QUADRIS final report

Mikael Ljung Aust
Jonas Bärgman
Malin Svärd

Datum: 2026-01-22

Table of Contents

1 Background, goals, and program relevance

A key question that should be asked repeatedly during the development of Advanced Driving Assistance Systems (ADAS) and fully Automated Driving (AD) applications is whether the technology, once deployed, actually will increase safety on the roads. Having a process and a toolchain that can deliver fast and accurate prospective safety benefit assessments for ADAS and AD is therefore critical to ensuring that we can consistently answer this question in the affirmative, allowing the Swedish automotive industry to deliver innovative, safe ADAS and AD technology and stay competitive.

Computer-based simulations are clearly faster and safer than field tests or experiments. As a result, virtual assessment is becoming the main evaluation method for both ADAS and AD. Examples include the previous projects P.E.A.R.S. and PEGASUS, as well as the recently concluded V4SAFETY. These projects are examples of where industry, institutes, and academia are collaborating to develop overarching standard frameworks for simulation-based safety assessments. Further, the NHTSA guidelines for AD development also require simulation-based assessments when new functionality is introduced or available functionality is modified, and simulations are being proposed as part of future legislation for the approval of self-driving vehicles.

So the automotive industry needs fast and accurate simulation-based safety assessment methods to a) even be allowed to bring AD to market (not least from a safety argumentation and liability perspective) and b) stay competitive by swiftly developing new, improved ADAS. However, since the area of simulation-based assessment for ADAS and AD is relatively new, several key knowledge gaps still exist, three of which were addressed in the QUADRIS project. Each of the three is described in further detail below, but for a fuller understanding of the challenges they represent, a brief summary of a typical simulation-based safety assessment process is first described.

The first step in this process is to create computer representations of the specific traffic safety problems to be addressed. This can for example be a particular crash scenario, such as a rear-end crashes on the highway or run-off-road crashes on rural roads. The data used to represent the problem typically comes from one or more in-depth crash databases with reconstructed pre-crash vehicle kinematics. Once this crash information has been digitalized, a set of these crash scenarios that can be run in a virtual simulation framework called a baseline batch: it represents the baseline (crash outcomes) that one wants to prevent.

Once a baseline batch has been created, simulations are run for each crash in the batch with the ADAS or AD technology under assessment activated, to answer the question, "What would happen if vehicle X had been equipped with technology Y in this crash situation?". These simulations, assessing what may have happened if a particular technology had been available in a particular crash scenario, are typically called "what-if" or counterfactual simulations.

QUADRIS was set up to close three specific knowledge gaps in this type of simulation-based safety assessment toolchain. The first concerns how to improve on the creation and validation of scenarios that can be used for virtual assessment of the effect of new safety systems. The second was a lack of a computational driver response models for assessment of ADAS addressing run-of-road crashes. The third was the need for computational driver reference models for use as safety targets in AD development. These three will be described in turn (Sections 2-4 below).

As part of the broader research activities within QUADRIS, and in close collaboration with the Horizon Europe project V4SAFETY. a comprehensive scoping review was also conducted to map and assess the landscape of computational human-behavior models used for virtual safety impact assessment. The aim was to complement the empirical and modeling-focused PhD work in this project by developing a systematic understanding of how human behavior is currently represented in virtual safety evaluations, where critical gaps remain, and what methodological foundations are needed to support credible and harmonized virtual safety impact assessment practices across Europe. This work is also reported (Section 5 below).

2 Scenario generation and validation for safety impact assessments of driving automation systems

Jian Wu is the Volvo Cars PhD student that has performed most of this research.

This part of the project addresses key methodological challenges in evaluating the safety impact of driving-automation systems using virtual simulation. As previously mentioned, because physical testing and real-world crash statistics are insufficient for assessing rare and severe crash outcomes, virtual safety-impact assessment has become an essential tool in both industrial development and regulatory evaluation.

The work focuses on how empirical pre-crash data can be transformed, extended and validated to support large-scale, simulation-based safety assessments. It is documented through a series of scientific publications that together develop an integrated framework spanning empirical data transformation, synthetic scenario generation, and representativeness validation.

2.1 Constructing representative empirical pre-crash reference datasets

Virtual safety impact assessment relies on comparing system performance in simulated pre-crash scenarios with and without driving automation systems. A prerequisite for credible assessment is the availability of empirical reference datasets that accurately reflect the real-world population of crashes. However, no single existing dataset satisfies this requirement. Naturalistic driving studies capture everyday driving behavior and low-severity crashes but contain very few severe events. In contrast, in-depth crash investigation databases provide detailed reconstructions of severe crashes but are strongly biased toward high-injury outcomes and lack exposure information.

This work addressed this challenge by developing a framework to construct representative empirical pre-crash reference datasets through model-based parameterization and data combination. Pre-crash scenarios from multiple data sources were first transformed into a common parameterized representation. Instead of working directly with heterogeneous time-series data, interpretable parameters describing lead-vehicle kinematics, following-vehicle behavior, and crash severity were extracted. This parameterization enabled systematic comparison and integration across datasets with differing structures, resolutions, and sampling strategies.

Because none of the available datasets alone could serve as a population-representative reference, data from multiple sources were combined to leverage their complementary strengths. Naturalistic data provided coverage of everyday driving and low severity rear-end crashes, while in-depth crash databases contributed detailed kinematic and behavioral information for severe crashes. However, naïvely merging these data sources would propagate their inherent biases, particularly the overrepresentation of severe crashes in in-depth databases. To mitigate this issue, sample-weighting methods were applied to align the combined datasets with known or more representative reference distributions for key parameters, most notably crash severity.

A k-nearest neighbors (KNN) sample weighting method was developed to address cases where conventional post-stratification is not applicable due to omitted strata or sparse parameter combinations. This approach enabled weighting even when certain regions of the parameter space were absent in the raw data, allowing the construction of reference datasets that better approximate the real-world crash population. The outcome of this work is not a single monolithic dataset, but a set of reference empirical datasets covering different subsets of parameters, each weighted to represent the real-world distribution as closely as possible given data availability. Together, these datasets form a robust empirical foundation for subsequent scenario generation and safety impact assessment, ensuring that simulated scenarios are anchored in real-world crash characteristics across the full severity range.

2.2 Generation of representative synthetic pre-crash scenarios

While the constructed reference empirical datasets improve representativeness, they remain too small to support large-scale simulation-based safety impact assessments directly. Severe crashes are rare, and robust estimation of safety effects—particularly for advanced automation functions—requires far more scenarios than are available empirically.

The second focus of this work was therefore the generation of large-scale synthetic pre-crash scenarios that preserve the essential characteristics of real-world crashes. To achieve this, parametric multivariate distribution models were built for the parameterized reference datasets. Parametric modeling was chosen to balance robustness, interpretability, and feasibility under limited and heterogeneous data conditions.

Marginal distributions of individual parameters were modeled using standard parametric families, while dependencies between parameters were captured through explicit modeling of linear correlations. Parameters exhibiting point-mass behavior were represented using mixture distributions to reflect distinct behavioral modes observed in the data.

Given the complexity and heterogeneity of real-world crash data, a single distribution model was often insufficient. Reference datasets were therefore partitioned into subgroups based on observed scenario characteristics, allowing simpler models to be fitted within each subgroup. The overall distribution was then constructed as a mixture of these sub-models, weighted according to their empirical prevalence. This approach enabled the synthetic data to reflect multiple patterns of real-world behavior while avoiding overfitting.

Synthetic pre-crash scenarios were generated by sampling from the fitted distribution models and reconstructing complete scenario trajectories using model-based representations of vehicle behavior. Lead-vehicle speed profiles were reconstructed using a kinematics model, while following-vehicle trajectories were generated using a behavior model that captures both nominal car-following and pre-crash braking responses. This model-based generation ensured that synthetic scenarios were not only statistically consistent with empirical data but also physically plausible and dynamically coherent.

Because discrepancies can still arise due to modeling assumptions, finite-sample effects or simulation constraints, a final sample-weighting step was applied to the synthetic scenarios. This step adjusted the contribution of individual synthetic scenarios to ensure alignment with the reference empirical distributions of key parameters. The resulting weighted synthetic datasets can be interpreted as representative approximations of the real-world population of rear-end pre-crash scenarios, serving as the primary input for virtual safety impact assessments of driving automation systems.

2.3 Validation of scenario representativeness using practical equivalence testing

Generating synthetic scenarios alone does not guarantee a credible safety impact assessment. The third component of this work, therefore, focused on development of a method for the validation of whether synthetic pre-crash scenarios are sufficiently representative of real-world crashes for the intended assessment purpose.

Traditional statistical significance testing is ill-suited for this task, as it is designed to detect differences rather than to assess practical similarity. To address this limitation, the work developed and applied a practical equivalence testing framework based on Bayesian inference and the concept of Regions of Practical Equivalence (ROPE). Instead of asking whether two datasets are statistically different, the framework evaluates whether their differences fall within bounds that are negligible for safety impact assessment. These bounds are defined in terms of assessment-relevant metrics, such as crash severity, temporal criticality, and braking behavior.

The validation framework proceeds through four main steps: selection of assessment-relevant metrics, definition of statistics and equivalence criteria, Bayesian inference, and overall equivalence assessment. Derived metrics are used in place of raw time-series signals to focus the validation on aspects of scenarios that directly influence system performance. Bayesian models are fitted to both empirical and synthetic datasets, allowing uncertainty to be explicitly quantified. Practical equivalence is established when the posterior distributions of the selected statistics lie entirely within predefined equivalence regions.

Building on this framework, the work further introduced binning-based statistics to support practical validation in high-dimensional and heterogeneous data settings. By partitioning reference distributions into bins and weighting them according to their relevance for safety outcomes, these statistics enable targeted assessment of representativeness in critical regions of the parameter space. This approach provides not only a validation decision but also diagnostic insight into where and why synthetic datasets deviate from empirical data.

Applications of the framework demonstrate that biased datasets—such as raw in-depth crash databases or unweighted simulation outputs—often fail practical equivalence tests on severity- or criticality-related metrics, even when they appear plausible at a superficial level. In contrast, synthetic datasets generated and weighted using the proposed methods can satisfy practical equivalence criteria across multiple metrics. These results highlight the importance of validation methods that are aligned with the objectives of safety impact assessment and that go beyond conventional hypothesis testing.

2.4 Summary

This work developed an integrated framework for virtual safety impact assessment of driving automation systems by addressing three interdependent challenges: constructing representative empirical reference datasets, generating large-scale synthetic pre-crash scenarios, and validating their representativeness using practical equivalence testing. By systematically integrating empirical data, generative modeling and validation methods, the framework supports transparent, interpretable, and credible safety assessments. The proposed approaches are directly applicable in industrial development and regulatory evaluation of driving automation systems and can be extended to other crash types and assessment contexts.

3 Computational modeling of driver responses to critical events

Malin Svärd is the Volvo Cars PhD student that has performed most of this research. She did most of her PhD under the FFi project QUADRAE, this covers the last part of her work.

This part of the project aimed to address the gap in understanding how drivers' visual perception influences their steering responses in critical run-off-road crashes. Fundamentals of driver perception and visually guided steering were examined to identify model components needed to construct a computational model of evasive steering in lane-departure situations. The work was documented in a scientific journal publication and formed the final contribution to a PhD thesis (Svard et al., 2023, 2024).

The work integrated three complementary data sources to capture a broad range of lane-departure severities and contexts: naturalistic crash and near-crash data from a large field study, baseline field-operational-test data representing everyday driving, and data from a high-fidelity driving-simulator experiment designed to induce severe lane departures during visual distraction. Because steering-wheel data was not available across all datasets, drivers' steering input was quantified using vehicle yaw rate relative to the road heading. All datasets included glance-behavior information from driver-facing cameras, enabling detailed analysis of drivers' visual attention.

In parallel, a complementary analysis of glance behavior was conducted based on data from a driving simulator study. The aim was to clarify how different visual stimuli capture attention and trigger gaze reorientation under distraction. Here, a looming-dominated lead-vehicle brake event was compared with a visually salient, close, but non-critical, cut-in characterized by lateral motion. The analysis explicitly addressed how data-selection choices affect measured glance response times and the results were presented at the Measuring Behavior International Conference on Methods and Techniques in May 2022 (Svard et al., 2022).

3.1 Visual perception in steering

Based on existing literature, angular measures were identified as the most influential perceptual variables for steering. These cues are directly available in the visual field and do not require complex cognitive processing, making them well suited for rapid decision-making in critical situations. Several angular metrics were examined in this project, all based on absolute yaw angle, relative yaw angle, and splay angle. Relative yaw angle is the difference between the vehicle's yaw and the road heading, while the splay angle represents the optical geometry of the lane edges as projected on the retina (a measure commonly used in aviation research). Both cues provide immediate feedback about lateral position and future trajectory, allowing drivers to estimate how far and how quickly they are deviating from the intended path.

In addition to angular cues, inverse time-to-lane-crossing (iTLC) was analyzed. Although iTLC is time-based and not directly perceivable, it has previously been effective for predicting steering interventions.

Drawing on previous research on intermittent control, steering was assumed to be executed in discrete adjustments rather than continuously. While a response to an unintended lane departure may involve several such adjustments, this project focused on the driver's immediate corrective action: the timing and amplitude of the first steering adjustment counteracting the departure.

3.2 Modelling the relationship between visual perception and corrective steering amplitude

Bayesian linear regression was used in the analysis of how the visual cues related to the amplitude of the primary corrective steering adjustment (i.e., the size of the steering correction). The most important finding was a quadratic relationship between relative yaw angle at the primary corrective steering adjustment and the corrective steering amplitude. This suggests the drivers scale their evasive steering to the severity of the lane departure situation, with small heading errors causing modest correction, whereas large errors trigger disproportionately stronger steering actions.

However, some differences were found between the data sets, with change in splay angle being an even stronger predictor than relative yaw angle in the naturalistic data sets. This suggests that different cues dominate the drivers' decision-making process depending on situational severity and context. A threshold model combining relative yaw angle and splay error achieved the highest overall performance. The insights from this analysis and the results of the linear regression models can serve as a basis in the construction of more complex computational models of driver behavior.

3.3 Timing of the corrective steering response

The timing of the primary corrective steering adjustment, measured from the end of the last off-road glance, was examined as a complement to steering amplitude. A central finding was that drivers across all datasets often initiated their first corrective adjustment before redirecting their gaze back to the road. This highlights the importance of peripheral vision in triggering evasive actions, even during visual distraction, and underscores the need to considering/incorporate peripheral-vision mechanisms in computational driver models.

3.4 Implications of data selection in glance studies

A complementary simulator study examined how visual cue saliency influences drivers' glance responses under distraction. The study compared a looming-based critical brake event with a non-critical but highly salient close cut-in characterized by strong lateral motion. Because distraction tasks often trigger incidental "check glances," several data-selection strategies were applied to filter out these unrelated glances and obtain stimulus-driven response times.

After removing check glances, drivers consistently redirected their gaze earlier to lateral motion than to looming, with some methods showing statistically significant differences. This pattern reflects the higher peripheral saliency of lateral motion compared to looming, despite the latter representing the greater true hazard. These findings underscore the importance of peripheral-vision sensitivity and the role of visual-cue saliency in triggering gaze shifts and initiating early evasive responses.

Together with the results on steering timing in unintentional lane departure situations—showing that drivers often initiate corrective actions before redirecting gaze—this complementary study reinforces the importance of explicitly modeling peripheral-vision sensitivity and visual-cue saliency in computational driver models. It also demonstrates that care should be taken to avoid masking true stimulus-driven reactions with unrelated glance behavior in future glance-response research.

3.5 Summary

This work investigates how drivers' visual perception shapes corrective steering in critical run-off-road situations. Using naturalistic crashes, everyday driving data, and driving-simulator experiments, the project analyzed glance behavior and yaw-based steering responses across a wide range of lane-departure severities. Angular visual cues—especially relative yaw angle and splay angle—were identified as key perceptual variables guiding steering. Bayesian modeling showed that corrective-steering amplitude scales nonlinearly with departure severity, while timing analyses revealed that drivers often begin corrective actions before directing their gaze back to the road, underscoring the important role of peripheral vision in triggering evasive maneuvers.

4 Reference driver models and safety targets for automated driving

Pierluigi Olleja is the Chalmers PhD student that has performed most of this research. Malin Svärd was the lead for the study on the relationship between glance behavior and inverse TTC.

This part of the project investigates how human behavior, data fidelity and contextual factors influence the safety performance and regulatory assessment of advanced driver-assistance systems (ADAS) and automated driving systems (ADS).
Across five scientific studies (reported in five papers), the work emphasizes the importance of incorporating real-world driver responses, realistic environmental conditions, and scenario-specific risk cues when designing safety-critical functions or evaluating automated systems. The research responds to the increasing reliance on virtual safety-assessment methods—a trend driven by the difficulty, cost, and limited representativeness of evaluating complex systems solely through physical testing or crash analysis (Olleja et al, 2022; Olleja et al., 2025).

Three central motivations guided the work. First, the representativeness of data used to generate virtual scenarios strongly influences the validity of safety assessments (Olleja, et al., 2022). Second, human reference-driver models (specifically those currently used in regulatory frameworks such as UNECE R157) sometime lack empirical grounding and may not reflect competent and careful human behavior in critical situations (Olleja et al., 2025). Third, driving behavior, including lane-change execution and visual attention, is strongly context-dependent. Any safety system or regulatory model that does not account for context therefore risks mischaracterizing both human and system performance (Jokhio et al., 2023; Olleja et al., 2026; Svärd et al., 2026).

The following sections elaborate on these themes and describe how the project advances methodological validity, behavior modeling, and the integration of human factors in ADAS/ADS development.

4.1 Which data to use when in virtual pre-crash safety assessment

A key contribution of this work is showing that data provenance and representativeness decisively affect virtual pre-crash safety-assessment outcomes. Paper I compares three common data sources used to construct crash and near-crash scenarios for evaluating automated emergency braking (AEB): a) GIDAS Pre-Crash Matrix (PCM) reconstructions, b) SHRP2 near-crashes, c) synthetic crashes generated from highD everyday-driving data (Olleja et al., 2022).

Synthetic crashes generated from highD (despite high resolution) systematically yield lower impact speeds, weaker lead-vehicle decelerations, and fewer true conflict events than reconstructed GIDAS crashes. This occurs because highD primarily reflects routine, low-criticality motorway driving. As a consequence, AEB algorithms appear substantially more effective when evaluated using highD-derived scenarios compared with scenarios based on real pre-crash kinematics.

Conversely, SHRP2 near-crashes capture authentic evasive behavior, sharper decelerations, and closer headways, producing synthetic crashes that align more closely with GIDAS PCM reconstructions.

The implication is clear: virtual safety assessments can substantially overestimate system performance if they rely on benign or unrepresentative datasets (Olleja et al. 2022). The work demonstrates that crash outcomes, scenario criticality and system-evaluated benefits vary markedly depending on the input data, thus challenging industry trends that prioritize dataset volume over behavioral richness.

4.2 Validation of driver reference models for ADS approval

Paper II extends the methodological critique to reference-driver models used in regulatory assessments. Under UNECE R157, an ADS must demonstrate performance at least equivalent to that of a "competent and careful" human driver. Two recommended computational benchmark models—the Competent and Careful Driver Model (CCDM) and the Fuzzy Safety Model (FSM)—were evaluated using SHRP2 cut-in near-crashes (Olleja et al., 2025).

The CCDM, which uses fixed reaction delays and a lateral-position threshold for cut-in detection, frequently responds too late, resulting in crashes that did not occur in the original dataset. This reflects the model's inability to incorporate urgency-related cues such as lateral speed or encroachment rate, which strongly influence human response timing.

The FSM exhibits the opposite behavior. Its fuzzy-logic threat assessment triggers braking too early, often before the lead vehicle has crossed the lane boundary, including events where human drivers did not respond at all. Such over-conservatism could lead to nuisance interventions and adverse effects on traffic flow.

Neither model consistently reflects the behavior of a "competent and careful" driver under cut-in conditions. The findings highlight the need for reference-driver models that incorporate both urgency, surprise and contextual cues, and that can be validated against real safety-critical data before being used in regulatory approval processes.

4.3 Characteristics of lane changes

This work also analyzed the characteristics of lane changes. The first study looked at more than a thousand lane changes recorded during baseline driving in the L3Pilot project (Jokhio et al., 2024). Using survival analysis to examine time-to-lane-change initiation (TTLCI) following turn-signal activation, the results provide additional information about signaling behavior and how to identify the initiation of a lane change.

Clear patterns emerged. Drivers signaled before initiating a lane change in only 60% of cases; in 33% they signaled after the maneuver had already begun, and in 7% they did not signal at all. Also, for those 60% of drivers who do signal prior to the lane change, they typically initiate the lane changes very shortly after signaling. Half of all maneuvers began within 0.7 seconds, and nearly 90% within 2 seconds. This is considerably shorter than the several-second signaling durations often assumed in design guidelines and suggests that signaling often reflects a moment of commitment rather than advance communication. Context also strongly affects TTLCI. Rightward lane changes take longer than leftward ones, multi-lane changes require longer waiting times, small lag gaps delay initiation, and higher speeds shorten initiation times.

These findings refine the understanding of natural lane-change behavior and have implications for safety, cooperation, and automated-vehicle prediction models. Specifically, they indicate a need for ADS prediction and control algorithms to incorporate realistic signaling patterns.

A key methodological element of this work was the use of Bayesian hierarchical modelling, enabling the analysis to capture variability across drivers and conditions while estimating how specific factors influence signaling behavior.

4.4 The importance of considering context in ADAS and ADS development

To better understand how context impacts glance behavior and safety, this work also analyzed off-road glance behavior in highway car-following as a function of time gap and looming, and estimated crash risk related to looming. It examined more than 6,000 steady-state car-following segments and shows that drivers adapt their glance behavior only weakly to risk variations induced by time gap changes. Although the very longest glances become shorter when time gaps are small, the median glance duration remains largely unchanged.

When these glance patterns are applied to a counterfactual crash-risk model, crash risk increases sharply as time gap decreases, demonstrating that natural driver adaptation is insufficient to maintain a constant risk level (Olleja et al., 2026).

Svärd et al. (2026) further identifies looming (inverse TTC) as a more sensitive cue guiding driver adaptation. Drivers rarely initiate off-road glances at loom rates above 0.2–0.3 s^{-1}, and long glances become extremely rare at higher loom levels. This indicates that drivers implicitly use optical expansion as a risk indicator and regulate their glance behavior accordingly.

Taken together, these findings show that contextual adaptation in drivers is present but often inadequate to fully control risk, that ADAS functions such as inattention-information systems (IISs) may be improved by incorporating contextual variables like looming or time gap rather than relying on fixed thresholds, and that reference-driver models should include realistic gaze behavior or explicitly justify decisions to use superhuman eyes-on-threat driver models.

Overall, the work shows that safety cannot be evaluated independently of context. Systems, drivers, and regulatory models all operate within dynamic environments, and overlooking this context can distort both risk estimation and evaluations of system performance.

4.5 The importance on including side check-glances in reference driver models

A pilot study (Olleja et al., 2022) also examined whether side check-glances (to mirrors or windows) should be included in reference driver models for side-swipe scenarios used in virtual safety assessment of automated driving systems. Using baseline highway data from L3Pilot around Gothenburg, the study manually annotated side-glance frequency and duration during straight, no-lane-change driving to quantify how often and how long drivers look sideways. These observations were paired with reconstructed Volvo Cars side-swipe crash kinematics, and counterfactual simulations were performed using simple driver-response models (steering S-maneuvers, braking and acceleration parameters, and assumed response times) to assess crash avoidance as a function of glance timing.

The analysis showed that crash avoidability can depend on whether the driver is glancing sideways at the critical moment; off-threat glances can delay responses enough to negate an otherwise sufficient evasive action. However, despite their behavioral relevance, the study found that including side check-glances changes safety-assessment outcomes by less than 2% for side-swipe scenarios. It therefore concluded that incorporating check-glances into ADS reference driver models is not necessary for this crash type.

4.6 A driver reference model for overtaking a cyclist

In another pilot work (Olleja et al., 2023), the project started the development of a reference-driver model for evaluating automated vehicles in cyclist-overtaking scenarios—an area that has received far less attention than highway car-following or cut-in situations. The motivation is that automated vehicles should perform at least as well as a “competent and careful” human driver when overtaking a cyclist. Building on earlier work, the study identified the safety-critical components of cyclist overtaking and formulated a model concept that sets a predetermined percentile of careful driving based on empirical data considered representative of careful and competent behavior. In this work, the 80th percentile was used.

The reference model builds on empirical findings and defines key safety metrics: minimum lateral clearance (MLC) as the primary indicator of cyclist safety; time-to-collision (TTC) to the cyclist when the driver initiates the overtaking; and the duration and timing of the steering-away phase. The comfort-zone boundaries (CZB) for these metrics were quantified independently for MLC and TTC, each using the 80th percentile to represent a careful driver. With these boundaries, the model initiates the lateral movement early and maintains sufficient clearance.

These boundaries were implemented in the esmini simulation platform, preliminarily assessing the 80th-percentile model against other percentiles to explore varying levels of conservatism. Simulation results show that the model can successfully complete overtaking maneuvers while adhering to cyclist-appropriate safety margins across a range of conditions. The work demonstrates that incorporating percentile-based comfort zones provides a structured and transparent benchmark for comparing automated-vehicle performance against human standards in cyclist-related scenarios.

However, it also highlights two important considerations. First, percentiles derived from everyday driving or experimental datasets—where drivers may satisfice rather than express their true comfort limits—are not equivalent to percentiles of the actual comfort-zone boundaries. Second, applying multiple criteria simultaneously (for example, requiring both a minimum lateral clearance and a safe TTC at overtaking initiation) reduces the probability that a maneuver will meet the “competent and careful” driver criteria defined in this way. Together, these points emphasize the need for careful interpretation of percentile-based models based on measured data, when using them as safety benchmarks.

4.7 Summary

This work examines how human behavior, data fidelity, and contextual factors influence the safety performance and regulatory assessment of ADAS and ADS. Across five main studies, it shows that virtual safety-assessment outcomes depend strongly on the representativeness of the underlying data, and that commonly used reference-driver models may not reflect competent human behavior in critical events. The research further demonstrates that lane-change behavior and visual attention are highly context-dependent, and that ignoring this can distort both risk estimation and system evaluation. Overall, the project advances methods for scenario generation, model validation, and integrating human factors into ADS safety assessment.

5 Scoping review on human behavior models for virtual safety impact assessment

As part of the broader research activities within this project—and in close collaboration with the Horizon Europe project V4SAFETY—a comprehensive scoping review was conducted to map and assess the landscape of computational human-behavior models (CBMs) used for Virtual Safety Impact Assessment (VSIA). The aim was to complement the empirical and modeling-focused PhD work in this project by developing a systematic understanding of how human behavior is currently represented in virtual safety evaluations, where critical gaps remain, and what methodological foundations are needed to support credible and harmonized VSIA practices across Europe.

The review shows that, although virtual simulation has become central to assessing pre-crash safety measures—including driver-assistance systems, automated driving features, infrastructure changes, and regulatory interventions—the behavioral models underpinning these simulations vary widely in maturity, scope, and validation. CBMs are already widely applied, but have evolved across multiple research traditions, producing a fragmented methodological landscape rather than a unified framework suited specifically to VSIA. To address this, the paper proposes a new categorization: everyday traffic-interaction models (ETIM), behavior-related crash-causation models (BRCCM), critical-event response models (CERM), and models of road-user responses and adaptations to safety measures (MRASM). This structure is intended to support both developers and assessors in selecting behavior models aligned with the safety metrics, objectives, and scenario types relevant to a particular VSIA.

Across the literature, the review identifies substantial imbalances. Most existing models describe routine, non-critical interactions—often developed originally for traffic-flow simulation rather than safety analysis. In contrast, models representing crash-causation mechanisms, realistic critical-event responses, or human adaptation to safety measures are comparatively rare. Vulnerable road users and heavy-vehicle drivers remain particularly underrepresented. Moreover, many models lack validation against real-world crash outcomes, limiting their usefulness for assessing injury or fatality risk. Validation is typically conducted only at the behavioral level (e.g., reproducing speed profiles or headway distributions), with very limited evidence that simulated conflicts or crash outcomes align with real-world severity distributions (e.g., relative speed at impact or injury risk).

A central conclusion is that robust VSIA requires three elements:

1. a clearly defined evaluation scope and safety metrics.
2. baseline scenarios that accurately represent the target traffic population.
3. sufficiently accurate models of how road users respond to, and adapt to, the safety measure being evaluated.

Many current VSIA applications fail at least one of these requirements, either by using CBMs outside their intended domain (for example, applying ETIMs to crash simulation), by relying on unvalidated scenario-generation pipelines, or by omitting behavioral adaptations despite their documented influence on real-world safety outcomes.

The review also identifies methodological challenges shared across model categories, including limited availability of high-resolution crash data, biases in crash databases, and the need for integrated modeling architectures that can transition smoothly from routine behavior to critical events. While emerging AI- and machine-learning-based approaches show promise, their ability to generate representative crash-level outcomes remains largely unproven.

5.1 Summary

This scoping review provides a conceptual foundation for developing harmonized methodologies for prospective, simulation-based safety assessments that rely on CBMs. It clarifies where standardized model requirements are needed, where new CBMs must be developed, and how validation strategies should be expanded to ensure credibility. As a complement to the PhD research, it offers a systems-level perspective that situates the empirical studies and modeling contributions of the thesis within the wider scientific and regulatory landscape of virtual safety assessment.

6 Dissemination and outlook

A main aim with dissemination of the results of the QUADRIS project were to make them available and to be applied by the project partners working with safety assessment in the automotive industry. This aim has been successful; Volvo Cars is already today starting to apply what has been learned in QUADRIS in both its ADAS safety assessment as well as within the safety argumentation work for Autonomous Drive.

Also, the results have reached a much wider audience through all the doctoral publications that have been part of the project work. A full list of project publications can be found below, including three finished or near finished PhDs.

The results from all three knowledge gap closure efforts have also been used in the particular contexts they were intended for. The reference driver modelling work has had an impact on the ongoing UN-ECE discussions on how to set up a regulatory framework for AD approvals within WG13. In this discussion, the QUADRIS work has clearly shown that the initial models proposed in the UN-ECE framework to capture the behavior of a careful and competent driver were flawed in several ways, and that improvements are necessary. As for the work on baseline batches and how to validate them, that has played a part in discussions with companies like Waymo regarding how the latter assess the safety of their driving technology. Finally, the work on visual perception in steering has found an important role in discussions with EuroNCAP on how to define and assess visual driver distraction.

Finally, it is worth mentioning that the work in QUADRIS provided an important foundation for a following project application called QUADRARUM, which was approved by FFI in the fall of 2025 and which continues to address some of the knowledge gaps identified for QUADRIS, in addition to tackling new challenges.

7 Published and/or presented work from QUADRIS

Jokhio, S., Olleja, P., Bärgman, J., Yan, F., & Baumann, M. (2024). Analysis of Time-to-Lane-Change-Initiation Using Realistic Driving Data. IEEE Transactions on Intelligent Transportation Systems, 25(5), 4620-4633. https://doi.org/10.1109/TITS.2023.3329690

Jokhio, S., Olleja, P., Bärgman, J., Yan, F., & M., B. (2023). Exploring Turn Signal Usage Patterns in Lane Changes: A Bayesian Hierarchical Modelling Analysis of Realistic Driving Data. IET Intelligent Transport Systems, 18(2), 393-408. https://doi.org/10.1049/itr2.12457

Olleja, P., Bärgman, J., & Lubbe, N. (2022). Can non-crash naturalistic driving data be an alternative to crash data for use in virtual assessment of the safety performance of automated emergency braking systems? Journal of Safety Research, 83, 139-151. https://doi.org/https://doi.org/10.1016/j.jsr.2022.08.011

Olleja, P., Bärgman, J., & Markkula, G. (2022, October 19-20). Quantification of driver's side-glance frequency and duration in straight highway driving. [Conference presentation]. The 8th International Conference on Driver Distraction and Inattention, Gothenburg, Sweden. https://ddi2022.org/wp-content/uploads/2022/10/3.1_Pierluigi_Olleja_DDI.pdf

Olleja, P., Markkula, G., & Bärgman, J. (2025). Validation of human benchmark models for automated driving system approval: How competent and careful are they really? Accident Analysis & Prevention, 213, 107922. https://doi.org/https://doi.org/10.1016/j.aap.2025.107922

Olleja, P., Rasch, A., & Bärgman, J. (2023, October 26–27). A reference driver model for overtaking a cyclist ICTCT, Catania, Italy. https://www.ictct.net/wp-content/uploads/35-Catania-2023/69-Olleja.pdf

Olleja, P., Svärd, M., Zhao, M., & Bärgman, J. (2026; almost complete, will be submitted Feb 2026). *Analysis of the safety impact of off-road glances in car-following driving scenarios.*

Svärd, M., Markkula, G., Ljung Aust, M., & Bärgman, J. (2024). Using naturalistic and driving simulator data to model driver responses to unintentional lane departures. Transportation Research Part F: Traffic Psychology and Behaviour, 100, 361-387. https://doi.org/https://doi.org/10.1016/j.trf.2023.11.021

Svärd, M., Olleja, P., & Bärgman, J. (2026; almost complete, will be submitted Feb 2026). *Drivers' glance-behavior adaptation in response to looming during car-following.*

Wu, J., Flannagan, C., Sander, U., & Bärgman, J. (2025). Model-Based Generation of Representative Rear-End Crash Scenarios Across the Full Severity Range Using Pre-Crash Data. IEEE Transactions on Intelligent Transportation Systems, 26(10), 15932-15950. https://doi.org/10.1109/TITS.2025.3573386

Wu, J., Flannagan, C., Sander, U., & Bärgman, J. (2024). Modeling Lead-Vehicle Kinematics for Rear-End Crash Scenario Generation. IEEE Transactions on Intelligent Transportation Systems, 25(9), 10866-10884. https://doi.org/10.1109/TITS.2024.3369097

Wu, J., Sander, U., Flannagan, C., Zhao, M., & Bärgman, J. (2025). Practical equivalence testing and its application in synthetic pre-crash scenario validation. arXiv:2505.12827. https://arxiv.org/abs/2505.12827

Wu, J., Sander, U., Flannagan, C., Zhao, M., & Bärgman, J. (2026; almost complete, will be submitted Feb-March 2026). *Practical validation of synthetic pre-crash scenarios.*

7.1 Licentiate thesis

Olleja, P. (2024). Towards more reliable pre-crash virtual safety assessment: The impact of the choice of data types and reference driver models on the assessment of vehicle automation (Licentiate thesis, Chalmers University of Technology). Chalmers Research. https://research.chalmers.se/publication/543009/file/543009_Fulltext.pdf

Wu, J. (2024). Generation of representative pre-crash scenarios across the full severity range using real-world crash data: Towards more accurate virtual assessments of active safety technologies (Licentiate thesis, Chalmers University of Technology). Chalmers Research. https://research.chalmers.se/en/publication/542271

7.2 Doctoral thesis

Svärd, M. (2023). Computational driver behavior models for vehicle safety applications. Chalmers University of Technology.

Olleja, P. (in prep). Driving Behavior and Safety Targets: A Naturalistic Perspective. Chalmers University of Technology

Wu, J. (in prep). "Generation and validation of representative pre-Crash scenarios: toward accurate and credible safety impact assessments of driving automation systems"

7.3 Other references

Svärd, M., Olleja, P., & Bärgman, J. (n.d.). *Drivers' glance-behavior adaptation in response to looming during car-following* (Paper V).