicle
HARA	Hazard Analysis and Risk Assessment
LLDC	London Legacy Development Corporation
ODD	Operating Design Domain
QEOP	Queen Elizabeth Olympic Park
R&D	Research & Development
TTC	Time to Collision
UCD	User Centred Design
V&V	Verification and Validation

1. Executive Summary

The aim of the Capri project was to deliver a pilot scheme for the use of automated and connected passenger transport 'pods on demand' (PODs) as a mobility service in 'campus' locations such as airports, hospitals, business parks, shopping and tourist centres. These areas may be entirely privately owned or comprise elements of both off-road and public highway usage. Four trials have taken place during the Capri project; the first in an off road site closed to the public, two further on private off-road land that are open to the public, and the fourth trial operating as dual mode with the pod demonstrating use in public areas both off-road and on a privately owned road.

A key part of the project was to build user and regulatory trust in these vehicles. A major concern of end users is whether or not autonomous vehicles (AVs) will be safe, which is why Work Package 4 of the Capri project was tasked with developing a robust verification and validation (V&V) process to assess the safety of the Capri POD. Primarily this V&V was done in simulation in a virtual world, and the role of Loughborough University was to design test scenarios for use in these simulations.

Additionally, throughout the project it became clear that real-world physical safety testing was also required to reassure any concerns over carrying out the public trials. Although the PODs are thoroughly tested in-house by the manufacturer, Westfield Technology Group, Loughborough was asked to carry out further pre-trial testing to provide an independent evaluation of the safety performance. Detailed test plans for physical testing were created by combining the scenario generation methodology we had already developed for virtual testing, alongside our expertise in carrying out road accident investigation, real-world vehicle trials, and studies of driver behaviour.

The scope of this report is to give insight into the scenario design methodologies that were developed by Loughborough throughout the Capri project, and how this methodology was utilised to generate scenarios for both virtual and physical validation of the POD. The AV sector has at times been criticised for not being transparent, leading the public to question if safety concerns are being hidden from them. It is hoped that by making this information available for Capri and demonstrating the robust safety assurance procedures that were followed during the project, that end users will gain more confidence in the safety of the technology.

Section 2 of this report gives the reader some background information on AV development as well as some of the current procedures for testing and safety assurance for other AV types. Section 3 goes into detail on the process for designing scenarios, including the challenges, the methods developed within Capri to overcome these, and the data sources that can be used in scenario development. Finally, the specific testing protocols utilised for assessing POD performance in the virtual and physical worlds are respectively given in Sections 4 and 5.

It is noted here that the testing protocols for the physical testing included in this report are significantly more detailed than those for the virtual testing. This is because the physical testing was carried out by Loughborough, so the protocols cover the implementation as well as the scenario design. For the virtual world, implementation was carried out by other partners in WP4, so only the scenarios and test parameters are described. Further information on how the simulations were carried out can be found in reports from T&VS, the University of Bristol, and the University of the West of England.

Furthermore, safety concerns and testing procedures relating to cyber security were not within the scope of this report. This element was covered in detail by the University of Warwick and Nexor.

2. Background on Safety and Testing of Autonomous Vehicles

In the world of autonomous vehicle technology development, techniques move very quickly, and it is necessary to look at the entire ecosystem before defining how the Capri project will move forwards with the task of developing test scenarios and performance requirements.

This section is intended to give a brief overview of what AVs are and more specifically what a POD is, the potential safety benefits AVs might provide, and approaches being taken by governments and industry to ensure the safety of AVs.

2.1 Autonomous vehicles

Autonomous vehicles and the technology involved in their design and development is big business. Google was spending upwards of $150M per year on automation between 2009 and 2015, and the majority of large automotive manufacturers are rapidly hiring staff and investing many billions of pounds in the area. It is predicted that by 2035 the autonomous technology market could be worth around £2.7bn for vehicles sold in the UK and over £60bn across the global market.

These huge sums of money, combined with the rapid development in the technology, may lead to detrimental effects to safety, in that a highly competitive market may encourage the quick release of technologies that are not always validated through a solid certification process.

When discussing AVs, it is important to note the level of automation being considered. Across the industry the SAE (Society of Automotive Engineers) taxonomy for driving automation systems (SAE, 2018) is most commonly used. It makes a distinction between whether the primary driving task is carried out by the system or a human, whether or not a human is required to take over at any points in the journey, and further describes 5 levels of automation ranging from level 1 (driver assistance) to level 5 (full driving automation).

Within the AV sector, many car manufacturers are adapting their vehicles to use varying levels of autonomous technology. The Capri POD, however, is not a car. It has no windscreen, no traditional seating positions, and no human driver. In terms of usability, this POD is intended to operate both on-road and in pedestrianised areas, introducing a range of new environments and unique pedestrian and cyclist interactions that cars do not typically encounter.

The Capri POD is considered to be SAE level 4, meaning that humans are not expected to take over the driving task at any point, but that the vehicle can only operate within certain conditions or areas. More detail on the Capri POD specifications is given in Section 3.3

For AVs in general, but in particular for PODs, there exists no standardised, systematic, or structured methodology for safety assurance. This gap may present a significant obstacle in autonomous POD development and consumer uptake from two sides; firstly, the scarcity of data for road safety practitioners to develop into appropriate test scenarios for safety testing, and secondly the lack of perceived transparency by the manufacturers could form the perception that there is a lack of manufacturer accountability, resulting in lack of trust from the end users.

2.1.1 Safety objective

According to the World Health Organisation (WHO, 2018), road traffic accidents are the eighth leading cause of deaths worldwide; it is notable that road traffic accidents are the only ‘injury’ related cause of death in the top 10 causes. Across the globe, 1.35 million people are killed in road traffic accidents every year; in the UK this number is around 1,500. A prime motivator for more widespread use of AVs is that they will save lives.

A key point in the argument for vehicle autonomy is that a high proportion of road accidents are caused by human error. A number of studies, including the UK Department for Transport report “Pathway to Driverless Cars: a detailed review of regulations for automated vehicle technologies” (DfT, 2015) have reported that human error is a factor in over 90% of accidents. Considering the common perception of computer control is that it eradicates the human component, the implication is that over 90% of road accidents can be avoided if traditional vehicles are replaced by AVs. However, this target is commonly seen as unachievable, as it is probable that machines will also make mistakes, particularly when machine learning or artificial intelligence is involved.

Outside of the engineering and statistical approach to defining a safety objective, there is the public perception, i.e., what do the end users want? For example, would someone be willing to switch to an AV if it offered a 90% reduction in the risk of death? What about 50% Or 10%? What level of benefit would be needed to convince the average person that it is worth any perceived risk? It is safe to assume that public expectation of autonomous driving technology is high, however at the same time their understanding of the systems may be low. This disconnect may lead to unreasonable expectations of what can be achieved with AVs and could even lead to over-reliance on the technology which could result in more accidents.

Rationally, even a 1% reduction in accidents should be motive for humans to use AVs, as this would illustrate that machine driving is safer than human driving. However, it is expected that users will hold machines to a higher standard than that demanded by fellow humans. The general public perception is that people typically have zero tolerance for deaths caused by machines, meaning that AVs will need to undergo the most intensive testing and validation procedures before being widely adopted.

2.1.2 Testing of AVs

The UK government has shown a willingness to get a range of autonomous vehicles into use and onto the public road before 2021. As of 2018 there were 22 connected and autonomous vehicle (CAV) research and development (R&D) projects that had been funded by the UK government, with a further tranche of 51 projects covered by an investment of £180 million, and this number continues to grow, Although not all CAV projects will focus on road going vehicles, those that do are already facing barriers in terms of regulation and legislation to allow legal road use. This is also a concern in Capri, and one area of work within the project is focusing on how regulations may need to be adapted to allow PODs to operate on public highways.

Beginning in early 2018, a 3-year review by the Law Commission of England and Wales and the Scottish Law Commission has been examining the legal obstacles to the widespread introduction of self-driving vehicles and the need for potential regulatory reforms (Law Commission, 2019). If the UK is to maintain its position as a world leader in self-driving vehicle R&D, it is important that the laws and regulations governing the use of these vehicles is aligned with this aim. Up to this point the UK government has not followed the United States (US) example of allowing autonomous vehicles (Level 4/5) into unrestricted service on public roads, though trials are permitted in certain situations.

State legislature in the US enables individual states and municipal governments to introduce legislation related to autonomous vehicle use on public highways. By 2017, 33 states had introduced such legislation. All legislation must be in accordance with the guidelines for automated driving systems issued by the National Highway and Transportation Safety Administration (NHTSA, 2017; Department for Transportation, 2018).

Nevada was the first state to authorise operation of AVs in 2011. At present over 59 separate bills are enacted across the various states to allow a wide range of different autonomous functions to be tested on public highways, from truck platooning up to full level 5 autonomy. However, in recent years a number of high-profile collisions have led to public and regulatory pressure on the major players in the field of autonomous mobility in the US. This has led to technology companies such as Uber putting on-road testing on hold and Tesla facing a range of challenges on the term ‘autopilot’ and the use of its self-driving technologies on public roads.

In the UK, during the Capri project the Department for Transport released a code of practice automated vehicle trialling (DfT, 2019), and although this is still relatively light on detail, it does provide useful guidance and insight and the later Capri trials were required to comply with this.

Other countries around the world are also taking their own approaches to AV testing, and, outside of national processes, independent organisations tasked with determining vehicle safety are already working on roadmaps for AV testing and verification. Car manufacturers will naturally seek to promote the benefits of autonomous technologies and consumer purchasing decisions may be made on this information. There is a vital need for a common independent framework that is transparent and understandable in order for consumers can make informed decisions.

When discussing AV testing, it is also necessary to consider exactly what is being tested. In traditional vehicles, safety testing such as crashworthiness testing is done on a particular make and model, and the safety performance reported. Alongside this, human drivers are required to pass a driving test before they can legally drive a vehicle. However, with AVs the vehicle becomes the driver, and testing will need to examine whether it is testing the ‘vehicle’, the ‘driver’ (autonomous control system), or both. It is entirely possible that different ‘drivers’ could be implemented into different vehicles, and similarly that a type of vehicle could be operated by different control systems, which further complicates the issue of safety assurance testing.

2.2 Steps towards a complete verification framework

Safety testing scenarios and the performance requirements of these are only one part of the POD verification and validation process. Understanding the whole testing life cycle illustrates that this crucial element is in fact a relatively small part of the overall verification framework.

Whether conducted in the real world or through simulation, the phrase ‘test scenarios’ typically implies that the test will by dynamic and may involve other vehicles and/or interactions. These scenarios can present particularly challenging problems for autonomous systems to deal with and as such are relatively high level in terms of the complete safety assurance of a new vehicle type such as a POD.

Figure 1 shows a simplified version of the testing ecosystem and will be used as a reference for this section.

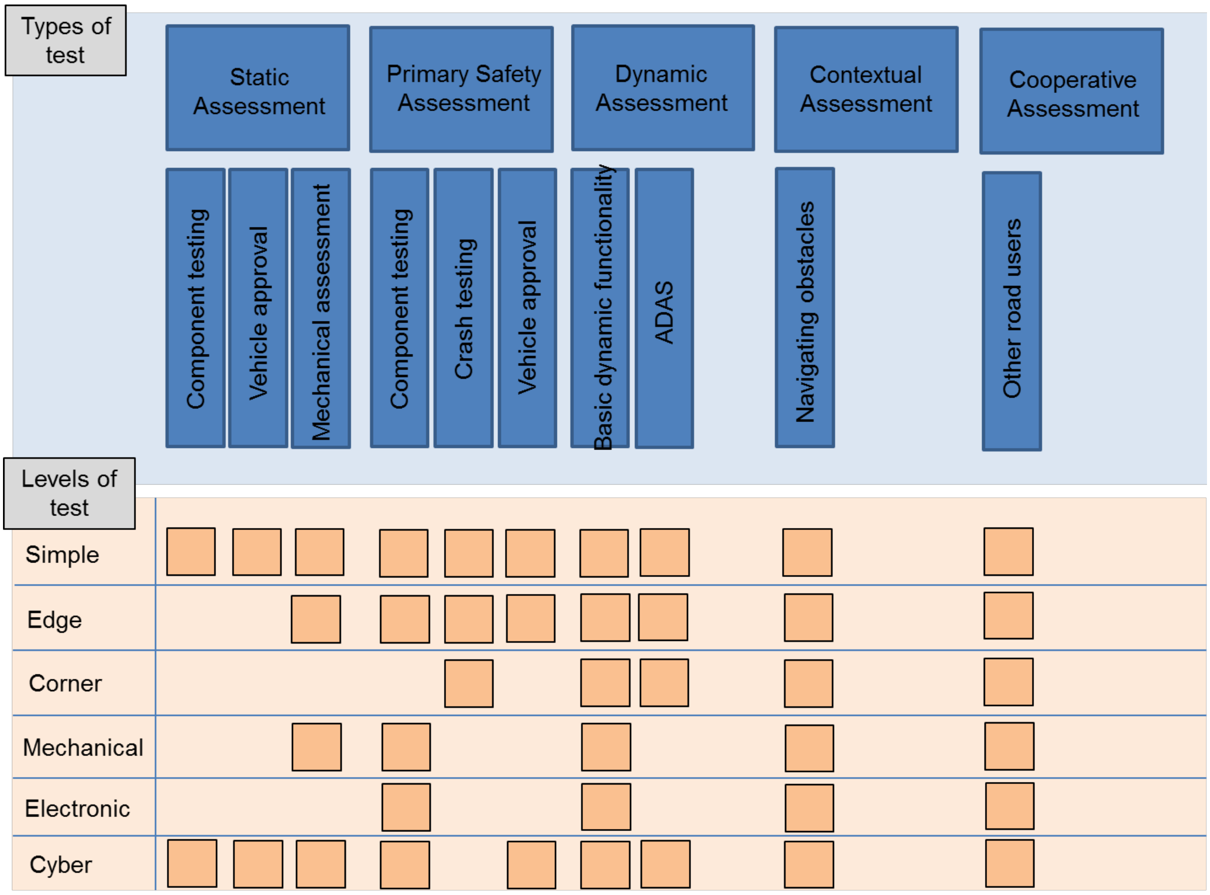


Figure 1: Simplified testing ecosystem for a complete verification framework

Before embarking on a testing plan the autonomous driving system (ADS) should first be specified across a range of safety elements. This provides an initial framework which constrains the testing across the subsequent types and levels of testing.

System safety: System safety is a specialism within systems engineering that supports program risk management. It is the application of engineering and management principles, criteria, and techniques to optimize safety. The goal of ‘system safety’ is to optimize safety through the identification of safety related risks, eliminating or controlling them by design and/or safety procedures, based on acceptable system safety precedence.

HARA: Hazard Analysis and Risk Assessment (HARA) forms the basis of a functional safety activity. The HARA determines the safety goals for the system and becomes the basis for deriving functional safety requirements. Automotive HARA is influenced by the ISO 26262 international standard

(ISO26262, 2018) and the process is imbedded within the assignment of Automotive Safety Integrity Levels (ASILs) dependent on the three factors: severity, exposure and controllability.

ODD: An Operational Design Domain (ODD) describes the specific conditions under which a given AV or autonomous system is intended to function. An ODD provides a definition of 'where'; such as what roadway types and speed and 'when'; under what conditions (day/night, weather limits, etc.) an AV or autonomous function is designed to operate.

2.2.1 Types of test

Figure 1 illustrates that the testing ecosystem could be broadly split into five main areas, with cyber security being overarching and interacting with all of these elements.

Static assessment: This includes purely static verification tests on anything from an individual vehicle component right through to whole vehicle approval testing. This section may involve some dynamic elements such as regenerative brake driveline testing on a rolling road, however the vehicle testing within this region should be aimed at determining the construction and use of the vehicle in question. In some ways the testing in this area can be viewed as similar to MOT level.

Primary safety assessment: Although predominantly dynamic in nature the tests for primary safety assessment could be designed to destructively test or determine safe limits for components; these tests can still be considered static as the vehicle is not driven under its own power or control. Tests in this area could be the replication of legislative/EuroNCAP crash tests or individual component tests to assess for example seatbelt anchorage points.

Dynamic assessment: This grouping is designed to provide initial validation for simple dynamic control involving the vehicle in isolation or with minimal infrastructure or other objects. Tests may involve determining the motor, steering or braking control around basic navigation tasks such as following simple paths or performing manoeuvres that may be expected in operation such as left/right turns or pulling to the side of a carriageway lane.

Contextual assessment: Once the vehicle is assessed dynamically in a simple environment the next type of testing should be the introduction of context. Context in the Capri case means dynamic POD tests with the introduction of environmental features, typically fixed or stationary, that the vehicle needs to detect. Tests may be similar to the layouts used in the simple dynamic assessments but with the introduction of parked vehicles, kerbs, traffic calming or control and other features typically seen adjacent to public roads (hedges, trees, walls, lighting poles, traffic lights, signage etc.)

Cooperative assessment: With the vehicle able to navigate and perform manoeuvres in an environment that represents the real world it then necessary to add other dynamic actors into the scene. These may be any of the normal road users a human driver would expect, but can also be scaled to test corner cases with, for example, stray or wild animals or vehicles being used in ways in which breach road rules or require adaptation from the AV; for example an opponent vehicle being driven on the wrong side of the road.

2.2.2 Levels of test

Not all trials will be or can be conducted to the same level. Some trials will, by necessity, be simpler than others while others will need to be relatively more detailed in order to test edge or corner cases. The extent of testing is not necessarily dependent on the type of test, for example a static test verifying propulsion motor operation is not necessarily less complex than a dynamic test examining right turning capability.

Testing conditions will need to take into account a number of different factors:

Simple: Despite the naming a simple test is not necessarily simple. In terms of set up or data recorded a simple test could involve as much work as a more complex trial. The term simple is used as these tests will typically not look at exploring edge or corner case conditions and nor will they involve forms of mechanical or electronic impediment. A test under this heading will simply be to understand whether a component or function performs to the prescribed level or meets expected performance requirements; for example a door operating system may be required to open and close the doors on a button command within a certain time frame.

Edge: Tests designed to investigate edge conditions are those where extreme (maximum or minimum) operating parameters are tested. This level is more demanding than a simple test as it is designed to investigate the boundaries of the system or component. Using the door operating system as an example, a test designed to meet edge case conditions may look at performing a series of tests at the maximum designed opening speed of the doors to investigate any weaknesses in the design.

Corner: Edge cases are only designed to test a system or component to the minimum or maximum design limits; corner cases on the other hand push the boundaries of the testing into areas beyond these basic design conditions. Corner cases are made more challenging compared to edge cases as they may also consider other variables that can contribute to exceeding design limits. Expanding on the edge case example of the door operating system, a corner case may be to investigate the maximum door opening speed plus 10%/20%/30% on design standards. The testing conditions set for an edge case test may also be made into a corner case with the addition of ambient temperatures of 40°C or -20°C or any other number of environmental conditions.

Mechanical: Systems or components that are designed to work as per design limits may be subject to unusual performance or total failure if they are mechanically impeded or hindered in some way. The result of this could be that a system fails to detect a potential mechanical malfunction leading to a dangerous situation or that a component fails in operation due to exceeding its design limits due to an outside mechanical influence. Considering the door operating system again, a mechanical impedance such as a force being applied to the door(s) from outside may cause the mechanism, control system or motor to become damaged and perform in a way that is unexpected. In addition, wear and tear to the system may lead the bearings and linkages to become stiff, damaged or worn which may enable to doors to come open while the vehicle is in use.

Electronic: Like mechanical impedances, electronic systems and components may also become damaged or faulty in use or may be subject to unusual performance due to outside influences, both malicious and benign. The result of an electronic impedance could be similar to the result of the mechanical example, however it is likely that a POD type vehicle will have a wider and more varied array of electronic sensors, control systems and power train components than a traditional vehicle. Again examining the door operating system, an electronic impedance may be that the sensor monitoring safe door operating conditions is obscured or damaged, or that the motor operating the door (if electric) may read incorrectly the amount of movement in the system allowing doors to be left open. In a more safety critical example, electronic hindrance may be the obscuration of cameras or sensors used for navigation and manoeuvring due to snow or heavy fog.

Cyber: Cyber testing level is in some ways similar to the five previous categories; however this test type does not exist in isolation. Instead cyber tests can impact systems and components across all of the previous categories from simple tests right through to corner or electronic impedance tests. Cyber tests should be aimed at determining the security of the POD in terms of a malicious hacking event or other external source that could render on board electronics inoperable or unpredictable.

2.2.3 Creating a robust test ecosystem

Outlined in Figure 1 and discussed above are the types and levels of testing that should be expected of any vehicle using publicly accessible areas, this should include autonomous POD vehicle types and subsequent generations of new vehicle types. Clearly there is a great demand on the testing for these vehicles; this is a huge task for the manufacturer but also an extensive and unwieldy issue for any legislative or third-party certification organisation.

Even creating a simple grid of the most basic tests that should be conducted can lead to a test plan running into the hundreds, this is understandably going to become a problem for any certification process as it will not be able to replicate every test under every condition while exploring every testing variable. There will need to be a method to reduce the burden of testing without oversimplifying the verification and validation regime outlined throughout this section.

One existing testing programme with many parallels to the issues being faced with autonomous vehicle certifications is the European New Car Assessment Program (EuroNCAP). In the same way autonomous vehicles currently have very little visible or external validation with which consumers can make educated choices, the same problem existed for traditional vehicle safety in the mid-1990s. The EuroNCAP programme is described in Ratingen et al. (2015).

At this time vehicles were not crash tested as routinely or as rigorously as they are currently and even when they were, the results and outcomes of these tests were not made available for consumers to examine or compare; testing was in effect behind closed doors. To break down this barrier an independent and transparent testing regime was developed based on the examination of real world crash data; the result of this was the development a series of crash tests that would reflect the real world situations and types of crashes that normal drivers were involved in. The testing was elective; manufacturers were not compelled to submit cars for testing; however it became apparent relatively quickly that demonstrating safety built user trust and increased sales, especially with the advent of the first 5 star scores and the easy comparison between performance of different vehicle models.

The result of this was twofold with two diametrically opposed effects. The first effect was that more people began buying better performing cars; this was a positive side effect of the testing. The second was that manufacturers, under pressure to develop 5 star cars, could optimise designs solely to give better performance in the specific EuroNCAP tests; this can be seen as a negative effect of the testing; in effect, designing to pass a test, not for overall safety.

The key lesson from the EuroNCAP example is that if you can prove something is safe, consumers will gain confidence to buy or use it.

3. Scenario Design

Test scenarios form the basis for all safety testing within the Capri project, and indeed are critical for wider AV safety assurance. A scenario in this context allows a safety hypothesis to be proven or disproven. For example, a scenario may be derived to test whether the POD is able to sense and stop for a pedestrian standing in its path. Each scenario when tested in the real world is in effect an experiment, and as such the combined outcomes of all the tested scenarios form a conclusion to the overarching question, "is the POD safe to carry out public trials"?

One of the overall aims of the Capri project was to develop a complete V&V framework for connected and autonomous vehicles. As has been shown, such a framework will by necessity need to comprise of many components, such as different types and levels of testing, to ensure is it robust and reliable. The challenge in Capri was to prepare a range of scenarios that are realistic, scale-able, repeatable, and representative of situations the POD will encounter.

A key building block of the overall framework will be the mixed method approach of carrying out both real-world and simulation testing. A simulated environment for testing is necessary as it is not feasible to achieve the number of miles of on-road testing that would be required to evidence the safety of PODs compared with human drivers. However, testing in the physical world is also needed to verify the accuracy of the simulation environment, and may also find issues that would not be apparent in simulation alone.

3.1 What is a scenario?

The term 'scenario' is a somewhat overused word in AV testing, as such it has evolved to mean a lot of different things. To clarify the terminology used in the Capri project the following section outlines how a 'scenario' is constructed and what types of information it contains. The common use of 'scenario' typically describes a composite view of a world detailing the interplay between different 'layers'. These layers describe fixed, static, or dynamic elements but can also contain information about 'actor' motivations and other transient factors which may influence actor behaviour.

The first step is to consider scenarios as experiments. An experiment is an overarching term used to describe the interplay between 'scenery', 'scenario', and 'actors'. As with all scientific experiments, each will have a defined reason for investigation in the form of a hypothesis or test requirements.

These layers are shown in Figure 2 and described in more detail below. For the purposes of this document the common term 'scenario' will be used throughout as it more closely aligns with work completed in other autonomous vehicle V&V processes.

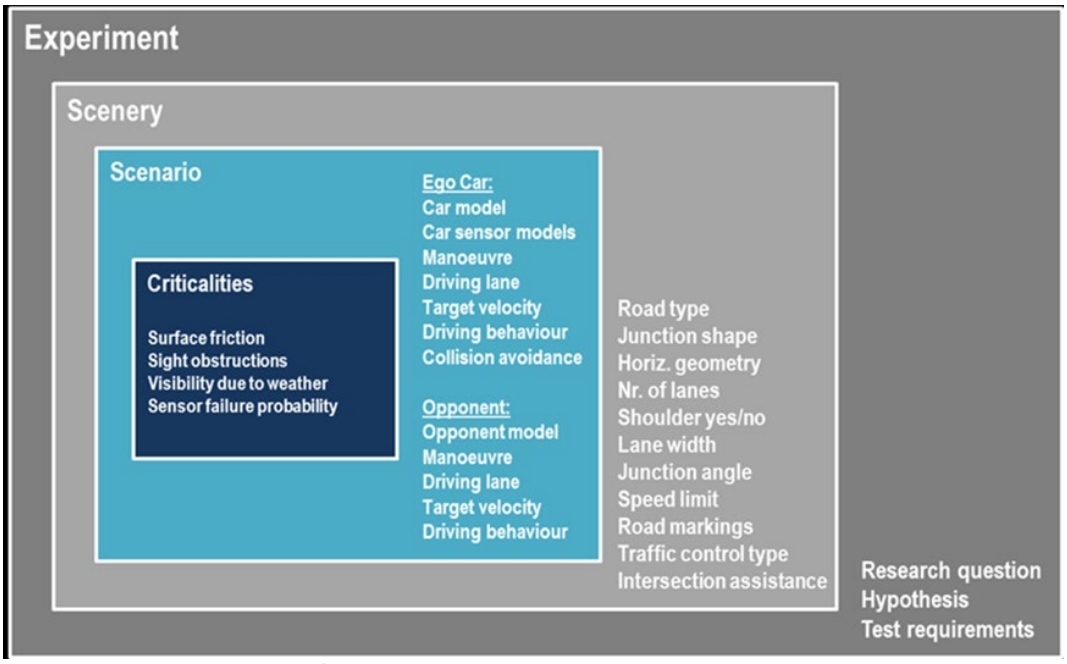


Figure 2: Layers for simulation experiments (source: Kienle et al., 2014)

Experiment: As with all experiments in any field of science, each autonomous vehicle test should be based on at least one, but preferably a number of the following points:

- **A Research Question.** The fundamental core that defines what we are attempting to discover. The research question defines what it is that we are looking to answer.
- **A Hypothesis**. Linked to the research question(s) but phrased as a prediction. The aim of the experiment is to prove the hypothesis to be true or false.
- **Test Requirements**. Can be either explicit or implicit.
    - Explicit; is the vehicle doing what it should do as described in its specification? However, some requirements may be hidden as claims or predictions; for example, our vehicle can do ‘this’ better than another vehicle**.**
    - Implicit; these are requirements that an end user will *expect.* These may not be explicitly defined but often cover ‘soft’ requirements such as comfort, performance, usability, security, etc/

Whenever a new research question is derived, hypothesis formed or test requirement identified, then a new experiment will be defined. This may mean that one piece of infrastructure (a simple T junction for example) may require a huge number of experiments to be performed, each focussed on addressing a different element of the points identified above.

A new experiment may also be formed in another way; as can be seen with subsequent levels in the schema (Figure 2) it may be necessary to define a new experiment when the parameters of a certain variable change to a degree which impacts the validity of a research question, hypothesis or test requirement. For example, a research question and hypothesis could be defined based around a vehicle navigating a T junction. If the parameters of the T junction were changed to such an extent that the research question or hypothesis was rendered invalid, then it may be necessary to define a new experiment to test this.

Within a particular experiment lie three layers which build together but individual describe the static, dynamic and critical elements within each experiment:

Scenery: describes the physical, static elements that constrain each experiment, and can be pictured somewhat like in a stage play. For transport simulations, the scenery will describe the physical layout and geometry along which the actors interact, and may also contain information on junction types, traffic control, lighting conditions etc. Exclusions to scenery exist where a static or transient element may play an adverse role in the experiment, for example reducing road surface friction/quality or changing other road user behaviour; these are classed as criticalities.

There may be a number of different experiments using the same scenery. Scenarios and/or criticalities will determine how an individual experiment is defined.

Scenario: describes the types, manoeuvres, paths, speeds and interactions of all actors at play within an experiment. An actor may be a pedestrian, cyclist or any other vehicle type, including the POD. The minimum number of actors is one while there is no maximum. An actor may have several different attributes, for example a start position, a given speed (or range of speeds) along with an intended path and/or intended manoeuvre. Although broadly similar to a description of a ‘road user’ there are exclusions such as animals or parked vehicles.

Criticalities: parameters that can adversely affect the performance of all actors (including the POD) across a wide range of elements. The parameters described are typically transient (they can come and go), volatile (appear without warning and disappear without trace) and unpredictable (their effects on the experiment can differ).

3.2 Challenges with scenario generation

In order for AVs and PODs to become more widely accepted and capitalise on the benefits that have been predicted, a logical process for safety validation using best practice is required. At present there exists no standardised, systematic, or structured methodology for the definition of AV test scenarios. Currently the overall direction behind this scenario generation is highly technical and makes little if any reference to established principles behind designing a better product.

The technical approach to designing vehicle test scenarios is in part driven by the historical methodological steps behind current automotive testing protocols. This methodology is problematic because though it may be well defined for conventional vehicles it is unable to provide a suitable basis for the development or testing of new innovative transportation solutions, such as PODs, since they do not conform to traditional vehicle performance or usability expectations.

The traditional process of designing test scenarios for manually driven vehicles typically involves the use of collision data to identify an individual, or range of, scenarios that encapsulates the real world 'worse case'. This process is almost certainly misrepresentative of the AVs being developed and presents a number of issues and gaps in knowledge initiated by the small number of vehicles with autonomous features that exist in the overall vehicle fleet. For example, it pays little regard to the new relationship other road users will have with vehicles with autonomous capability, such as autonomous pod style vehicles operating in pedestrianised areas.

Furthermore, while we can design scenarios that evidence has shown humans can fail to deal with, there is no guarantee that a machine will fail in the same way. Indeed, the expectation is that they will avoid these failures, but what new failures of their own will they have that will need to be tested?

3.2.1 Physical testing mileage

Any future AV testing programs will need to find a solution to determine whether or not are safe enough for public use. Such programs will need to find a way of determining that they are capturing all events in all conditions and across all variations; the problem with using a real world approach for this is that, for example on EU highways, it is possible to drive manually for up to 200 million kilometres between fatal accidents. If this is a good surrogate for an unexpected and rare event, then on-road verification of AVs will require billions of test kilometres to encounter all eventualities. The bottom line is that for a robot to prove it is a better driver than a human it will need to drive billions of perfect test kilometres – perhaps even more.

Realistically the only way of ensuring a vehicle covers the required mileage (and therefore events) would be to move testing into the virtual world. This step, although beneficial and convenient in terms of determining whether safety standards, may not be enough to allay users concerns and inevitably will fail to capture some of the fine details that are only found in real life; this is where some level of visible and understandable physical testing is required.

3.2.2 The complexity of the transport system

Automated driving systems used on AVs should not be likened to current automotive software. The huge complexity determined by the variety of driving tasks, vehicles, environments, and behaviours of other users results in an overall system functionality that is difficult if not impossible to describe in a specification document. Take for example a relatively simple driving task such as entering a roundabout. How would a specification document handle the complexity that human drivers can (normally) cope with? Setting aside the 'normal' roundabout scenario, which is complex on its own, humans are also trained to deal with any number of abnormal scenarios, for example:

- There is a broken-down vehicle at the entrance to the roundabout.
- A vehicle on the roundabout blocks the entrance to the roundabout.
- An emergency vehicle on response travels the 'wrong way' round the roundabout.
- There is a grey and dirty truck at the back of the queue entering the roundabout and it is raining.

The AV must respond to a huge number of factors; it must be able to detect, classify and locate every actor in the visible scene, it must be able to read the whole road layout, it must be able to detect and interpret signals and road markings. Having a prescribed specification will likely not be sufficient as different decisions may be required for each possible combination. It is also not possible to write a test script or define a test scenario for each eventuality as the number of combinations is almost incalculable. This complexity is why current test processes are unlikely to provide complete safety assurance.

It is not just the physical environment which has near infinite variation, drivers and other agents will also act in ways which are unpredictable or do not fit into the AV test scenarios. In other words, it is not practical to build a robot that only works if all other cars are robots, human behaviour must be incorporated to have a safe self-driving vehicle. With simulation scenarios there is a need to optimize for human behaviour on the road and the optimal way of achieving this effectively is with real training data from real world conditions or data.

This problem could become more prevalent with more powerful simulations using more and more variables to represent the real world. In this case each parameter can be adjusted arbitrarily and potentially falsely, especially if this is conducted without some expectation or intuition as to the likely or approximate outcomes; simulation could in effect predict whatever is required.

3.2.3 Achieving and measuring the very low failure rates required

Ultimately any performance requirement that comes out of the AV testing process should be measurable against a set target or range or values. Many of these requirements will be relatively easily assessed as they will be in the form of outputs from sensors or processing, i.e. was the sensing performance within the accepted tolerance range during the test. However, who will define and set these targets, and ensure that AVs comply with them, is yet to be decided.

Many of these targets are as yet undefined for Capri but one of the key outcomes is likely to concern overall vehicle safety and whether or not the vehicle is significantly safer than a human driver in a given condition. The problem of measuring this is that the failure rates for human drivers may already be exceptionally low; take the example of a right turn into a minor road across traffic. For human drivers there will be a measurable number of failures, this can be seen from the national accident data and, framed against the context of how many successful right turns are completed, this will provide the failure rate which the AV must better.

The problems come in determining; i) what constitutes a failure, ii) what level of reduction proves safety, and iii) how many tests the autonomous vehicle must complete successfully (physically or virtually) to prove its safety advantage.

3.2.4 Verifying systems containing artificial intelligence or machine learning based components

As AV system complexity builds almost exponentially in order to achieve the ever-increasing demand for autonomy, so the processing and decision making used in the vehicle becomes ever more advanced. These advances are beneficial to opening up the performance envelope of the autonomous features but could be detrimental to human understanding of the systems.

The move to electronic systems from the predominantly mechanical operating systems of past vehicles placed much fewer design constraints compared to the physical systems it replaced. In response, software quickly grew in complexity, making it increasingly more challenging to fully understand a given system. Autonomous features provide an additional spectrum of software behaviour which may or may not be easily understandable to vehicle engineers or designers; even a vehicle with prescribed autonomous characteristics or behaviour (i.e. it just follows commands from lines of code) may perform differently to how it is specified given the variety of diverse situations and environments it is likely to encounter.

It is estimated that a modern vehicle with some autonomous capabilities will have approximately 150 million lines of code used to operate all vehicle functions. One line of bad code within this can make the vehicle perform outside of design requirements and it could be an almost endless task to identify the error. Subsequent design changes or over-air updates to autonomous software may also have significant ramifications to the vehicle behaviour and it will be difficult to know how these design changes operate under all conditions.

Artificial Intelligence (AI) and/or machine learning technologies also bring a new level of complexity to verifying AV systems. These technologies do not necessarily follow prescribed code but will actively change their operations and modify vehicle behaviour based on what stimulus the systems have been given in driving. Knowing how an AI system interrogates the world it views, processes the information and ‘learns’ from its mistakes is a challenging field of science in itself let alone when applied to the already complex topic of autonomous vehicles.

3.2.5 A new type of vehicle

The Capri project is tasked with bringing marketable and transferable ‘last mile’ autonomous POD vehicles to the real world. Currently there isn’t a vehicle type classification for a ‘pod’, with most AVs typically built on an existing platform from a classified vehicle; for example, a heavy quadricycle or heavy/M1 passenger carrying vehicle.

The pod used in Capri is relatively compact in terms of external measurements while maximising interior space. Overall dimensions are similar in length to a small city car however it is considerably narrower enabling it to offer ‘dual mode’ operation; meaning that it is able to travel in both pedestrianised areas and on public roads. The vehicle seats 4 although unlike traditional vehicles these are in ‘conference’ format (two pairs of seats facing each other).

By SAE definition the Capri Pod is level 4 automation. This level of automation is predominantly due to the lack of driver controls in the vehicle; there is no physical driver present or a way of handing back control to a driver, however it will only operate within a strict ODD. All of these elements provide a range of issues for the determination of test scenarios for the Capri POD.

As the vehicle does not fit easily into an existing vehicle category it is unlikely that there will be a lot of crossover between this new vehicle type and vehicles already using the public road. Traditionally, for investigating a specific road safety aspect for a particular vehicle type, the national or in-depth accident data would provide a large amount of useful and relevant information. However, these datasets contain few, if any, directly relevant cases for fully autonomous vehicles; the Capri POD doesn’t look like vehicles that populate the dataset, it has different vehicle dynamics compared to traditional passenger vehicles, the technology used in the POD is unfamiliar to a lot of current vehicle systems, the vehicle may be operating in areas that motorised vehicles typically aren’t allowed to, and they may not have the traditional raft of driver behaviour issues associated with them.

Equally, finding relevant standards or testing protocols will be more challenging as these will likely have been derived based on existing technology. It is therefore likely that some level of extrapolation or interpolation between the types of tests and scenarios that already exist and what the POD can achieve in testing will be required.

3.2.6 Feedback loop

Current AV verification and validation plans, although limited in number and completeness, appear to collectively have one major element missing: any form of feedback loop from the real world.

Feedback in a system of this type is incredibly important as without new knowledge about how AV systems are working and failing it will be difficult to achieve full safety assurance for them. New types of events, interactions and errors will be uncovered at a rapid rate with each full-scale implementation of new systems, and it is important that these are captured and disseminated amongst practitioners in order to improve the collective safety. The traditional passive safety systems employed in current human driven vehicles do not necessarily suffer the same high levels of secrecy and confidentiality that is required to maintain AV technological advancements and market share. However, access to the crash performance and subsequent sharing of results through in-depth data has undeniably improved knowledge of the subject and performance of the vehicles.

As POD and wider AV trials start to occur, having a reliable mechanism for capturing and using field data is especially important in order to share lessons learned. However, it is anticipated that collecting and interrogating this information for gaps and defects will be potentially very difficult; this is especially the case with electronic systems, which on an AV are manifold. Added to this is the complexity of how the information is held within the vehicle systems, these may include those related to software programming, intermittent electronic hardware faults and electromagnetic disturbances amongst others, all of which may not leave physical evidence to aid investigations into observed or reported unsafe vehicle behaviours.

3.3 The Capri POD specification

The Capri pod is intended to operate at a SAE automation level 4. It is the next generation evolution of the POD design already in use in dedicated guideways at Heathrow airport and manufactured by project partner, Westfield Technology Group. The GATEWAY[1] project evolved this design into one capable of operation across a ‘campus’ site but not on the public road. Capri will take this a step further with the aim of providing a ‘last mile’ transport service for passengers, capable of operation both on pathways and pedestrian areas but also on defined sections of roads. The project refers to this concept as a ‘dual mode’ POD.

[1] A project part funded by the Centre for Connected and Autonomous Vehicles (CCAV). Gateway’s main aim was to carry out a user research trial, with members of the public invited to ride a pod on a 3.4km off-road route in Greenwich and share their opinions https://gateway-project.org.uk/.

The Capri POD main characteristics are detailed below in Table 1.

Table 1: Capri POD characteristics

Attributes	
Passenger Capacity	4
Gross Vehicle Weight	1430kg (980kg unladen)
Maximum Speed	24mph (40km/h)
Turning circle	13.5m
Wall to wall turning radius	6.75m
Emergency deceleration rate	~3m/s^2
Battery range	6-8 hours
Dimensions	
Overall length	3.7m
Overall width (excl sensors)	1.47
Height	1.8m
Body height	1.4m
Front/rear overhang	0.778m
Wheelbase	2.145m
Track width	1.05m
Minimum ground clearance	0.134m
Min Battery state	50% to 100% charge
Tyre inflation	Between 26psi and 32psi
Tyre condition	>1.6mm tread depth

Operational design domain

In general, the Capri ADS is designed around some high-level domain requirements. These are outlined below to frame the scope of the scenario generation. The creation and specification of an ODD nominally implies that the vehicle using the ADS is operating below SAE level 5. If the vehicle was to have an SAE automated driving level of 5 then an ODD would be much less restrictive as, by definition, the vehicle would have the same mobility as a human driver (i.e. the ODD would be defined as everywhere a human could drive the same type of vehicle while still allowing for some restrictions such as private land etc.).

Overall domain requirements for the Capri POD:

- Dual mode (both on and off road).
- Accessible to all users.
- Detect and respond to all anticipated hazards it may encounter, for example vehicles / bicycles / pedestrians / animals / road furniture etc.
- Follow pre-planned routes.

The table below gives more detail on some of the operating conditions which are expected to be in and out of scope for the Capri POD. It is noted that these are the capabilities as understood by Loughborough University, however these have not been verified by Westfield.

Table 2: Capri POD operating conditions

Property	In scope	Out of scope
Area and road type		
Urban	✓	✗
Extra-urban	✓	✗
Rural	✓	✗
Motorway (and 'special' roads)	✗	✓
A roads (including PRN and SRN)	✗	✓
A roads (excluding PRN and SRN)	✗	✓
B roads	✗	✓
Unclassified	✓	✗
…		
Road elements and access		
Pedestrianised areas	✓	✗
Private roads	✓	✗
Publicly accessible roads	✗	✓

Private parking	✗	✓
Service yards	✗	✓
Emergency access points	✗	✓
Tunnels	✗	✓
Enclosed parking structures	✗	✓
Roundabouts	✗	✓
Signalised junctions (traffic light controlled)	✗	✓
Uncontrolled junctions (give way/stop)	✗	✓
Closed private campus sites	✓	✗
Shared use paths	✓	✗
Construction zones	✗	✓
…		
Road descriptions		
Slope (driving)	<10%	>10%
Slope (stopping/parking)	<10%	>10%
Slope (setting off)	<7.5%	>7.5%
Approach depart angle	<15°	>15°
Accessible width	>1.5m	<1.5m
ACS Navigable width	>3.5m	<3.5m
Vertical edge	<30mm	>30mm
Curve radius	>6.75m	<6.75m
Negative height object	<30mm	>30mm
Contaminants	✗	✓
…		
Environmental conditions		
Seasonal restrictions	All seasons	
Time restrictions	All times (unless specified by site owners)	
Min Temperature (operational)	>0°C	<0°C
Min interior temperature (comfort)	>16°C	<16°C
Min Temperature (parking/charging)	>5°C	<5°C
Max temperature (operational)	<30°C	>30°C
Max interior temperature (comfort)	<30°C	>30°C
Max Temperature (parking/charging)	<45°C	>45°C
Standing water depth (wading)	<30mm	>30mm
Surface temp	-10°C - +40°C	<-10°C or >+40°C
Surface ice	✗	✓
Surface snow	✗	✓
Falling snow	✓ within sensor capabilities	✗ Unless sensors deteriorate
Precipitation	✓ within sensor capabilities	✗ Unless sensors deteriorate
Fog/mist	✓ within sensor capabilities	✗ Unless sensors deteriorate
Windspeed	<30mph	>30mph
Humidity	Are there limits	
Light conditions	Are there any	
…		
…		
Manoeuvres and operation		
ACS Forward travel	✓	✗
ACS Right/Left turns	✓	✗
ACS Reversing	✗	✓
ACS stop for hazard	✓	✗
ACS swerve for hazard	✗	✓
ACS pull into stop	✓	✗
….		

Traffic density	*Unclear how to define*	
Load inside	✓ (450kg)	✗
Load external	✗	✓
Trailer or carrier use	✗	✓
…		
…		

3.4 User centred design methodology

Scenarios developed for Capri testing were derived through a user centred design (UCD) approach, and more specifically through the application of the Design Double Diamond, a well-established design architecture. This approach is described in greater detail in Reed et al.

Typically, the overall direction behind scenario generation is purely technical and makes little if any reference to established principles behind designing a better product. Products designed on the basis of technical merit or theoretical performance are often highly capable but prone to errors of implementation, use and 'fit' to the real world.

The philosophy behind UCD is that at all key decision points in the design process, the needs, preferences and characteristics of the user are prioritised over other design drivers such as aesthetics or technical 'push'. By using a UCD approach for Capri scenario generation there will be an increased probability of the product succeeding commercially since there will be a greater confidence that the product or service meets the required needs.

Within the AV testing domain there are traditionally two ways of validating vehicles or functions through scenarios; i) by generating these scenarios from a theoretical basis, for example basing the testing program on a list of assumed accident or manoeuvre types (often derived from simple manoeuvre vectors), and ii) from a practical retrospective 'beta-testing' basis - for example putting the vehicles into service to see what they encounter. These two methods both offer a 'top down' approach to scenario generation which begins with the foundation of a manoeuvre type (a left turn for example) before defining the types of vehicle behaviour and possible permutations available for this manoeuvre.

As described in the previous section, for all types of AVs, the lack of process in scenario generation is compounded by minimal real-world collision or near miss data meaning that much like other innovative consumer products, where rapid return on investment is sought, testing and development is undertaken on technical and hypothetical constraints alone. For Capri a more systematic approach is employed which aims to understand factors in scenarios on an evidential basis. This replicates entirely the UCD philosophy employed successfully in other areas of technology, but which is not apparent in the technically dominated AV development sector.

The different data sources outlined in this section go some way to evidencing the application of the UCD design approach for scenario definition.

3.5 Data sources for scenario generation

Building a scenario for simulation requires knowledge and input concerning many parameters including vehicle systems, sensors, road environment and driving behaviour. The space of possible scenarios is spanned by many dimensions, such as road geometry, behaviour of the driver and other participants, weather conditions, vehicle characteristics and component faults. For Capri, a combination of different data sources were required in order to develop scenarios that would best assess the safety performance of the POD.

3.5.1 In-depth accident data

For the Capri project, in-depth road accident data collected as a part of the OTS and RAIDS studies was analysed. The Road Accident In-Depth Studies (RAIDS) programme and associated database was commissioned by the UK Department for Transport in 2012, to collect new in-depth data and additionally to consolidate data gathered from previous in-depth accident investigation programmes (including the On-The-Spot (OTS) study) dating back to the year 2000. The database has subsequently grown with further investigations since 2012, and investigations are still on-going.

Accident investigation teams are responsible for collecting detailed information at the scene of the accidents which is later enhanced through liaison with emergency services, hospitals and local authorities. Access to this data is provided through permission from the Department for Transport.

The RAIDS database contains several hundred coded variables covering all aspects of the accident scene, vehicles, occupants, injuries and contributing factors. It is therefore possible to identify collisions that fit within the ODD of the POD and challenge the autonomous control system (ACS) to perform with a better outcome than the human driver. Clearly the collision data does not currently contain accidents involving actual PODs for reference; however, by using a carefully defined ODD, relevant cases from within the in-depth data can be identified.

3.5.2 Naturalistic driving data

By definition, accident data concerns interactions which resulted in an accident occurring. However, there are many other occasions when interactions between road users are hazardous, and an accident is only narrowly avoided; these are still interesting interactions as they provide the bigger picture of everyday driving situations and potential challenges. This type of data is captured through naturalistic driving data. Naturalistic driving, also known as naturalistic observations, is a relatively new approach within transport safety research. Such studies offer an objective view of driving behaviour in both conflict and non-conflict situations using video data as an evidence base.

UDRIVE is a recently completed research project funded by the European Commission which collected naturalistic driving data over a 2 year period in 7 EU member states including the UK. Whilst a high proportion of this data will be outside of the ODD for the Capri POD use case, there are some data that has direct relevance and that can be reviewed to build up a good picture of the likely interactions between PODs and pedestrians. For example, several vehicles within the study regularly travelled across Loughborough University campus. This environment could be considered akin to a low speed semi-pedestrianised environment typical of the ODD for a POD.

Analysis of the UDRIVE data within the project showed that speed, infrastructure, and surprise were the three components most associated with conflicts involving pedestrians with at least 2 of these being present in all incidents involving pedestrians (Jansen et al., 2017). Reviewing the UDRIVE data, or other naturalistic studies, could provide a range of pedestrian behaviours for input into the scenarios, thereby increasing the credibility of the behaviour model.

3.5.3 Established test protocols and standards

Although previous sections have highlighted the problems with trying to transfer established test protocols for manually driven vehicles to the AV sector, that does not mean there is not useful information that can be drawn upon.

For example, reference has already been made to the EuroNCAP programme. Of interest for Capri, EuroNCAP have a range of test protocols to assess Autonomous Emergency Breaking (AEB) technology installed in vehicles. These protocols include scenarios for car-to-car, car-to-pedestrian[2] and car-to-cyclist, as well as a pedestrian emerging from behind obscuration, all of which are relevant for the Capri POD. Whilst these cannot be carried out word for word in the Capri testing, they are easily adapted to suit the POD's ODD (e.g. performing the test at a much lower vehicle speed than is typically specified within the EuroNCAP protocols).

Safety standards for trucks are also useful to reference. Though trucks may not be what first comes to mind when the POD is considered, they are prone to a number of accident scenarios involving pedestrians and cyclists and as such a number of safety regulations have been introduced to better help the driver avoid these accidents. For example, UNECE Regulation 46 specified a minimum vision standard for drivers of HGVs where mirrors have been replaced with camera monitors. Test protocols were designed to test this, including a scenario where a truck has to make a left turn and avoid colliding with a cyclist that is coming up alongside it. Although the POD has no driver, the same scenario can be used to assess whether the POD sensors can 'see' the cyclist and act appropriately to avoid a collection.

Industry standards, such as those developed by the International Organization for Standardization (ISO) are also relevant and their use is promoted on a global scale as opposed to a national law or standard. In particular, ISO 26262: Road Vehicles - Functional Safety is commonly referenced in

[2] https://cdn.euroncap.com/media/26997/euro-ncap-aeb-vru-test-protocol-v20.pdf

conjunction with AV technology development. This voluntary industry standard is the first comprehensive and voluntary automotive safety standard that addresses the functional safety of electrical and/or electronic (E/E) and software-intensive features in today's road vehicles. ISO 26262 addresses the needs for an automotive-specific international standard that focuses on safety critical components. ISO 26262 is predated by, and a derivative of, IEC 61508. IEC 61508 is the generic functional safety standard for electrical and electronic systems. Although this is less relevant for generating specific test scenarios, it does promote safety of individual systems within the POD and should be considered as part of the overall V&V framework.

3.5.4 Site surveys

The three data sources already described are more generic to developing scenarios for any AV, or, for Capri purposes, for any use case of the POD. However, these only form one part of the development process. It is also important to consider the very specific use cases the POD will be demonstrating during the project, and develop scenarios based on hazards and conflicts that will be found at the public trial sites.

Public trials will take place at the Mall at Cribbs Causeway and Queen Elizabeth Olympic Park (QEOP). These two sites provide a range of common hazards as well as some that are unique to each, and a thorough site survey was completed at each location.

Queen Elizabeth Olympic Park

Using the insight from the previously analysed data sources as well as expertise in human factors and the investigation of road accidents, the team were able to visit QEOP to analyse how the other park users behave and predict how they would interact with the pod. This was supported by information provided from staff at LLDC such as footfall in different areas of the park.

Based on the expected use of the POD the scenarios are primarily designed with non-motorised road users such as cyclists and pedestrians in mind. In addition to the hazard specific information an analysis of hazard route and path choice was conducted to understand the likely pod versus vulnerable road user interactions. A series of junctions were selected for site surveys as they represented the most complex pod manoeuvres (e.g. a tight 90-degree turn) and the highest frequency of other users. This data helped inform the scenario design with respect to the positioning of paths for moving hazards, the angle of approach and the velocity of the hazards.

The trial planned to only use the north park, which generally has less foot traffic than the south park, and contains mostly wide, open paths which will make it easier for other users to avoid the POD. The main areas of concern were the Timber Lodge café and Tumbling Bay play area. These areas had the highest concentration of pedestrians, and in particular high numbers of children which present a higher risk. Children are smaller and often dart to and fro in erratic paths. They may also pay less attention to their surroundings than an adult would or have lower perceived risk of danger. Given that this is a park, it was not uncommon to see parents allowing their children to play freely, not expecting that they would need to watch for dangers as they would on road.

In terms of hazards, generally pedestrians are present throughout the park in greater numbers than other hazards. the variety of movements also appeared to be challenging with multiple and unpredictable walking paths seen. These movements are often as a result of 'desire lines'[3] being created or the nature of pedestrians being able to access areas where cycle users cannot, for example taking short cuts through planted areas.

Cycle traffic throughout the park is perhaps more predictable based on the results from site surveys. Routes tend to be more established, for example using the wider paths linking destinations on the park and movements tend to be along the paths rather than across or at angles to the path, i.e. cyclists do not tend to cut across established paths from grassed areas. Cyclists present a more challenging element to the POD in that they travel substantially quicker than pedestrians. Typically, leisure cyclists appeared to travel at between 8mph (3.5m/s) and 12mph (5.3m/s), however this is also a popular commuter route to Here East and most commuters were noted to travel faster, perhaps as fast as 20mph (8.9m/s) on wider paths.

[3] A natural path created by human traffic. The path usually represents the line of least distance or most easily navigated route between locations or adjacent paths.

Finally, it was noted that Bird scooters had begun operating within the confines of QEOP. Scooter riders present yet another different challenge to the POD, with e-scooters capable of travelling up to 15mph but generally weaving between other users more than cyclists were seen to.

The Mall Cribbs Causeway

The second location for public trials is at Cribbs Causeway. This is a large out-of-town shopping area including adjacent retail parks that also features an enclosed shopping centre known as The Mall. The proposed public trial will use publicly accessible land (paths, carparks and service roads) immediately adjacent to the Mall building.

Preliminary site visits to this location identified similar hazards to those identified at QEOP, those being mainly pedestrian hazards. Although cycle provision is situated around The Mall it is expected that cycle traffic will be significantly lower than QEOP considering its location, access and use.

One particular hazard identified was pushchairs. Much like other hazards these cover a wide range of sizes and construction however they can present an insubstantial or lower density object compared to a pedestrian for example. Pushchair traffic was significantly more widespread than that seen at QEOP and this hazard was also restricted to narrower pathways which could lead to closer passing distances between the POD and pushchairs.

Because the main retail site is accessed on all sides from large surface car parking, the majority of pedestrians will be crossing into the Mall and therefore going directly across the path the POD, as opposed to in the same direction as it which will be more common at QEOP. There is an additional hazard of pushchairs (or other lower hazards) emerging from behind raised planters and hedges bordering the car parks. One particular scenario is made more challenging with the placement of a lift shaft adjacent to one of the Mall entrances. This obscuration will result in adult pedestrians being obscured from an approaching POD.

The design of the Mall and the park surrounding the main building also necessitates the pedestrian paths to negotiate tight bends around the corners of the building. At these locations smaller pathways radiate outwards leading to car parking or onto elevated footways to other retail sites. This design creates an additional hazard when a POD negotiates the main pathways near the Mall building as common pedestrian routes will interact with the pod when it is turning around a blind corner.

As an example, Figure 3 shows a segment overlaid on a representative corner at The Mall (yellow line) that could be replicated in testing. An example of a POD path through the turn (blue line) is shown alongside two common pedestrian paths that were observed during the site observations. Path B illustrates a path most likely to be taken from the North/South path adjacent to The Mall into the Southern parking structure and Path A leading across the walkway at the furthest point of the turn into the Western parking structure.

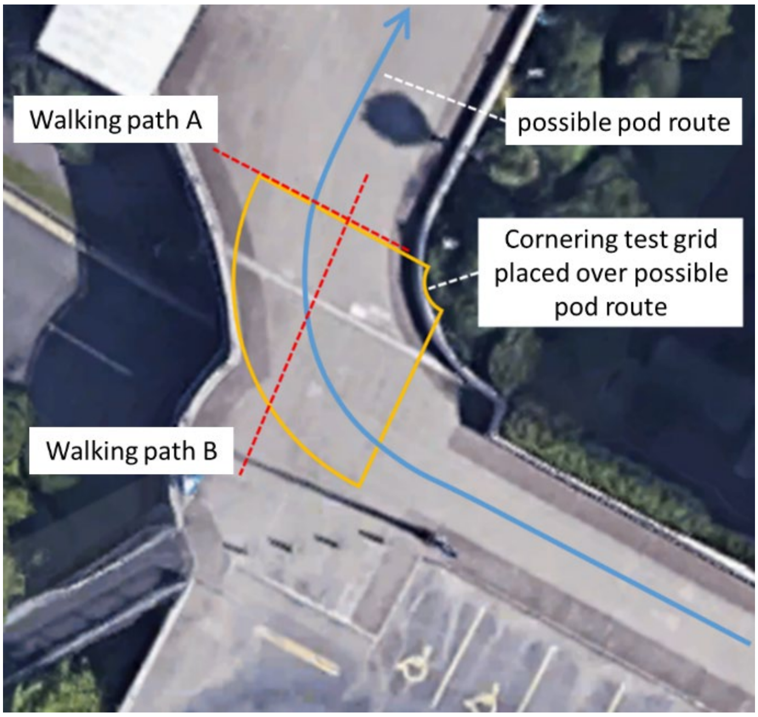


Figure 3: Identification of corner radius and pedestrian paths for cornering tests

3.5.5 Insurance and landowner requirements

Finally, it is also important to consider any requirements that insurers (in this case AXA) and/or landowners (in this case The Mall and LLDC) have on the safety case before they allow the public trials to take place.

Generally, such requirements are on the ODD and general management of the trial rather than the safety performance (e.g. maximum speed allowed within the park is 5mph). However, one issue that was highlighted was the need for a minimum safety performance in terms of low-level hazard detection. Low level detection (and negative height detection i.e. holes in the road surface) is a particularly challenging area for AVs so it was necessary to design scenarios with suitable low height objects with which to test the POD while also making the pod development achievable.

Small objects, or those which in one dimension are ‘low height’, are particularly difficult to determine and define. It was noted that during site visits to QEOP that dogs were often seen off leads, however dogs cover a wide range of sizes, only some of which could be considered a low object. Ducks were also present quite often, and, although it is not against the law to run over a duck or similar wildlife, it would certainly not be desired in the trials. Concerns were also raised that a small child could fall into the path of the pod therefore suddenly presenting a low-level hazard.

A composite approach was identified which covered a range of different heights while also ensuring that tests could be repeatable and replicable. The most serious incident of a child falling into the pods path was to be covered by a child sized mannequin while a Trunki[4] provided a surrogate for dogs and other low objects.

[4] A Trunki is a wheeled plastic hardcase suitcase designed for a child to ride. In this case it provided a robust object that could be struck without damaging the pod or the object.

4. Virtual Testing Protocols

The primary aim of scenario generation was to provide inputs into the simulation testbench. In some ways, virtual testing can be considered more generic than the physical testing, with the testbench architecture and scenario design being the key aspect, for which any ACS could be plugged in. A goal in the Capri project was to create a methodology and framework for virtual testing, which would be tested on the Capri POD but that was transferable to other AV types. Therefore, the scenarios developed needed to be relevant to Capri but the methodology to create them needed to be scale-able and adaptable.

In practice, using the UCD methodology described in section 3.4, and drawing on the range of data sources described in section 3.5, it was possible to develop a range of scenarios, covering simple, edge and corner cases.

A very simplified version of the overall virtual testing process is shown below in Figure 4. This is intended to show the high-level flow and does not capture the intricate detail of areas such as the testbench and assertion testing which are described in reports from other partners.

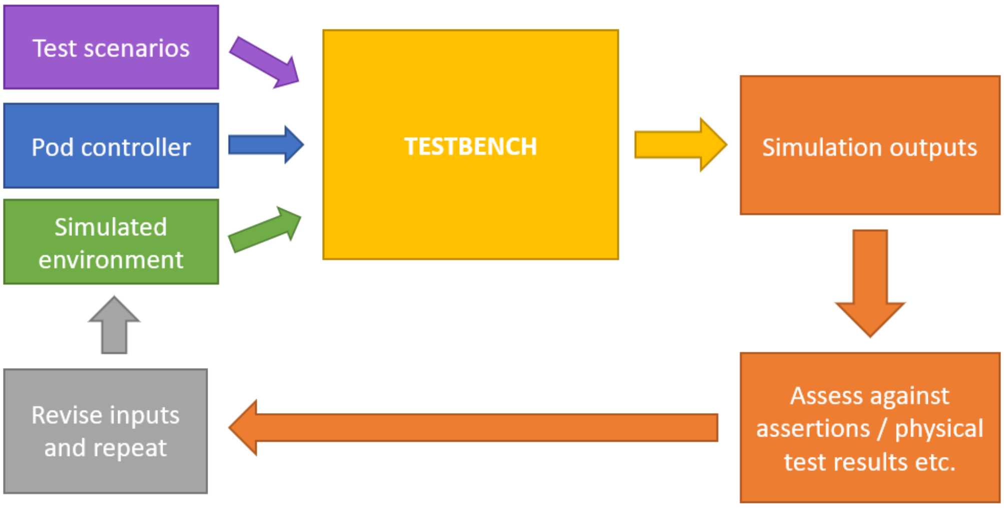


Figure 4: Virtual testing process flow chart

A key element of this process, as already alluded to in earlier sections of this report, is the need for a feedback loop. AV testing, certainly in its early stages, will not be a linear process. Through the early testing new insights will be gained, and these lessons should be fed back into the design process to refine both the ACS and scenarios, and the entire process repeated until no new lessons are learned and optimum safety performance is achieved.

Assessing the performance of the POD in each scenario is also a critical part of the process. This can be done via a variety of methods, such as comparing results of identical tests carried out in simulation and in the physical world, or by measuring performance against a series of 'assertions', i.e. expected behaviour. The complexities of this are discussed elsewhere in the project, however one assertion for example would be that the POD should always avoid impact with other actors.

4.1 Test parameters

One of the benefits of testing in a virtual world is that all parameters can be controlled for. Some factors, such as weather conditions, cannot be controlled in the real world, which can make it difficult to perform the exact same test as specified each time it is run.

For Capri simulation experiments, scenarios are described with relatively light detail compared with what might be expected. The purpose of the Loughborough input was to provide the overall picture, the basic "scenario" as described earlier, detail any fixed parameters, and specify ranges for variable parameters. The testbench would then be able to randomise these parameters within the constraints and perform any number of permutations and combinations of those values.

With this method, a single scenario (or experiment to use the proper term) provided by Loughborough could yield 10s or even 100s of individual simulations if all combinations of values were assessed. If this were to be attempted in the real world it would take incalculable hours, but the benefit of using a virtual world is that this can be achieved in a far shorter timeframe. For key scenarios however, it was also possible to provide specific values of the variable parameters that should be assessed together in order to engineer the critical event, i.e. to provide the exact start positions, speeds, timings and paths of each actor that would result in a collision if no action was taken by the POD.

For early scenarios, which were simple in terms of actor paths but required precise timings to achieve the desired interactions, a more technical presentation was used involving to-scale drawings set on a 50x50m grid, and tables of value combinations for variable parameters. Figure 5 and Figure 6 below show excerpts from a scenario where pedestrian(s) move out from behind a wall to cross the path of the POD, to replicate the challenging lift shaft area seen at the Mall. The calculations allow for multiple pedestrians in convoy with varying separations.

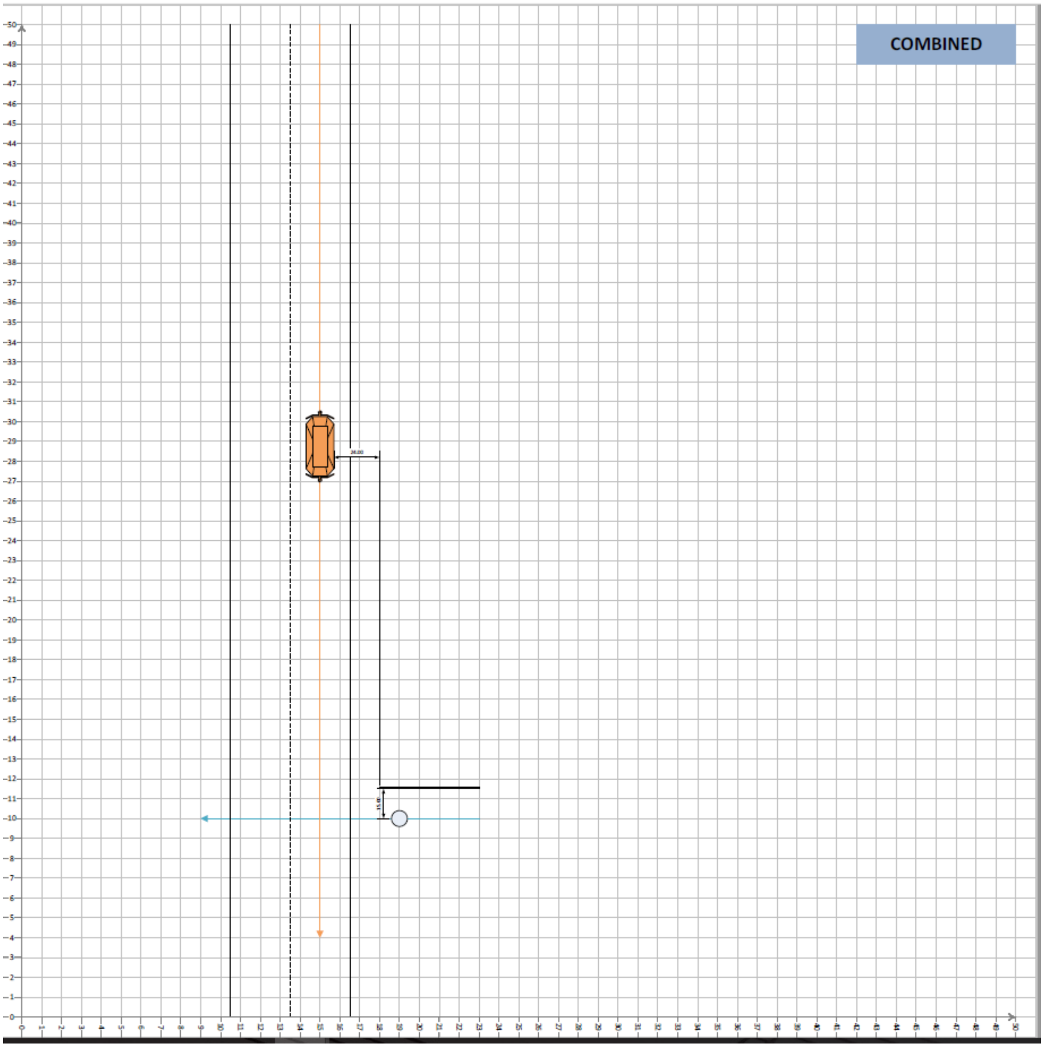


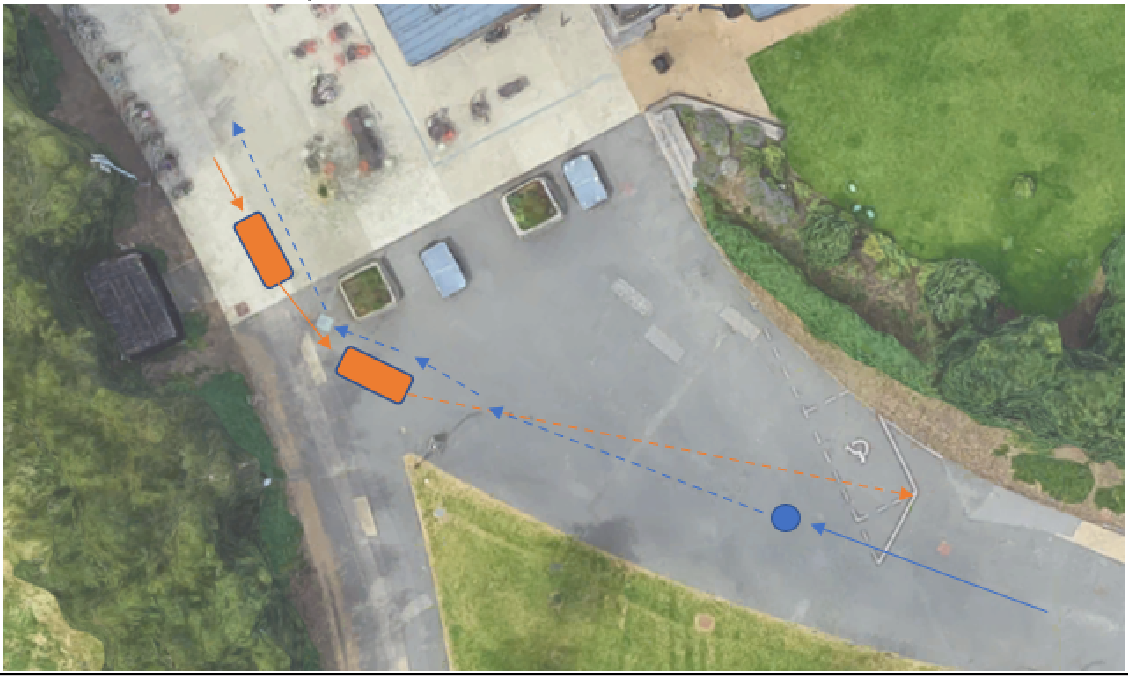
Figure 5: Example scenario drawing

Sub-Experiment A: Pedestrian movement from Left to Right of pod path		Lower bound	Upper bound	Priority 1	Priority 2	Priority 3	Priority 4	Priority 5	Priority 6
Actor Ref: 1	Start time (s)	T0 (fixed)		T0	T0	T0	T0	T0	T0
	Speed (mph)	5 (fixed)		5	5	5	5	5	5
Actor Ref: 2	Start time (s)	T0 (fixed)		T0	T0	T0	T0	T0	T0
	Speed (mph)	2.2 (fixed)		2.2	2.2	2.2	2.2	2.2	2.2
Actor Ref: 3	Start time (s)	T0+2.5 (fixed)		No actor 3	T0+2.5	T0+2.5	T0+2.5	T0+2.5	T0+2.5
	Speed (mph)	2.2 (fixed)			2.2	2.2	2.2	2.2	2.2
Actor Ref: 4	Start time (s)	T0+2.5	T0+10	No actor 4	No actor 4	T0+5	T0+6	T0+7	T0+8
	Speed (mph)	2.2 (fixed)				2.2	2.2	2.2	2.2

Figure 6: Example scenario timing calculations

However, within the course of the project, it soon became apparent that for the more natural real-world scenarios it is difficult to describe precise road layouts and actor paths in a written form that could easily be interpreted by the simulation team, which is why for some scenarios a more visual approach was adopted. Figure 5 below shows one such scenario, which, although appearing simple on paper, quickly becomes complex if all combinations of variable parameters are considered.

Experiment Number:	**LU005 - QEOP -** Partially obscured left turn with oncoming scooter
Experiment Description:	
- Pod coming towards timber lodge stop, executes gradual left turn between grassed area and square raised planter - Man and child on scooter coming in opposite direction, on a head-on collision path with pod - Scooter rider gets quite close then forced to swerve around pod to their relative right	
Grid reference: 51.547292, -0.015139	



Required Performance
Pod will stop when oncoming scooter rider gets too close, continue its path when clear

	Parameter		**Value(s)**		
Fixed Parameters	Road layout		[as diagram, see grid ref]		
	Roadside obstacles		Obscuration on approach, vision clear once turned		
	In-road obstacles		None		
	Actor 1	Type	Capri pod		
		Path	[as diagram, orange lines]		
		Manoeuvre	Gradual left turn, intending to continue straight		
	Actor 2	Path	[as diagram, blue lines]		
		Manoeuvre	Straight path, intending to turn gradual right		
Variable Parameters	Light condition		Any		
	Weather condition		Any		
	Actor 1	Speed	1mph to 5mph		
	Actor 2	Type	Scooter	Bicycle	Pedestrian
		Speed	5mph to 20mph	5mph to 20mph	1mph to 10mph
Additional Notes	This is an observed scenario from T3 at QEOP The start positions and relative timings for each actor to move can be varied, but the randomisation within the test bench should ensure to include runs where a collision or near miss would occur if the pod did not react. Specific timings to achieve this for different actor speeds can be provided if needed.				

Figure 7: Example real-world test scenario

For parameters such as actor speed, a range of speeds would be specified for each scenario. For other parameters, particularly those relating to scenery or criticalities, a taxonomy of possible values was developed, and in any given scenario either a specific value was selected if the parameter was fixed, or “any” was stated if it was variable. It was assumed that the testbench would randomise and assess each of the possible values if “any” was stated.

The tables below list the main parameters and their possible values. However, it is noted that this list is not exhaustive as it is constantly expanding and being refined as new scenarios are developed and lessons learned. As already mentioned, parameters such as road geometry and actor paths were presented visually (either via scale drawing or aerial image), though descriptive notes could be added if there was a need to convey important information (e.g. a steep slope).

Table 3: Scenery parameters

Name	**Description**	**Values**
Light condition	Broad description of the light conditions at the time of experiment	Daylight Twilight Darkness Darkness with artificial lighting
Roadside obstacles present?	e.g. buildings, parked car, street furniture, vegetation etc.	Yes / no, (only yes if relevant to experiment) If yes, specify type
In-road obstacles present?	e.g. vehicle load, debris, animals etc.	Yes / no, (only yes if relevant to experiment) If yes, specify type
Junction type	The type of the junction in terms of actors involved, type of manoeuvre and intention for each experiment.	Not at junction Roundabout Mini roundabout T or staggered Slip road Crossroads Private drive / entrance Other (specify)
Junction control	Including any constraints (e.g. traffic light phase timings fixed or variable?)	Authorised person (e.g. school crossing patrol, steward) Traffic light Stop sign Give way None

Table 4: Scenario parameters

Name	**Description**	**Values**
Number of actors	Number of actors involved in the experiment	*Integer*
Type of actor	Actor 1 will always be the AV (Pod). Subsequent actors can be any from the list.	POD Pedestrian Pedal cycle Motorcycle*[5] Car (including taxi) Minibus (8-16 seats) Bus / coach (17+ seats) Tram / light rail LGV (<7.5 tonnes) HGV (>=7.5 tonnes) Other

[5] Can be split into further categories if needed (by engine size)

Table 5: Criticality parameters

Name	Description	Values
Weather conditions	The ambient weather conditions, specifically those which may have an adverse effect on the safety performance of the AV.	Dry Rain (light / moderate / heavy) Snow (light / moderate / heavy) Fog / mist High wind (specify in mph) Other (specify)
Road surface friction	Road surface conditions which may affect stability, handling, or object detection.	Normal - dry Reduced due to rain Reduced due to snow Reduced due to contaminants (e.g. oil, gravel, leaves)
Construction / maintenance	The presence of a temporary construction or maintenance zone, not already specified in the route programming.	Temporary traffic lights 'Stop / Go' boards Traffic cones Holes in the road
Vehicle defect	A pre-existing or sudden defect in the AV which may affect safety performance. E.g. sudden sensor failure.	Yes – mechanical (specify) Yes – electronic (specify) Yes – sensor related (specify) No
Abnormal behaviour of non-AV driver	e.g. interacting actor driving under the influence of alcohol or drugs, speeding, driving wrong way down the street.	Yes (specify) No

5. Physical Testing Protocols

The Capri project set out to showcase the POD technology and service across a series of public trials taking place at the Queen Elizabeth Olympic Park (QEOP) in London and the Mall at Cribbs Causeway in Bristol. In advance of each public trial the technology had to be tested in order to assess and build confidence in the safety performance of the POD before it operates in a public environment.

In an ideal world, a vast amount of testing would be carried out in simulation, with a small selection repeated in the physical world to verify the accuracy of the simulation environment. However, due to time constraints in the Capri project, the trials occurred before much simulation was possible, and as such a larger amount of physical safety testing was carried out.

The scenarios and methodologies developed through the process described in the previous section were adapted to be used in physical testing. Detailed test plans were created to assess the possible hazards and scenarios that had been identified through the site surveys of the trial sites. These test plans were primarily focussed on the off-road operation, in pedestrianised areas, as this would form the majority of the public trials. However, protocols for assessing on-road performance were also developed, though this report does not cover these in great detail.

The plans described in this section were developed to cover as wide a range of scenarios and hazards as possible, whilst accepting that it is impossible to predict every single scenario which may occur in the real world. These protocols were then used to carry out over 100 hours of testing with the POD, assessing a number of different software versions for different use cases. In total, over 1000 individual safety tests were carried out, and, though it was not possible to complete every possible test that had been defined, a clear understanding of the safety performance of the POD was gained.

It is noted here that the testing protocols for the physical testing included in this report are more detailed than those for the virtual testing. This is because the physical testing was carried out by Loughborough, so the protocols cover the implementation as well as the scenario design. For the virtual world, implementation was carried out by other partners in WP4, so only the scenarios and test parameters are described. Further information on how the simulations are carried out can be found in reports from the other partners.

5.1 Test setup

5.1.1 POD manoeuvres

The test plans aimed to assess detection and response to hazards while the POD is operating in two main manoeuvres: straight line and turning.

Turning tests involve a 90-degree turn (or maximum possible turn radius the POD is able to complete) to either the left or the right, with the respective test plans for these manoeuvres being effectively a mirror image of each other. Assuming the sensor arrays are identically mirrored for each side of the POD then it would be expected that results for the left turn would be the same as those for the right turn, however this cannot be assumed. If there is limited testing time available then it is preferable to carry out all tests for one turn rather than half of the available tests on each turn, however it is recommended that all tests be completed for both the left and right turn manoeuvres for due diligence.

Within each manoeuvre there are three broad categories of tests:

1) POD approaches static hazard. The POD should stop if the hazard is within or close to the side of its path.

2) POD is static and a hazard is placed within or near its path. The POD should not move off if separation distances are too close, only moving off if there is enough space to do so and stopping again if the hazard comes into closer proximity.

3) Moving hazard intersects the POD path. The POD should stop to avoid any contact with the hazard, and only continue its route when it is clear to do so.

Broadly, all tests fit into four general groups shown below in Figure 8. These simplified groups cover all types of scenario that the POD could be expected to avoid using its sensing and braking capabilities. The huge variety of situations presented to a POD in real-world use will in all likelihood result in scenarios which exist outside of these groups, however the intention is for other scenarios and manoeuvres to be tested in simulation where there are fewer time constraints.

Figure 8: General test type groups

All hazards identified in the following section are tested under static and/or moving conditions for both straight running and turning tests.

5.1.2 Test grids

In order to accurately position hazards relative to the POD path, and to carry out tests in a systematic and repeatable way, testing grids were designed for each POD manoeuvre.

Straight line tests:

A square grid is used as shown in Figure 9. The centre vertical line of this grid is the centreline of the path followed by the POD; this should be determined by following the POD through its route and marking its centreline.

The grid then expands 2.8m to either side; first with two increments of 400mm then two increments of 1000m. The purpose of the initial smaller increments is to create a ‘corridor’ in the centre of the grid which is roughly the same width of the POD, to allow hazards to be accurately placed immediately adjacent to the POD path.

Finally, perpendicular horizontal lines are added at 1000mm increments, completing the grid. The plans nominally specify 5 increments for a total of 5000mm (5m), in order to give a large testing area within which to assess responses to hazards. However, this number can be increased or decreased depending on the required tests, and similarly the vertical lines can continue further out if needed.

A cartesian co-ordinate system is used to describe the specific hazard placements, with (0,0) being the centre of the grid at the lowest point of the vertical lines; i.e. the point at which the centre of the pod enters the testing grid. Figure 9 is annotated with a selection of co-ordinate points for reference; it should be noted that the first two increments of 400mm represent +/- 0.5 and +/- 1 on the x axis.

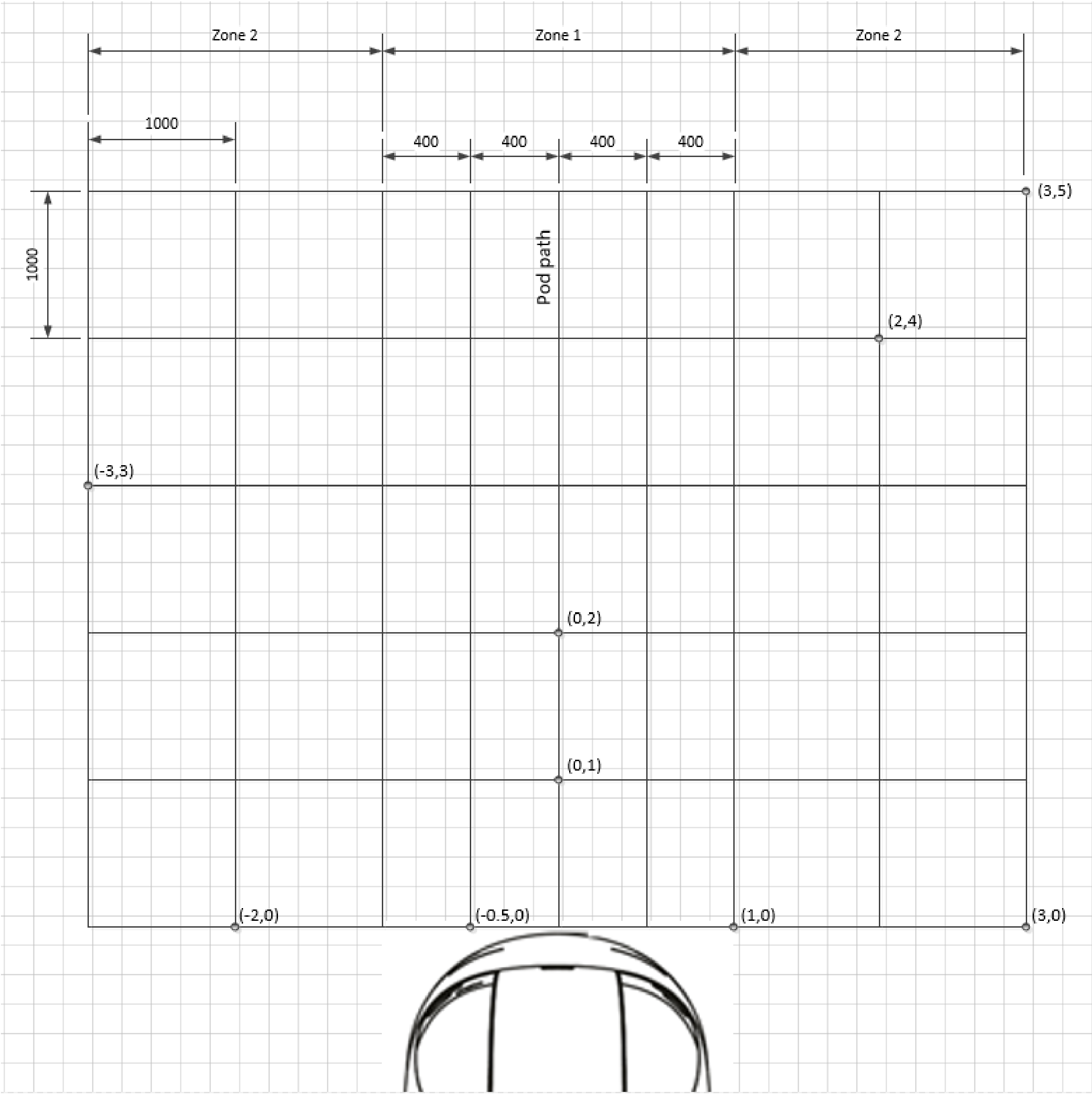


Figure 9: Straight line test grid with example co-ordinates

Turning tests:

For turning tests, a similar process is used, though it is slightly more complex due to the dynamics of a turn manoeuvre. Figure 10 below shows an example left turn grid, it is started in the same way by following the POD and marking the centreline as it turns – this becomes the '4' line of the grid.

Next the 'origin' of the circle radius is determined (an approximate calculation by eye), and string is used to mark straight-line spokes around the turn, intersecting the centreline at 1000mm intervals.

Finally, for each spoke intervals of 800mm are marked either side of the curved line, with these points being able to be joined to form the additional curved lines which will mark the POD 'corridor' (3 and 5 lines) and then intervals of 1000mm are marked along each line going further in and out as needed.

For the turning grids a labelling system more similar to map references is used, with numbers for the curved lines and letters for the intersecting spokes. This allows specific description of hazard placement, for example A4, C5, D6 etc.

Figure 10 shows a 'quarter circle' turn with spokes A to H, and curve lines 1 to 7, however as before this can be increased or decreased as needed. This can be mirrored to produce the right turn grid.

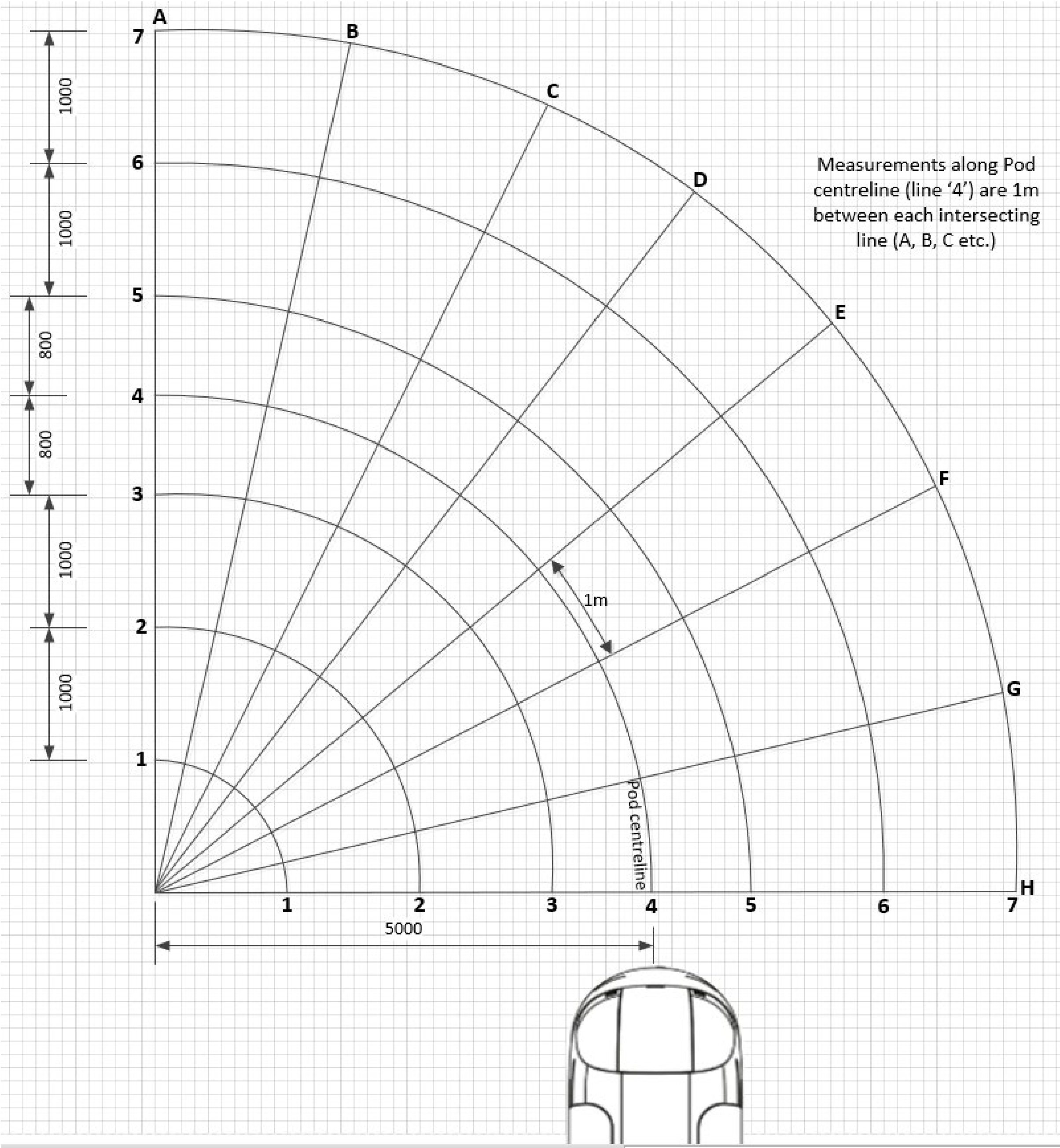


Figure 10: Left turn test grid with example co-ordinates

For turning tests, the turn radius was set to be a balance between what was needed to navigate the routes planned for the public trials and the capability of the pod. For all tests carried out during the pre-trial testing the turn radius was 7m.

The tests plans describe a variety of static positions and moving paths in relation to these grids and the list of possible hazards, all of which were derived from the scenario development as described in the previous sections.

Test scenarios for static hazards are relatively simple due to the lack of dynamic movement. However, they are extremely valuable, as specific systematic placing of the hazards around the testing grid allows a testing team to determine the exact bounds of capability and highlight any 'blind spot' areas in a repeatable and clearly identifiable way.

Scenarios for dynamic hazards, whilst following prescribed paths, are more 'naturalistic' and aim to replicate real world scenarios. These are generally more challenging and require more detailed setup.

5.2 Hazards

Static Hazards

For scenarios involving static hazards, a range of possible targets were included in the test plans:

- Adult pedestrian (dummy or live person)
- Child pedestrian (dummy) – 7-year-old, 5-year-old, toddler
- Simple pushchair
- Trolley
- Trunki suitcase
- Scooter
- Bicycle
- Car

This list was drawn primarily from site observations; adults, children, pushchairs and trolleys were initially specified for pre-trial testing for Cribbs Causeway, and the list was expanded to include scooters, bicycles and dogs after observations at QEOP. The Trunki suitcase as included to represent a 'low level' hazard; the test plan also specifies a lying down pedestrian dummy to assess low level detection, however the Trunki offers an option that is taller than a lying down pedestrian but still relatively low, which can be used if there are poor responses to the lying down pedestrian to determine the exact levels of low level detection capability.

When devising test scenarios, one or more of these hazards, or multiples of a hazard, can be used in a variety of different combinations of positions / orientations.

Dynamic (moving) hazards

Dynamic hazards were again chosen based on observations at the trial sites. The majority of scenarios for dynamic hazard testing involve moving human pedestrian hazards. The specifications include a range of variations on this, specifically:

- Adult pedestrian walking
- Adult pedestrian jogging
- Multiple adult pedestrians walking (opposite directions, crossing path)
- Multiple adult pedestrians walking (same direction, convoy, varying separation distances)
- Adult pedestrian(s) pushing pushchair
- Adult pedestrian(s) pushing trolley
- Adult pedestrian(s) riding scooter

The final dynamic testing hazard is a moving cyclist, for which a variety of scenarios were devised.

5.2.1 Hazard positioning

The exact positioning of the hazard for each test is reported in terms of a general 'position relative to POD' as well as the previously described more specific grid co-ordinate position. The relative position generally falls into one of the following categories:

1. 'Front' – directly along the centreline of the POD path. I.e. '0' on the x-axis for straight line tests and along the '4' line for turning tests.

2. 'Front left' / 'Front right' – offset from the centreline to either the left or the right by 0.8m, to achieve the effect of being directly to the side of the POD or just caught by the corners as it passes. I.e. -1 or 1 on the x-axis for straight line tests and along the '3' or '5' lines for turning tests.

3. 'Left / right' – outside the 'corridor' travelled by the POD, to achieve the effect of potentially being close to the POD but would not be impacted as it passes. I.e. -2 or 2 on the x-axis for straight line tests and along the '2' or '6' lines for turning tests.

4. 'Far left' / 'far right' – well outside the 'corridor' travelled by the POD, would be a comfortable passing distance if POD passed by. I.e. -3 or 3 (or greater) on the x-axis for straight line tests and along the '1' or '7' lines (or further out) for turning tests.

For humanoid, pushchair and bicycle targets, a large number of tests were carried out and as such a 'standard orientation' for static tests was defined Orientations for less commonly used targets were described in the test results on a case by case basis. Similarly, if a test uses a non-standard orientation that is different to that which is described below this should be noted in the test results.

Humanoid Targets

It was decided to orient the child dummy and adult human targets in a 'side-on' position (shown in Figure 11 below), as this presented a more challenging target for the sensors and is also more realistic of what would be expected in a public trial with pedestrians crossing the path of the POD.

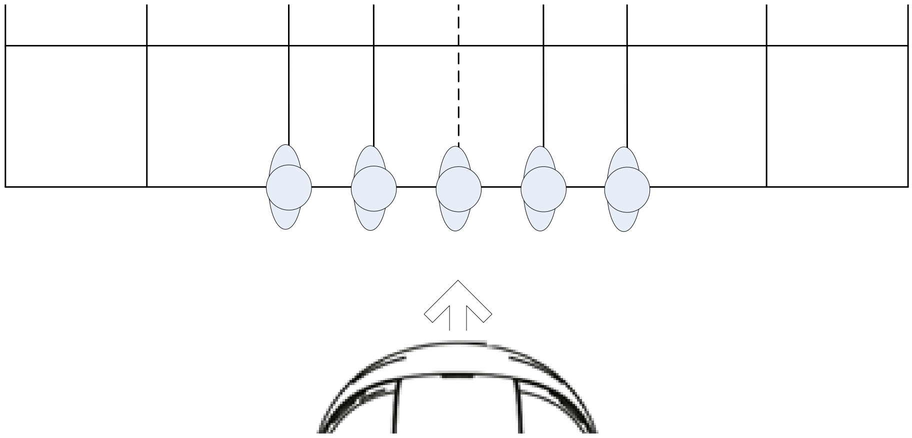


Figure 11: Orientation of humanoid targets for straight line running tests

Pushchair

For the straight running tests the pushchair was placed on the front line of the testing grid at a 45° angle to the approaching POD path. The decision to place the pushchair at this angle was to cover the widest range of potential paths a pushchair might reasonably take across the path of the POD with minimal test iterations. For the straight-line running tests with a static pushchair the orientation of the pushchair remained the same with the front of the pushchair always pointing towards the 'offside' of the POD (as if crossing from the relative left to right of the POD path).

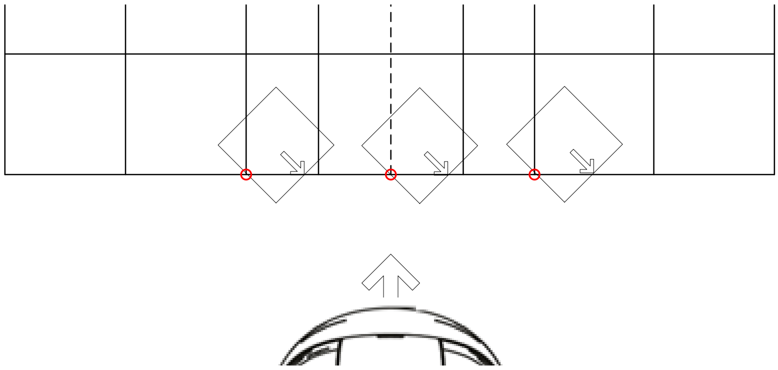


Figure 12: Orientation of pushchair target for straight line running tests

In order to place the pushchair in a repeatable location on the test grid an easily identifiable point is used. Using the side view of the main body of the pushchair (see Figure 13) and creating a grid over this a rivet on the central hinge was identified as a suitable reference point. This point does not represent the centre of mass or centre of the object but purely the centre of the side plane of the object which can be used to align the pushchair over any given point in a consistent manner.

For turning tests, the same methodology for consistent placement and orientation was used, with the pushchair placed as if crossing from the relative left to right of the POD path.

Figure 13: Pushchair design and centre-point used for positioning

Bicycle

The static bicycle target was placed in either a ‘head on’ position, as if riding directly towards the POD, a ‘side on’ position, as if taking a perpendicular path to the POD, or at a 45-degree angle, in the same way as the pushchair was oriented. For consistent positioning accuracy, the pedal hinge was used as a datum point as this was roughly in the centre of the bicycle

Figure 14: Bicycle design and datum point for positioning

The bicycle was held upright by a tubular metal bike stand on each wheel. Figure 15 shows the ‘as tested’ condition of the bicycle positioned at 45 degrees on a test grid.

Figure 15: Bicycle positioned at 45 degrees to testing grid and in 'as tested' condition

5.2.2 Hazard timings for dynamic testing

Other than the different hazards over the four groups of test and their placement around the grids, the only other major variation in testing relates to timing, specifically the timings needed to get the POD and moving hazard into a collision or near miss type scenario at the intersection of their paths. This methodology ensures that all tests are repeatable, reliable and comparable and, in the case of cyclist and pedestrian hazards, safe.

In order to release a hazard, timing marks calculated from the POD speed and path length were used. These timing marks indicated when a hazard should begin moving towards the point where its path intersected with that of the POD (the 'collision point') – a member of the testing team stands by this mark and shout 'go' when the POD reaches it. These timing marks also allowed for different separations between POD and hazard, for instance it was possible to conduct tests where the hazard would pass through the collision point one, two or three seconds before the POD reached the same point. A schematic of example timing marks and calculation is shown in Figure 16.

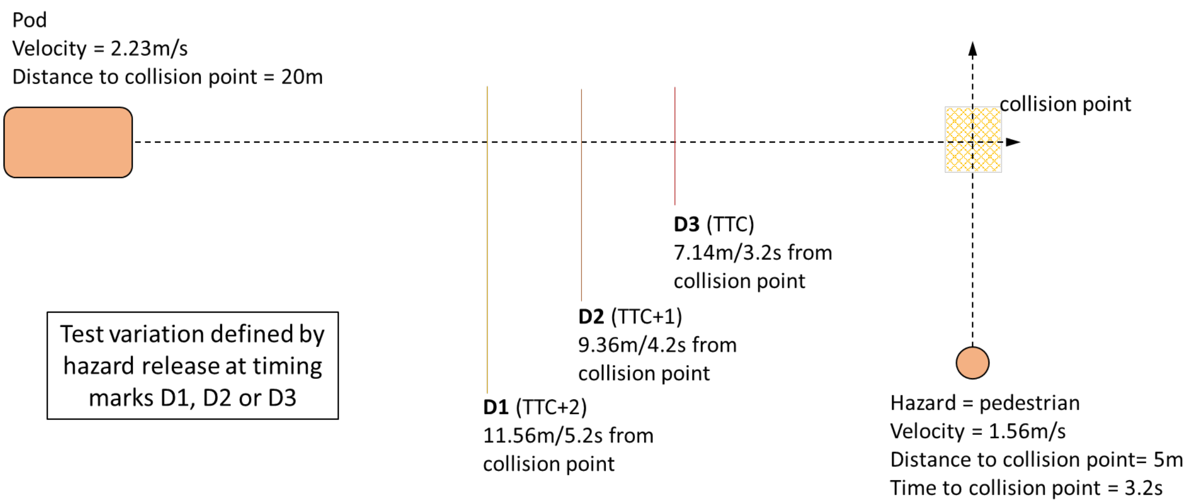


Figure 16: Schematic showing timing marks to determine hazard release

This variation allows each block of testing for a particular hazard to be split into finer subsets of tests giving a clearer understanding of how the POD sensing and decision-making systems operate at different vehicle separations (i.e. for accidents and for near misses). An outline of how the timing categories were derived in this example and the terminology used is shown below:

- Timing mark ‘D1’ was derived for the hazard to pass through the POD path and collision point 2 seconds before the POD reached this point. This can also be seen as ‘Time to Collision’ plus 2 seconds (TTC+2).
- Timing mark ‘D2’ was derived for the hazard to pass through the POD path and collision point 1 second before the POD reached this point. This can also be seen as ‘Time to Collision’ plus 1 second (TTC+1).
- Timing mark ‘D3’ relates to a point where, should no braking occur, the POD and hazard will meet at the intersection of the path of the POD and the path of the hazard (point of collision). This can also be seen as ‘Time to Collision’ (TTC).

In the example 3 timing marks are calculated for a pedestrian hazard travelling at a 1.56m/s (approx. 3.5mph). This process can be repeated for a range of hazard speeds, and with any TTC separations.

5.2.3 Hazard construction

Some elements of the test plan require specific construction of hazards and/or a simulated environment. This enables the testing to be repeatable and more representative of the real-life scenarios the POD will encounter. Construction included:

Pedestrian dummies
A 50th percentile adult male, 50th percentile 7-year-old male child and 50th percentile 5-year-old male child were constructed using standard measurements from EuroNCAP testing protocols[6] or verified anthropometry measures such as the ‘Childata’[7] reference book.

The purpose of this was to allow ‘edge case’ testing or testing of potentially more risky scenarios in a safe manner. For example, in the early stages of testing, dummies were used first to gain confidence and demonstrate suitable POD performance before live human targets were used. Child dummies were constructed purposefully as children represent a harder target for the POD to detect but testing with real children is not practical.

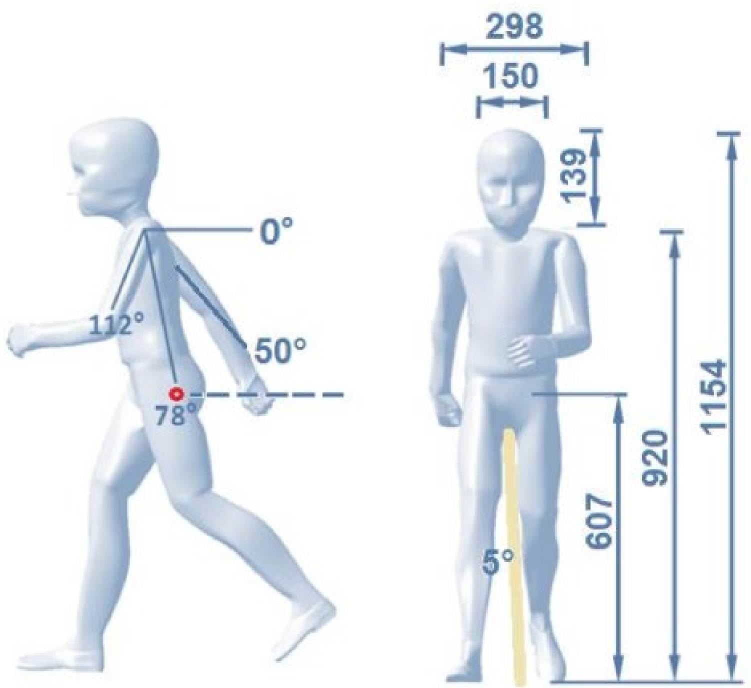


Figure 17: 7yo child pedestrian dimensions for AEB testing (source: EuroNCAP)

[6] https://www.euroncap.com/en/for-engineers/protocols/vulnerable-road-user-vru-protection/ and https://www.acea.be/uploads/publications/Articulated_Pedestrian_Target_Specifications_version_1.0.pdf
[7] Childata. The handbook of Child Measurements and Capabilities – Data for Design Safety. DTi. In addition, British Standards Institution (BSI) data has been used where available.

The dummy build material was carefully chosen as it had to be durable in case the dummy was struck by the POD but light and flexible enough that it would not damage the pod in these instances. The material also had to be repairable in the field and widely available/cheap enough that spares could be easily bought and constructed.

One construction material fitted within these constraints and subsequently the bulk of the material used for the pedestrian dummies is flexible tubular polyethylene foam (PEF). This material is typically used as pipe lagging for buildings and will conform with British Standard BS EN 14313:2015.

As shown in Figure 18 the arms, legs and torso of the dummies are constructed from the PEF tubing with any ‘hard points’, for example the shoulder and hip joints, made out of plastic pipe. Articulation of the arms at the elbow and the leg at the knee are achieved by foam covered wire ties imbedded within the PEF tubing.

Figure 18: Pedestrian dummy construction

Before the dummy is dressed a final layer of bubble wrap or foam rubber sheet is applied over all surfaces to achieve the correct dimensions specified in the sections above, this is held in place with paper masking tape. Hands and feet are made of closed cell foam, shaped to the correct dimensions before being wrapped in tape.

Hardwearing clothing in a dark colour, typically blue or black, is used to cover the dummy. The clothing consists of a pair of trousers (typically a thick cotton) and a sweatshirt/hoody (cotton or heavy-weight jersey material).

Built environment

A built environment, consisting for instance of walls and bollards. was used to replicate challenging areas of the route so they could be assessed prior to the pod going on site. For example, as previously discussed, at Cribbs Causeway the route involved the POD passing a lift shaft that would obscure pedestrians approaching, so testing involved building a wall and having pedestrians move out from behind it into the pod path. During a site survey measurements were taken of this area so it could be accurately recreated in testing.

The major ‘wall’ sections of the built features were designed to be modular so features could be built at a variety of heights or lengths or to form a corner or a building or wall. Sections were constructed from a frame of timber faced with 12.5mm plasterboard. This facing material was chosen as it presents the densest material to the sensors and should therefore give a reliable POD response and close comparison between the built features and a real wall/building. Plasterboard is also easily worked and simple to construct.

In addition to the flat features such as walls Cribbs Causeway has a large number of steel bollards protecting doorways and other areas from vehicles. To simulate these bollards sections of stainless-steel ventilation ducting were cut to the correct size. The reflectivity of the ducting closely matches the stainless-steel finish of the ones at cribs causeway even if the density of the material is considerably lower (0.6mm for the ducting compared to ~4mm for the bollards).

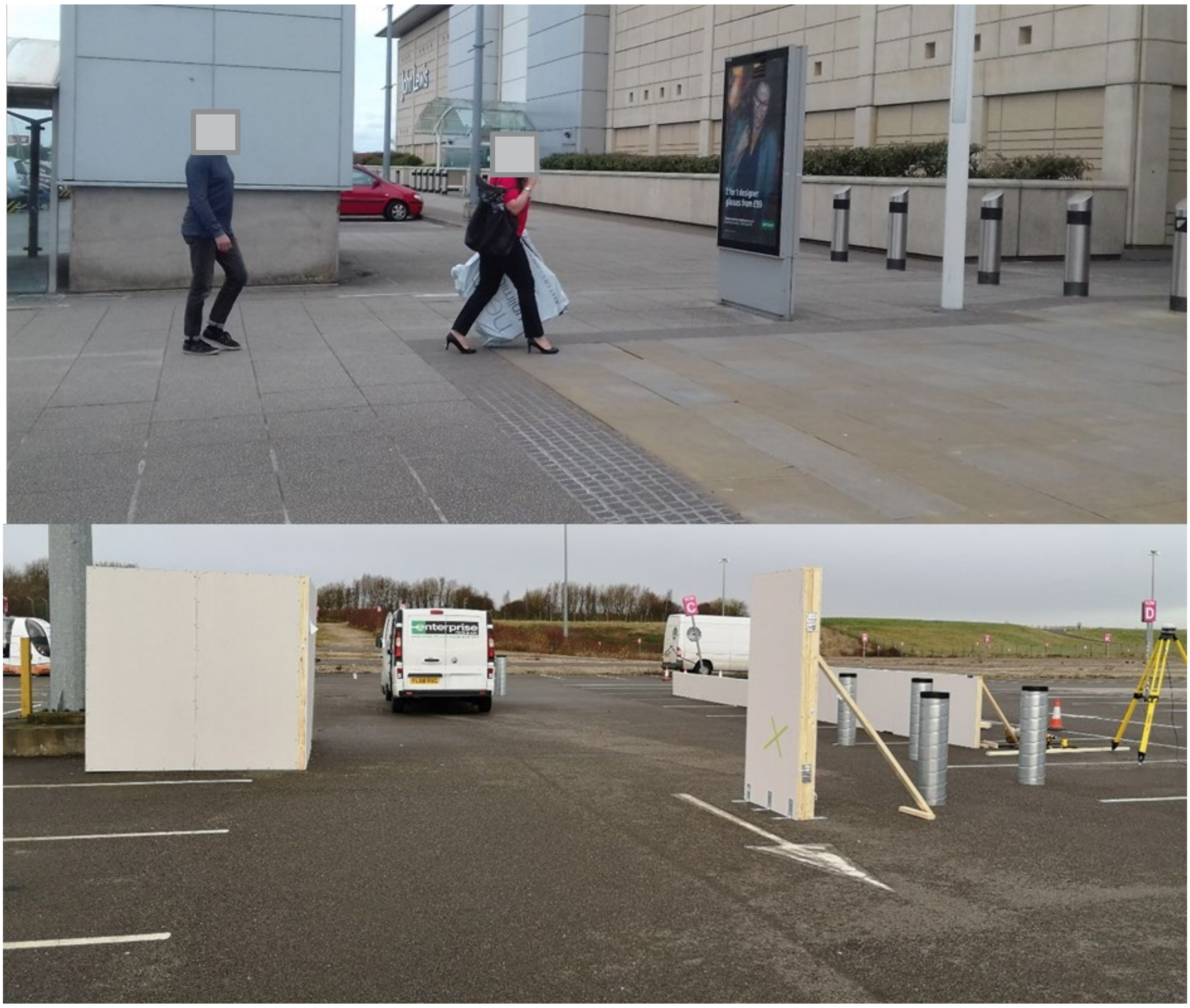


Figure 19: Comparison of Cribbs and built environment during pre-trial testing.

5.3 Test scenarios

5.3.1 Static hazards

Test scenarios for static hazards are relatively simple due to the lack of dynamic movement. However, they are extremely valuable, as specific systematic placing of the hazards around the testing grid allows a testing team to determine the exact bounds of capability and highlight any ‘blind spot’ areas in a repeatable way.

POD straight line manoeuvre

For straight line manoeuvres, the recommended minimum protocol is to take each hazard in turn and carry out tests with the hazard placed in the positions listed below in order to assess detection and response to hazards for all areas both within and alongside the POD path. The expected responses for each test are shown below.

Table 6: Tests for static hazards (pod running straight)

Hazard position	Expected pod response
(0,0)	Stop for hazard
(0.5,0)	Stop for hazard
(1,0)	Stop for hazard *Note: depending on the hazard used, pod not stopping with this test may not result in contact with hazard but would be very close passing*
(2,0)	Slow for hazard, move past if safe to do so *Note: for larger hazards, e.g. side on bicycle, this positioning will leave part of the hazard closer to the pod and it should be ensured that pod response is safe and*

	comfortable for the hazard
(3,0)	Continue past, possibly slowing if operating at higher speed *Note: stopping for hazards in this position is not a safety concern, but could present operational issues*
(-0.5,0)	Stop for hazard
(-1,0)	Stop for hazard *Note: depending on the hazard used, pod not stopping with this test may not result in contact with hazard but would be very close passing*
(-2,0)	Slow for hazard, move past if safe to do so *Note: for larger hazards, e.g. side on bicycle, this positioning will leave part of the hazard closer to the pod and it should be ensured that pod response is safe and comfortable for the hazard*
(-3,0)	Continue past, possibly slowing if operating at higher speed *Note: stopping for hazards in this position is not a safety concern, but could present operational issues*

Additional testing beyond (-3,0) and (3,0) may be useful if wider braking envelopes are being used, in order to determine what size corridor needs to be clear for the POD to progress through. Additionally, closer intervals (e.g. (2.5,0)) can also be used if the 'breakpoint' is between points and it is useful to know the exact point when the POD will / won't stop for a hazard. For complex hazards, such as the pushchair, hazard positions can be repeated with a different orientation if this provides further insight.

For each of the hazard positions listed there is also a set of '**moving off tests**' that can be carried out with each hazard. As a minimum 'moving off' tests should be carried out with at least one hazard for the '0', '1' and '-1' lines of the y axis.

The process for 'moving off' tests is to 'creep' the POD to the start of the test grid, with the centre point at (0,0), and place it in 'pause' mode. A hazard is then placed 0.5m in front of the POD along the line of the y axis to be tested. The POD is then placed in ACS mode to see if it moves off or not. This process is repeated with the hazard moving further out in 0.5m increments until the POD moves off (where it is expected it will then stop for the hazard again as it gets closer to it).

An example set of moving off tests is given below, which can be repeated for any hazard and horizontal placement. It should be noted that the separation distance required by the POD between itself and the hazard will be determined by the braking envelopes, and the size of this separation may be larger or smaller depending on operational requirements. For safety purposes, it is recommended that the POD should be able to keep a minimum 1m clearance between itself and a hazard.

Table 7: Tests for static hazards (pod moving off)

Hazard position	**Expected pod response**
(x,0.5)	Will not move off from stationary
(x,1)	Will not move off from stationary
(x,1.5)	Will not move off from stationary
(x,2)	Moves off from stationary if wide enough separation, stops again for hazard within a safe and comfortable distance *Note: for larger hazards, e.g. head on bicycle, this positioning will leave part of the hazard closer to the pod and it should be ensured that pod response is safe and comfortable for the hazard*
(x,2.5)	Moves off from stationary if wide enough separation, stops again for hazard within a safe and comfortable distance *Note: for larger hazards, e.g. head on bicycle, this positioning will leave part of the hazard closer to the pod and it should be ensured that pod response is safe and comfortable for the hazard*
Etc…	

POD turn manoeuvre

For turn manoeuvres, the recommended protocol is to take each hazard in turn and carry out tests with the hazard placed in the positions listed below in order to assess detection and response to hazards for all areas both within and alongside the POD path. The expected responses for each test are shown in Table 5.

Note that with the turn manoeuvre there is a much larger number of tests to carry out, due to assessing each lateral position on each spoke of the grid in order to assess responses at each stage of the turn. If testing time is limited, then as a minimum tests should be carried out for every other spoke (A,C,E etc.) for the '3' '4' and '5' lines. Should any issues occur then the uncompleted tests should be carried out to fill in the gaps and determine the exact break point of capability. Also note that the 'inside' of the turn is generally more challenging and tests on the inside should be prioritised.

Additionally, it is recommended to start testing with the hazard placed on the spoke at the end point of the turn, moving incrementally inwards. This is because a point will be reached where the POD stops before its wheels have started to turn, and this can in effect be considered a 'straight line' test scenario and there is then no need to continue testing at any remaining spokes.

'Moving off' tests are not minimum protocol for turn tests as long as satisfactory responses have been observed for straight-line tests. However, if desired they can be carried out with a similar process.

As a final note, it should be borne in mind that due to the mechanics of the turn manoeuvre, the rear wheels do not follow the exact same path as the front wheels. As such, a situation can occur where the front of the POD will pass clear of an object, but the rear will impact if it does not respond. Hazard placement can be adjusted, or smaller increments of intervals used, to account for this and ensure the desired test is achieved.

Table 8: Tests for static hazards (pod turning)

Hazard position	Expected pod response
A4	Stop for hazard
B4	Stop for hazard
C4	Stop for hazard
D4	Stop for hazard
E4	Stop for hazard
F4	Stop for hazard
G4	Stop for hazard
H4	Stop for hazard
A3 through H3	Stop for hazard *Note: depending on the hazard used, pod not stopping with this test may not result in contact with hazard but would be very close passing*
A2 through H2	Slow for hazard, move past if safe to do so *Note: for larger hazards, e.g. side on bicycle, this positioning will leave part of the hazard closer to the pod and it should be ensured that pod response is safe and comfortable for the hazard*
A1 through H1	Continue past, possibly slowing if operating at higher speed *Note: stopping for hazards in this position is not a safety concern, but could present operational issues*
A5 through H5	Stop for hazard *Note: depending on the hazard used, pod not stopping with this test may not result in contact with hazard but would be very close passing*
A6 through H6	Slow for hazard, move past if safe to do so *Note: for larger hazards, e.g. side on bicycle, this positioning will leave part of the hazard closer to the pod and it should be ensured that pod response is safe and comfortable for the hazard*
A7 through H7	Continue past, possibly slowing if operating at higher speed *Note: stopping for hazards in this position is not a safety concern, but could present operational issues*

5.3.2 Adult(s) walking

Test parameters

General human walking speed, including people of all ages, ranges from around 1mph to 6mph. The average adult walking speed is around 3mph. For testing with walking adults, whilst a range of speeds would be useful, it is most important to get consistent timings and as such it is recommended to use the same person(s) for all tests within a block, and for them to walk at a speed that feels natural and comfortable to them as that will be more replicable.

To carry out dynamic testing with adult(s) walking, timing marks are calculated as described in section 5.2.2. For each test scenario, it is recommended to start with TTC+2 (for safety reasons to assess response in a non-collision scenario), then move incrementally forward through TTC+1, TTC, and then to TTC-1, TTC-2 etc.

For these later separations (TTC-x) the purpose is to test a person walking out right in front of the POD at the last moment or walking effectively into the side of it. In these cases it is anticipated that the POD may be very close to the person when it stops, but the purpose it to test if a response is given, not that it maintains a clear 1m distance for example.

As a minimum testing protocol, TTC+1, TTC and TTC-1 should be tested. Additional separations, including ‘half second’ (e.g. TTC+1.5), can be tested where they will provide useful further insight.

POD straight line manoeuvre

For the POD operating in a straight-line manoeuvre, the walking path taken by the walking adult(s) is perpendicular to the pod path.

Whilst this can be anywhere on the grid, it is recommended to be along a higher line as this will result in the POD stopping within the testing grid area which is easier for measuring and recording. Figure 20 shows an example marked path, walking from (3,5) to (-3,5), or the reverse of this.

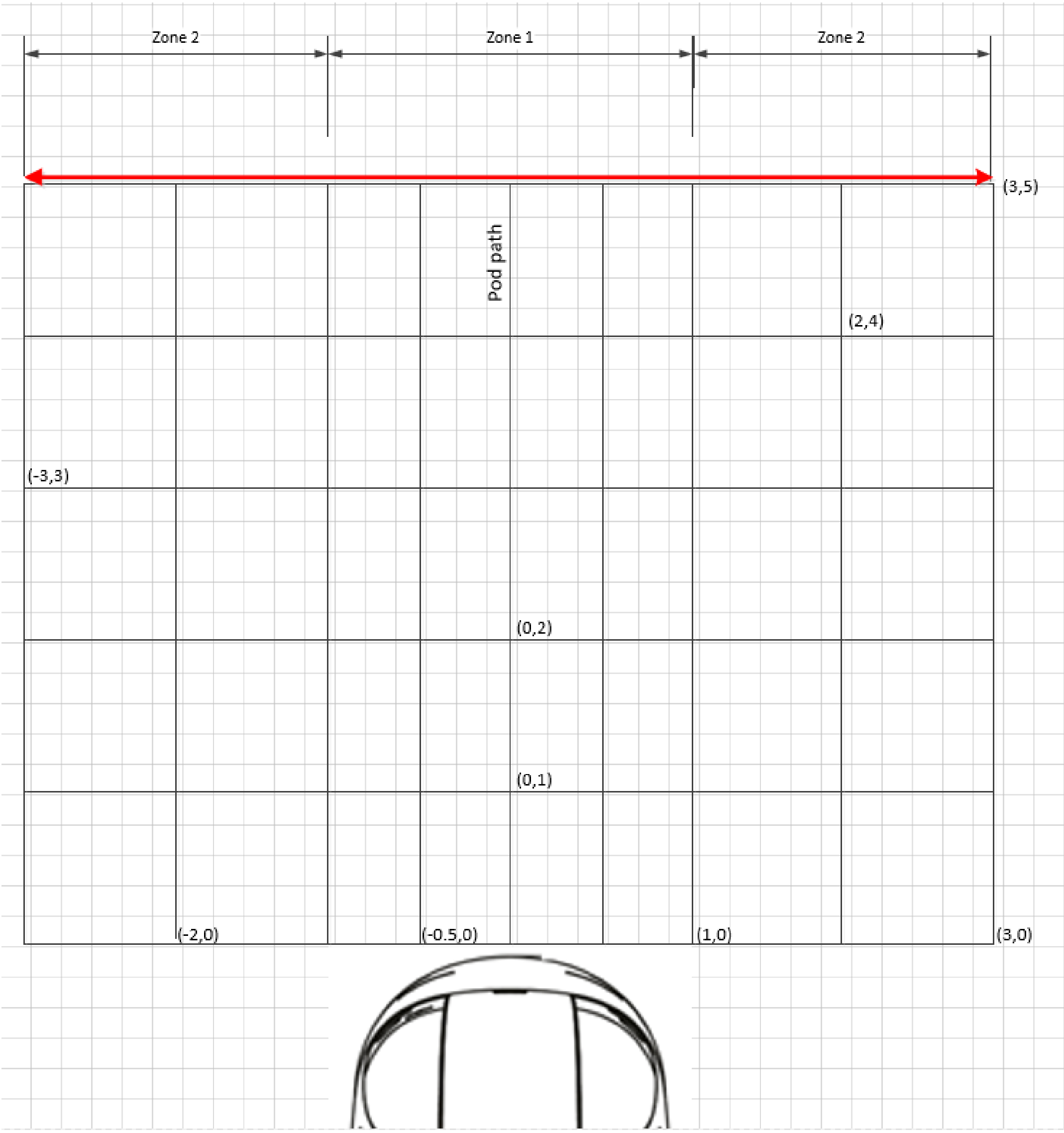


Figure 20: Straight line test grid with marked walking path

This scenario is an adaptation of established EuroNCAP testing protocols; specifically, CPNA-25 and CPNA-75, which are designed to test AEB in cars with an adult walking from the nearside in a perpendicular path. EuroNCAP protocols are described in detail on their website[8] and will not be detailed further here but can be referenced for further information and justification of this scenario.

The scenario shown in Figure 20 can be tested with any of the described dynamic hazards. As a minimum it is recommended to test a single walking adult and 2 adults in convoy, varying both the timing marks to the first hazard, and the separation distances between hazards in convoy. Other walking paths could also be tested using the same methodology.

POD turn manoeuvre

For the POD operating in a turn manoeuvre, two walking paths taken by the walking adult(s) have been developed based on site observations at Cribbs Causeway, shown in Figure 21.

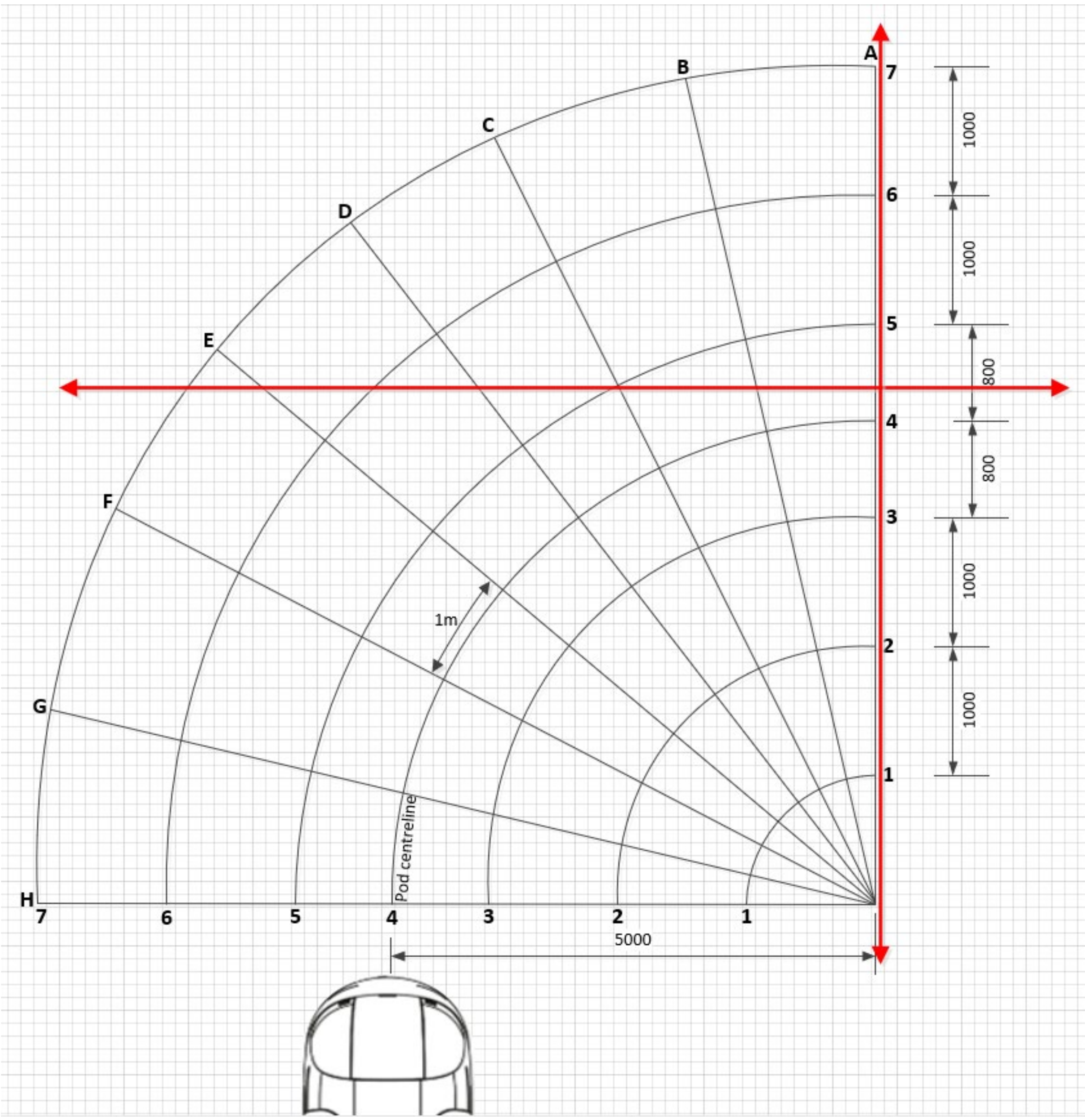


Figure 21: Turn test grid with marked walking paths

For these tests, timing marks can be calculated using a slightly more ad-hoc process. As the POD turns it will slow and the speed will vary, making it more challenging to calculate exact timing marks. Instead a systematic method is used by utilising the grid spokes, with the testing team shouting 'go' to the hazard as the pod reaches the A spoke, B spoke, C spoke etc. Additional timing marks can be added prior to the POD reaching the grid by measuring 1m intervals back along its path.

[8] https://cdn.euroncap.com/media/26997/euro-ncap-aeb-vru-test-protocol-v20.pdf

For cycling tests, where the hazard will need to start moving when the pod is much further back, the usual TTC method should be used, though some variance will be expected due to the changing POD speed and this should try to be factored into calculations.

As with the straight-line tests, the scenario shown in Figure 21 can be tested with any of the described dynamic hazards. As a minimum it is recommended to test the ‘A7-A1’ path (the more challenging scenario) with a single walking adult and 2 adults in convoy, varying both the timing marks to the first hazard, and the separation distances between hazards in convoy.

Other walking paths can be drawn on the grid to be tested, and both paths shown here can be tested in either direction, though caution should be used when walking from A1 to A7 as this will place the hazard in the blind spot of the POD until it turns to face them.

5.3.3 Adult cycling

Bicycles were observed throughout QEOP travelling at a range of speeds and trajectories, both as single hazards and travelling in groups, and ranging in ability from young children to professional riders. Since these were determined so be a significant and challenging hazard at QEOP, a large number of test scenarios and parameters have been developed.

Test parameters

Based upon site observations and feedback from QEOP, it is anticipated that the pod could encounter cyclists travelling around the Olympic park at speeds within the range of 5mph to 20mph, with the average speed being nearer the lower end of this range.

Testing of all possible cyclist speeds would not be practical in the real world, however it is recommended to test the scenarios with a ‘slower’ and ‘quicker’ moving cyclist as a minimum. Other speeds can then be tested in simulation.

For safety reasons, it is not recommended to carry out the cycling scenarios with “TTC” separation distances, as it is harder for a person on a bicycle to manoeuvre at the last moment if the POD does not react. It is also harder for a cyclist to maintain a constant speed for accurate timing mark calculations, so TTC+2 or TTC+3 separations are recommended to allow for slight inaccuracies.

Table 7 below gives suggested approach angles for the scenarios described below. These angles are based on analysis of path intersections at QEOP.

Table 9: Cycling scenario angles

Test	Minimum approach angle	Maximum approach angle
A1 & A2	10 degrees	25 degrees
B1 & B2	20 degrees	160 degrees
D1 & D2	20 degrees	40 degrees
D3 & D4	10 degrees	25 degrees
E1 & E2	n/a	n/a

Test scenarios

These scenarios have been designed primarily with cyclists in mind, however there is no reason they could not be used /adapted to test with other faster moving targets such as an e-scooter rider. The descriptions below should be read alongside the diagrammatical representations in Appendix A.

Scenario A (A1&A2)

Designed to replicate another road user passing the pod and turning on/off another path as the pod continues straight ahead. This test will be designed to replicate the common pathway angles seen at QEOP and will also build on system coverage.

A1) road user coming from behind the POD, passing the nearside of the POD and turning right across the POD path.

A2) road user coming from behind the POD, passing the offside of the POD and turning left across the POD path.

Note: Angle of road user turn and time separations are independent variables which are determined for each test based on the requirements.

Scenario B (B1 & B2)

Designed to replicate another road user crossing in front of the POD. Building on previous testing, perpendicular crossing[9] with a variety of hazards is planned, along with road users passing in front of the POD on an angled path as the POD continues straight ahead.

B1) road user crossing ahead of the POD from nearside to offside.
B2) road user crossing ahead of the POD from offside to nearside.

Scenario C (C1 & C2)

Designed to replicate the random naturalistic behaviour of pedestrians and other road users but in a repeatable and measurable way.

C1) Three areas approximately 4m x 4m on either side of the POD path (randomly applied) and two linear paths between these can be used to randomly apply hazards (both static and dynamic).
C2) Three areas approximately 4m x 4m on either side of the POD path can be used to randomly apply hazards. Hazards can move between areas.

Scenario D (D1 to D4)

Designed to replicate the behaviour of a road user passing on the inside of the POD as it turns.

D1) Road user coming from behind the POD on an angled path crosses in front of POD as it completes a right turn.
D2) Road user coming from behind the POD on an angled path crosses in front of POD as it completes a left turn.
D3) Road user coming from in front of the POD on an angled path crosses in front of POD as it completes a right turn.
D4) Road user coming from in front of the POD on an angled path crosses in front of POD as it completes a left turn.

Scenario E (E1 & E2)

Designed to replicate a road user travelling on a straight path passing the POD on the inside of the pod as it turns. This differs from Scenario D in that the location of the road user path will be within the PODs side detection.

E1) Road user coming from behind the POD on a straight path crosses in front of POD as it completes a right turn.
E2) Road user coming from behind the POD on a straight path crosses in front of POD as it completes a left turn.

5.4 Building a complete test plan

The complete test plan of all possible tests is generated by cross tabulating every possible hazard, position, walking path or scenario, and timing mark, for each possible POD manoeuvre. With just the parameters described in this section, this would lead to over a thousand possible tests and as such will not be shown in full here. The general methodology is shown below however, and this can be applied to any or all of the parameters described in this document or new ones specific to other AVs and use cases to develop a personalised testing plan to assess any required performance.

[9] During the QEOP site visit, a particular intersection of paths near the Copper Box Arena was identified as a high-risk area due to cyclists crossing at high speeds directly perpendicular to the pod path.

Static hazards: (repeat for both 'pod approaching static hazard' and 'pod stationary moving off with static hazard')

Hazard	Position	Straight line manoeuvre	Right turn manoeuvre	Left turn manoeuvre
Hazard type 1	Position 1			
	Position 2			
	Position 3			
	Etc…			
Hazard type 2	Position 1			
	Position 2			
	Position 3			
	Etc…			
Hazard type 3	Position 1			
	Position 2			
…	…			
Hazard type n	Position n			

Dynamic hazards:

Hazard	Scenario	Timing	Straight line manoeuvre	Right turn manoeuvre	Left turn manoeuvre
Hazard type 1	Scenario 1	Timing mark D1			
		Timing mark D2			
		Etc…			
	Scenario 2	Timing mark D1			
		Timing Mark D2			
		Etc…			
	Etc…	Etc…			
Hazard type 2	Scenario 1	Timing mark D1			
		Timing mark D2			
		Etc…			
	Scenario 2	Timing mark D1			
		Timing Mark D2			
		Etc…			
…	…				
Hazard type n	Position n	Timing mark n			

6. References

Department for Transport (2015), The Pathway to Driverless Cars: A detailed review of regulations for automated vehicle technologies. ISBN 978-1-84864-152-5. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/401565/pathway-driverless-cars-main.pdf

Department for Transport (Centre for Connected and Autonomous Vehicles). Code of Practice: Automated vehicle trialling, 2019. Available online: https://www.gov.uk/government/publications/trialling-automated-vehicle-technologies-in-public/code-of-practice-automated-vehicle-trialling (accessed 20/11/19).

Department of Transportation (2018) Automated Vehicles 3.0. Preparing for the future of transportation. U.S. Department of Transportation

ISO 26262-1:2018, Road Vehicles – Functional Safety, International Organisation for Standardisation. Available online: https://www.iso.org/standard/68383.html (accessed 29/01/19).

Jansen, R., Lotan, T., Winkelbauer, M., Bärgman, J., Kovaceva, J., Donabauer, M.,…Van Nes, N. (2017). Interactions with vulnerable road users. UDRIVE Deliverable 44.1. EU FP7 Project UDRIVE Consortium. https://doi.org/10.26323/UDRIVE_D44.1

Kienle, M., Franz, B., Winner, H., Bengler, K., Baltzer, M., Flemisch, F., Kauer, M., Weißgerber, T., Geyer, S., Bruder, R., Hakuli, S., Meier, S., 2014. Concept and development of a unified ontology for generating test and use-case catalogue

Law Commission (2019) Automated vehicles: Analysis of Responses to the Preliminary Consultation Paper. London: OGL

NHTSA (2017) Automated Driving systems 2.0. A vision for Safety. U.S. Department of Transportation

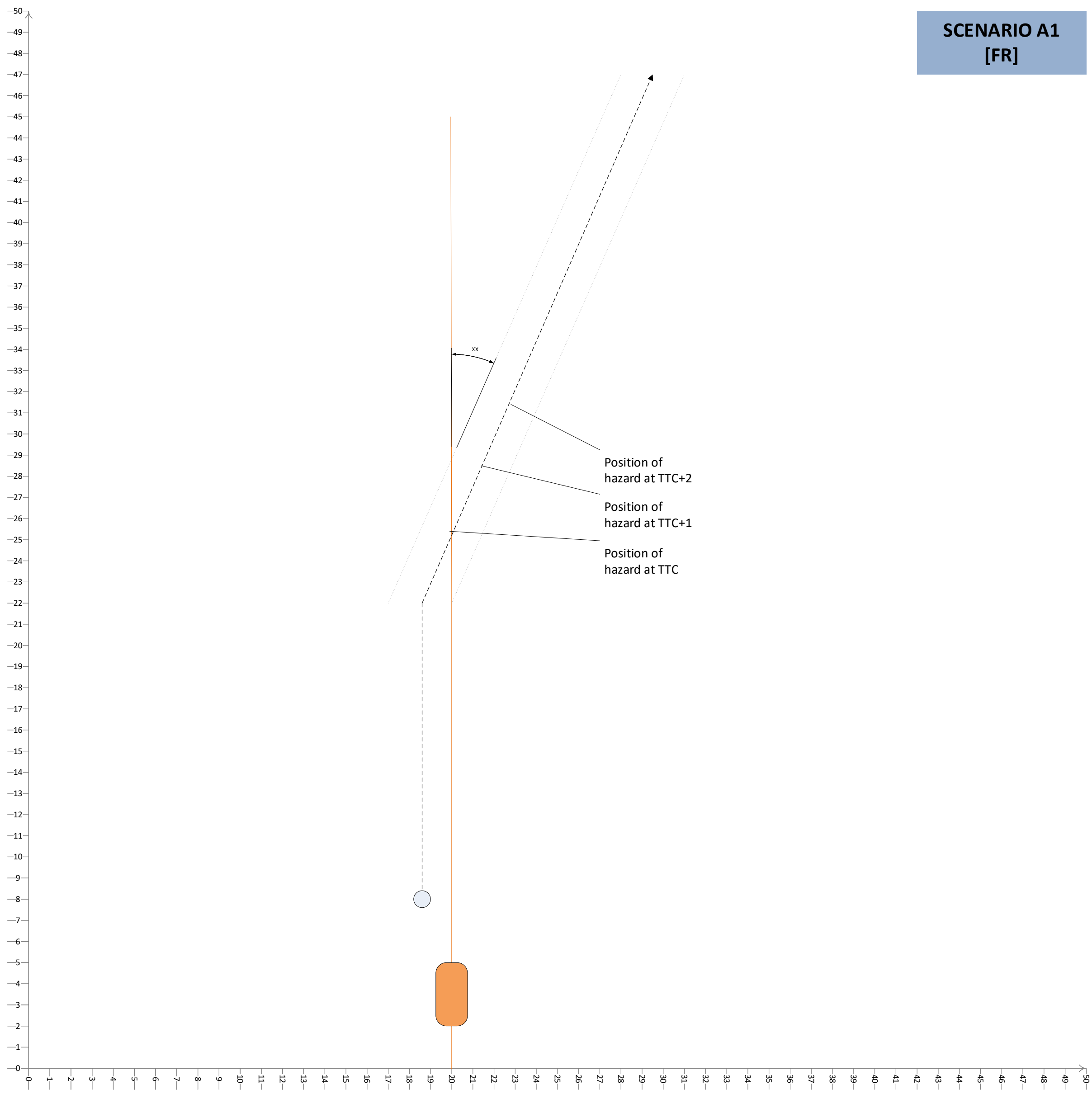
Ratingen, M.; Williams, A.; Lie, A.; Seeck, A.; Castaing, P.; Kolke, R.; Adriaenssens, G.; Miller, A. The European New Car Assessment Programme: A historical review. CJT 2015, Vol 19, pages 63-69.

Reed, S., Clift, L., Brown, L., Welsh, R. (unpublished). A User Centred Approach to Designing Scenarios for Autonomous Vehicle Validation.

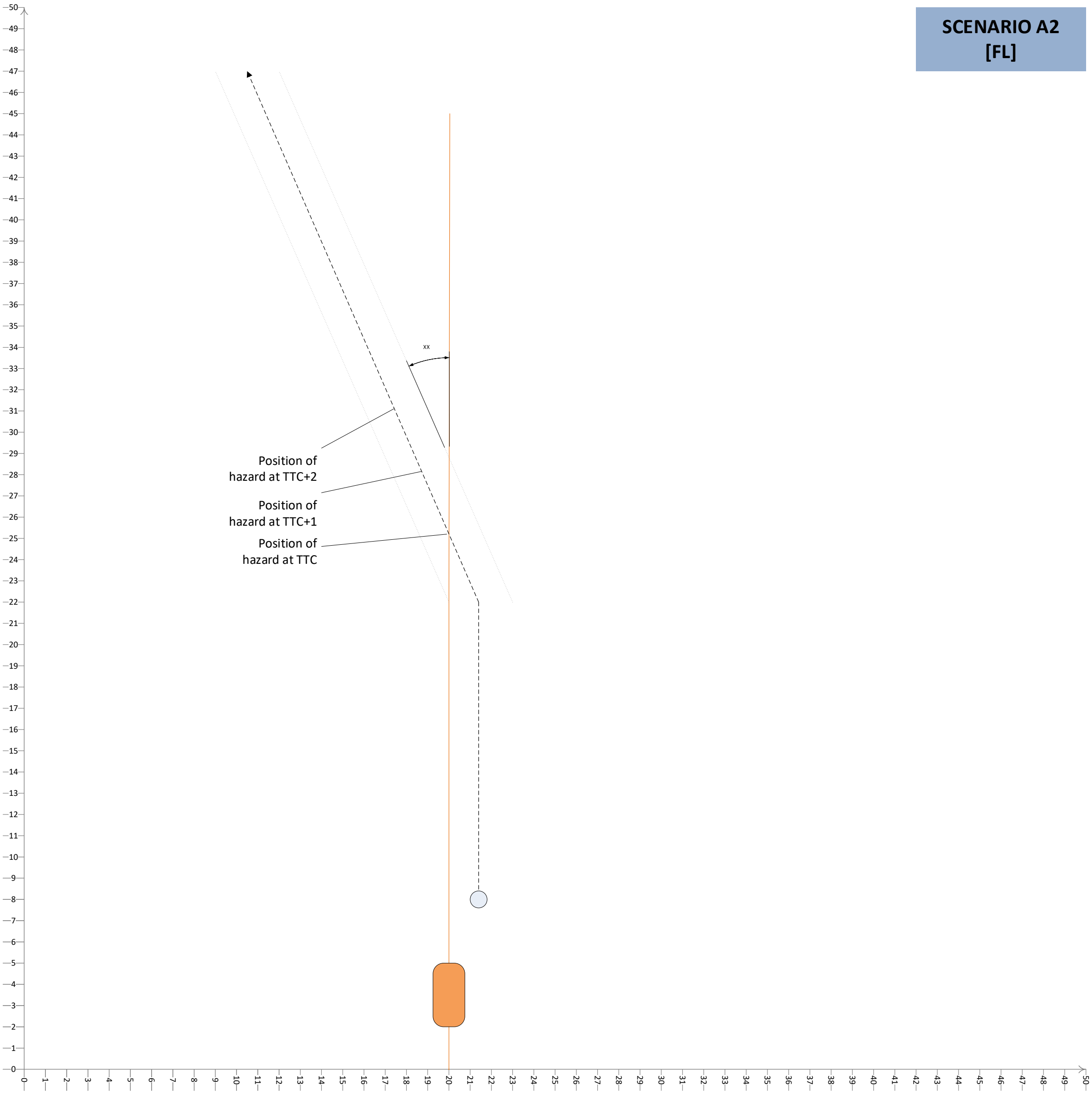
SAE J3016, Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, 2018. Available online: https://www.sae.org/standards/content/j3016_201806/ (accessed 29/01/19).

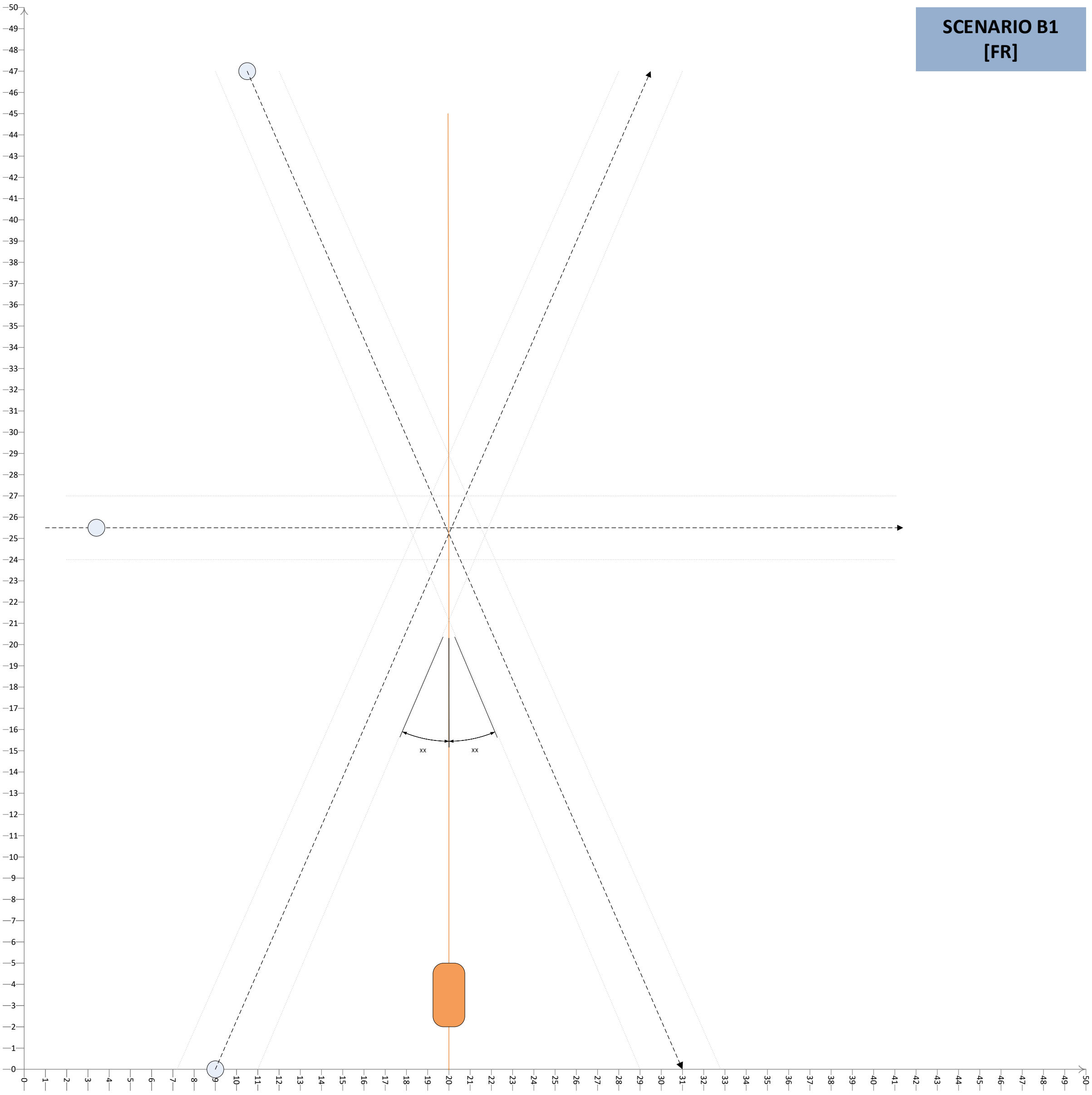
World Health Organization. (2018). *Global status report on road safety 2018: Summary* (No. WHO/NMH/NVI/18.20). World Health Organization.

Appendix A Dynamic Cycling Scenarios

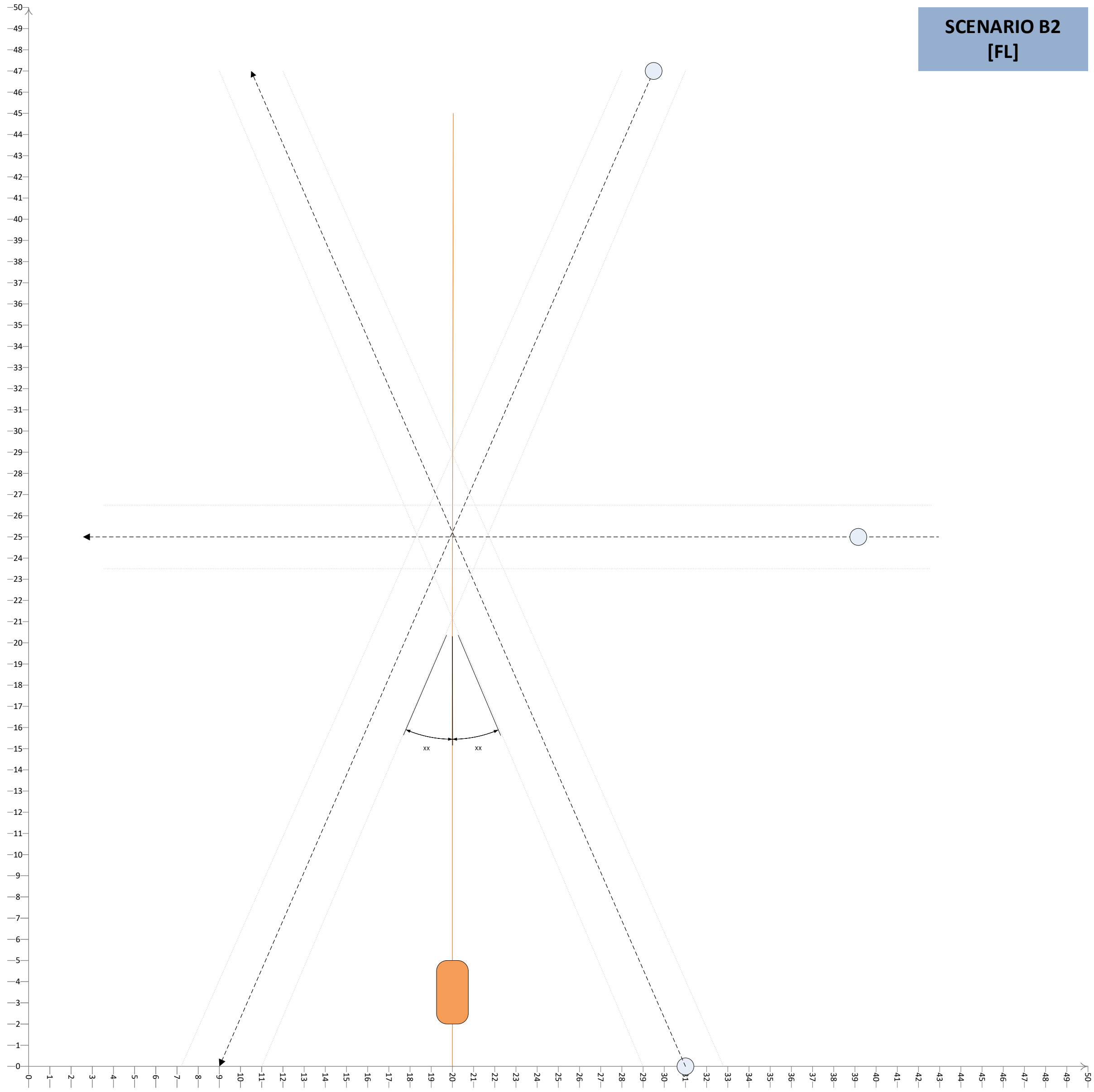


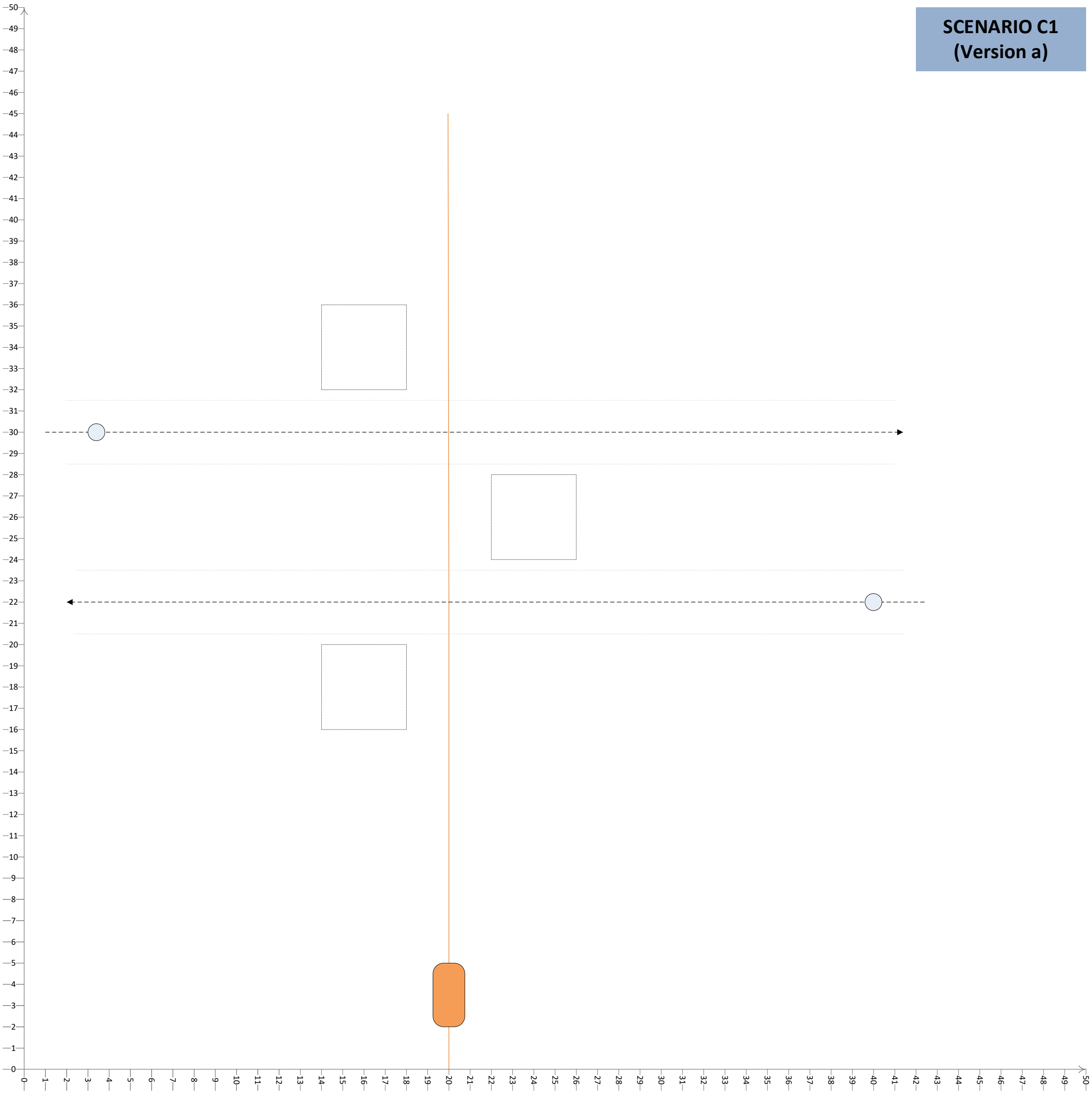
SCENARIO A2 [FL]

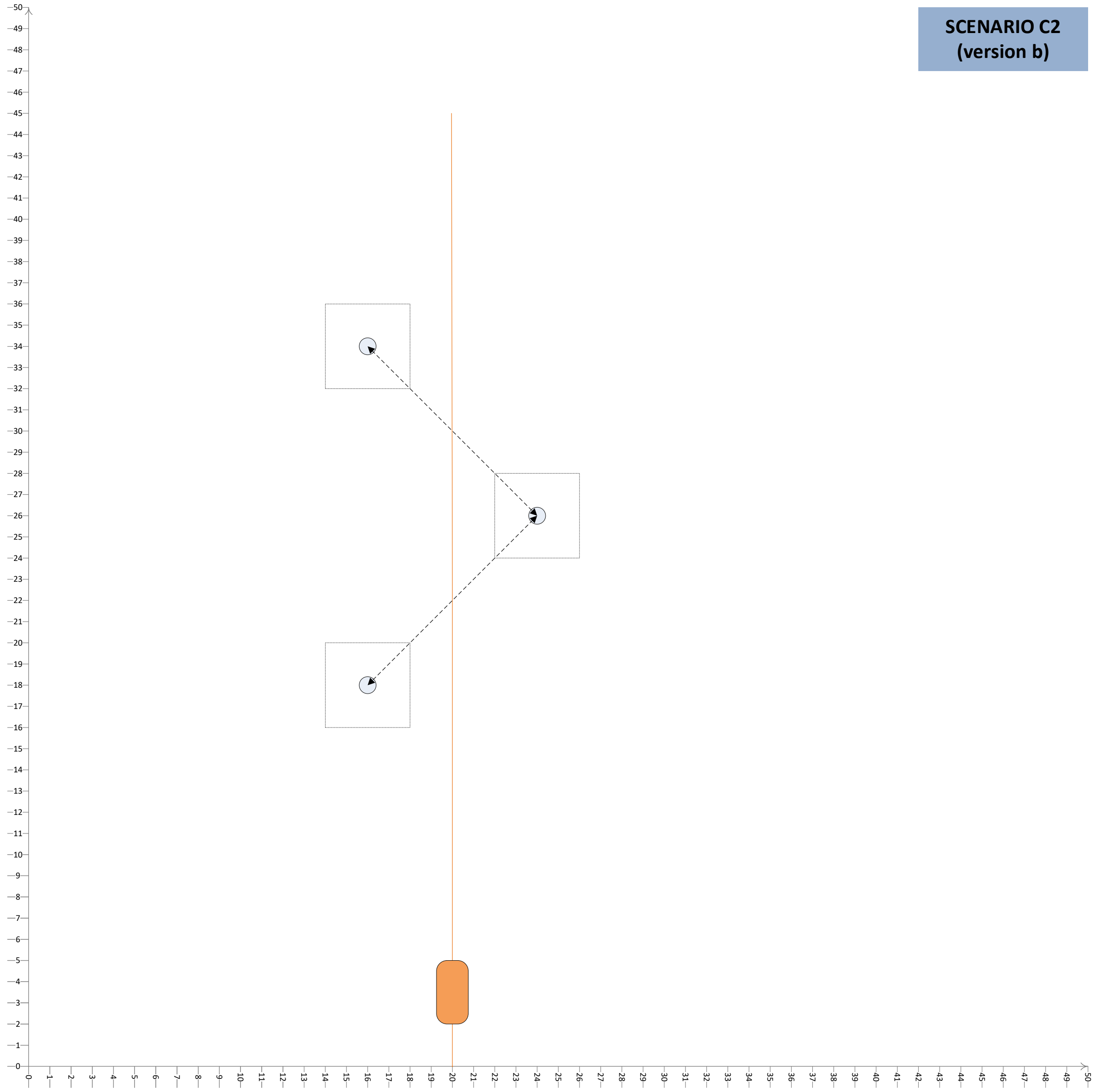


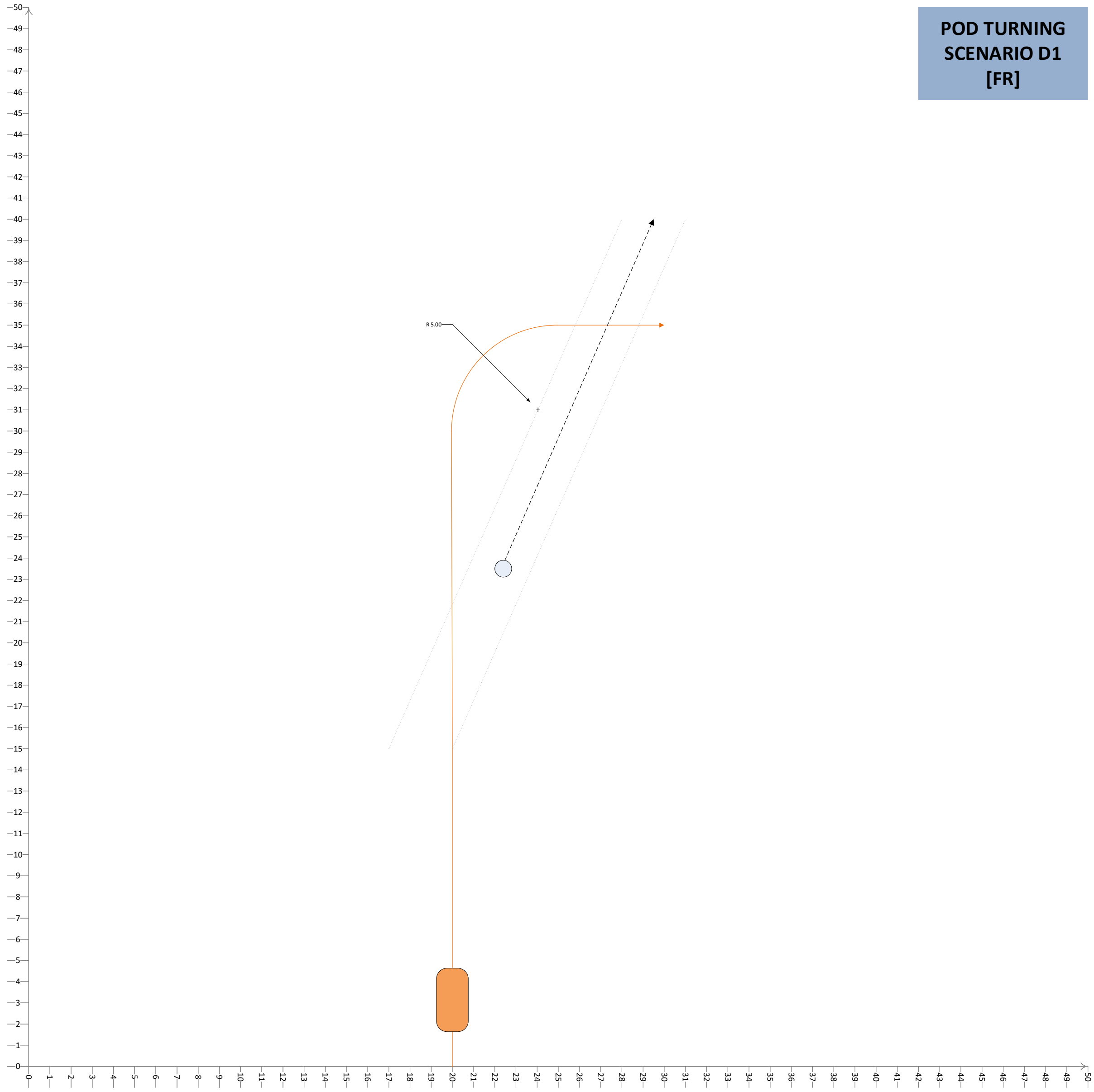

SCENARIO B1
[FR]
xx
xx

SCENARIO B2 [FL]

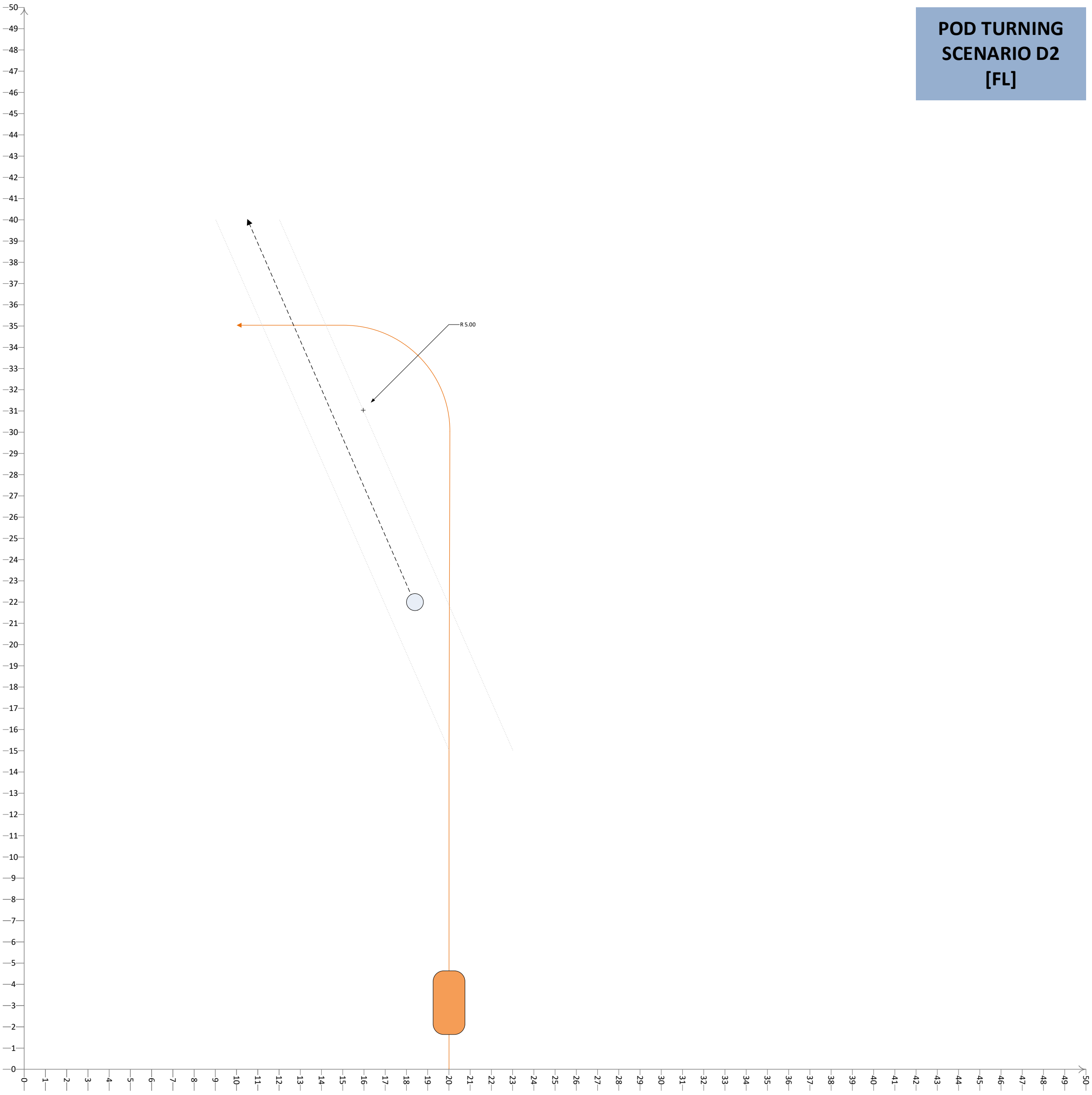


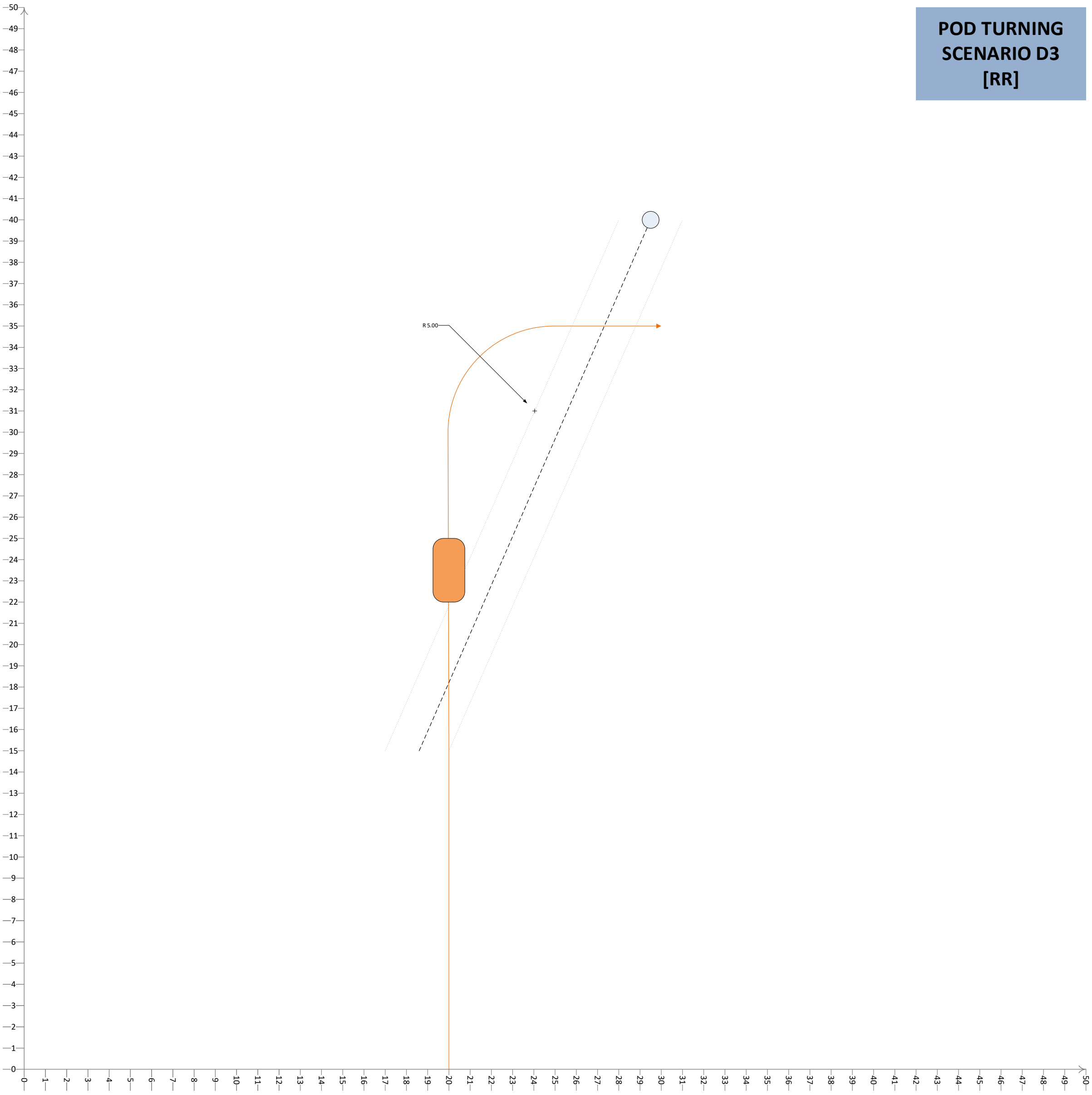

SCENARIO C1
(Version a)

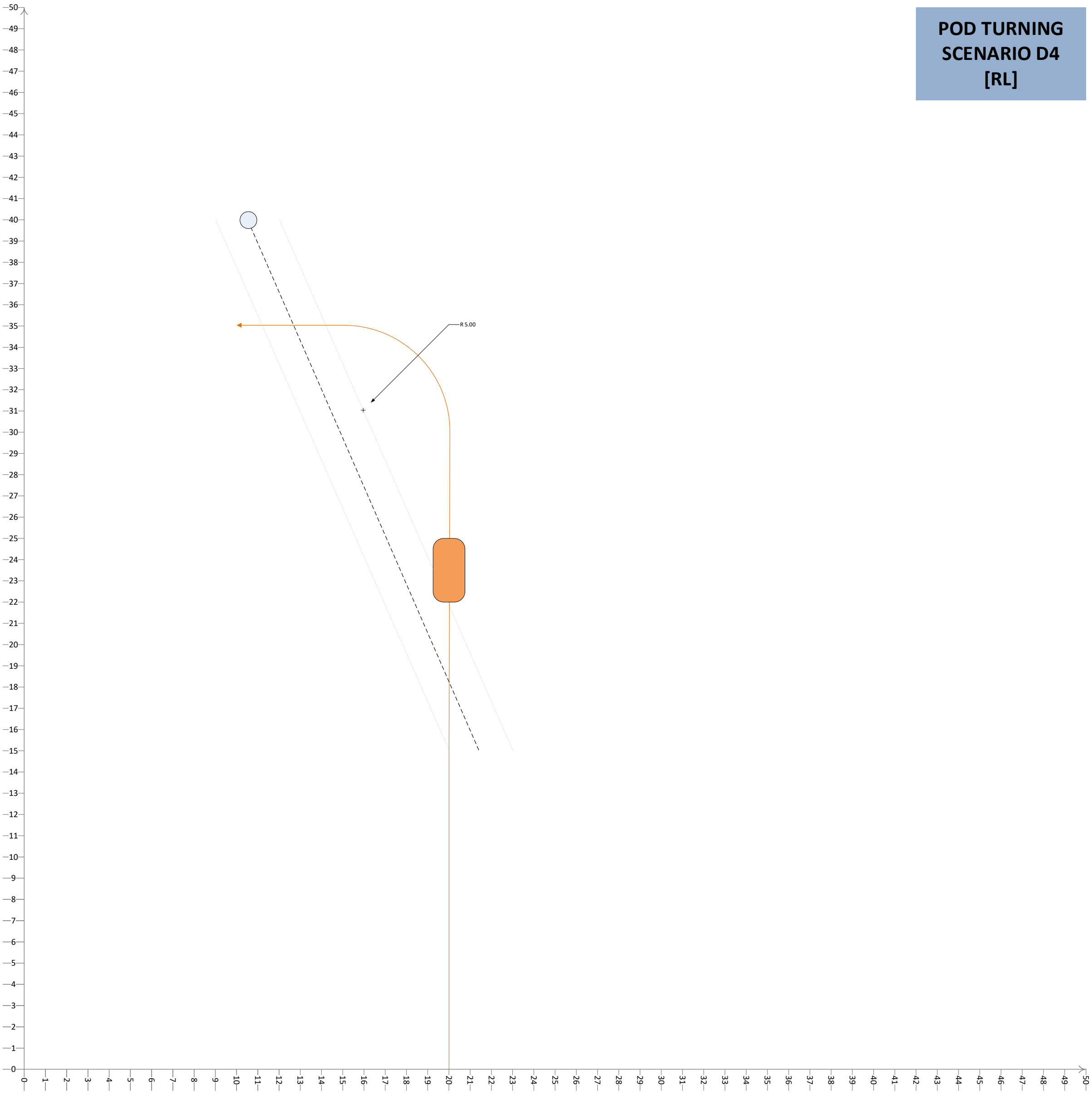

SCENARIO C2
(version b)

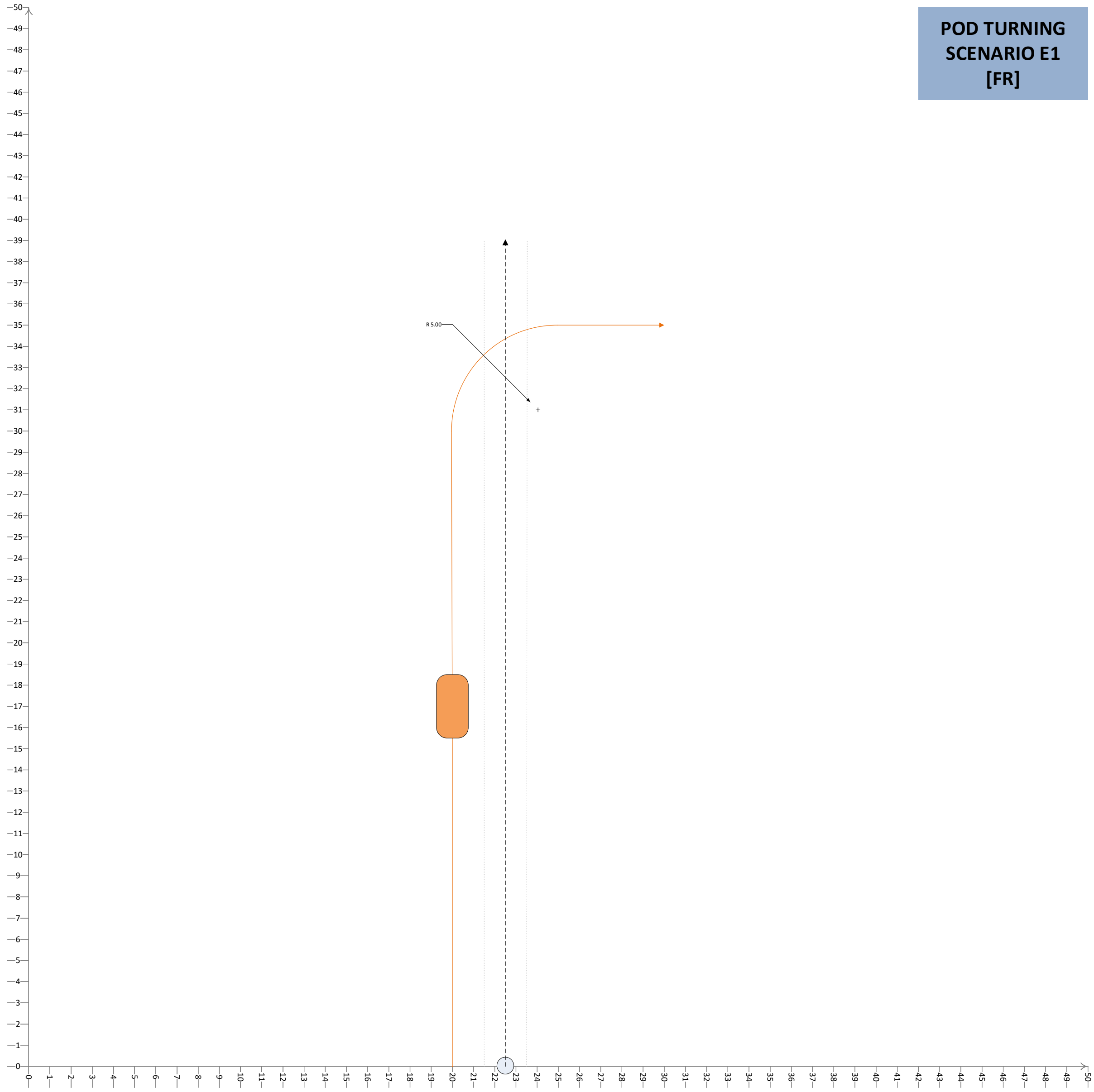

POD TURNING SCENARIO D1 [FR]
R 5.00

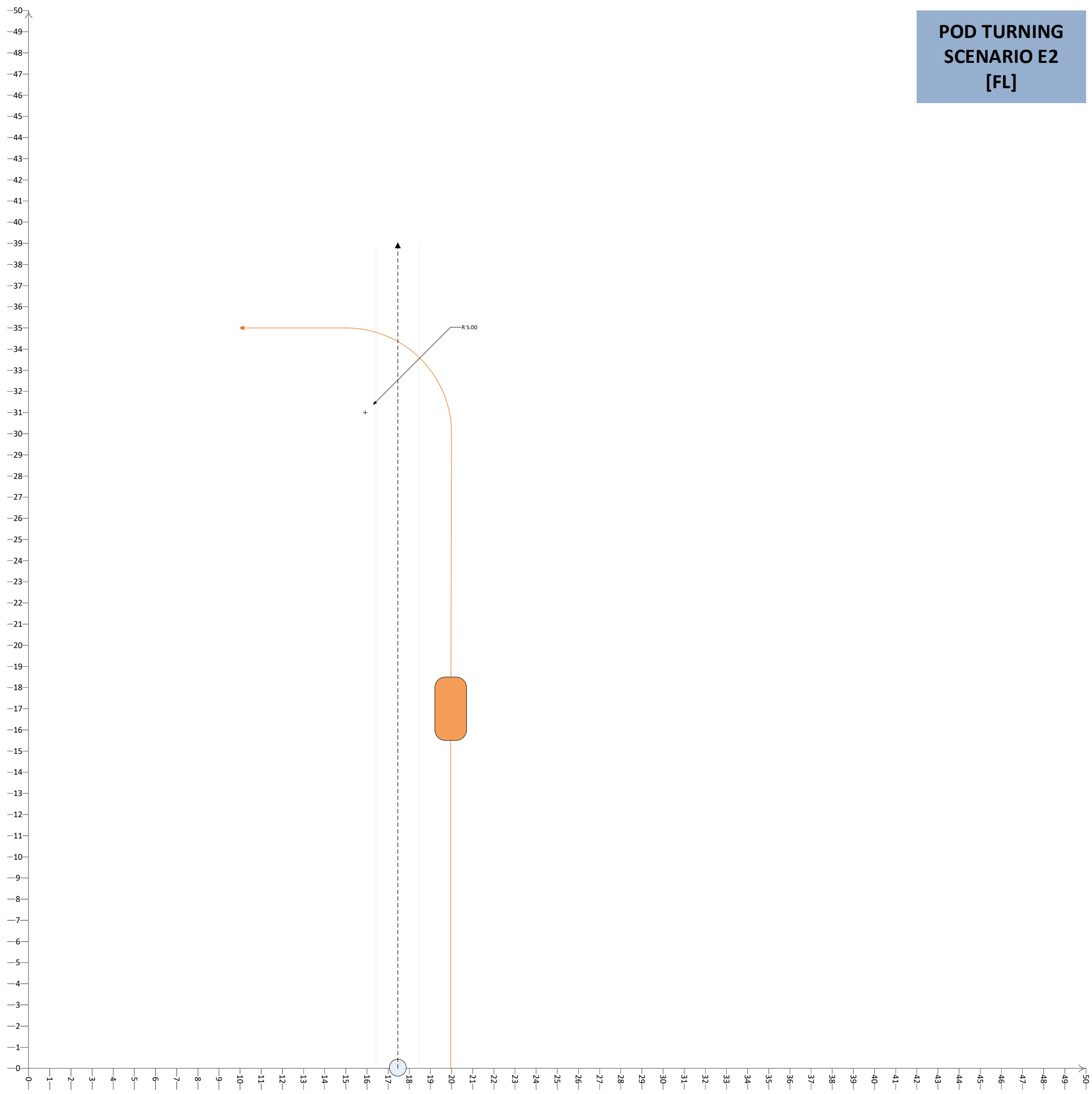
POD TURNING SCENARIO D2 [FL]




POD TURNING SCENARIO D3 [RR]
R 5.00


POD TURNING SCENARIO D4 [RL]
R 5.00


POD TURNING SCENARIO E1 [FR]
R 5.00


POD TURNING SCENARIO E2 [FL]
R 5.00

About Capri
Capri is an Innovate UK funded project that is being delivered by an AECOM-led consortium of 19 partners.

The purpose of the project is to capture the value of POD (pod on-demand) fleet services and develop an integrated solution for POD deployment, while building user, client and regulatory trust in POD vehicles that can travel on both pedestrianised areas and public roads.

For more information please either visit the website, tweet us or get in contact via the details below.

Contact

George Lunt
Project Manager
T +44 (0) 7921 646 425
E George.Lunt@aecom.com

Meghan Evans, MEng
Graduate Consultant, Transportation
T +44(0) 117 240 2974
E meghan.evans@aecom.com