A Framework for Precautionary Safety for Automated Driving Systems—Toward Tactical Decisions Fulfilling Quantitative Risk

Citation for the original published paper (version of record):
Gyllenhammar, M., Rodrigues de Campos, G., Sandblom, F. et al (2026). A Framework for Precautionary Safety for Automated Driving Systems—Toward Tactical Decisions Fulfilling Quantitative Risk Acceptance Criteria. SAE International Journal of Connected and Automated Vehicles, 9(4). http://dx.doi.org/10.4271/12-09-04-0028

ARTICLE INFO
Article ID: 12-09-04-0028

doi:10.4271/12-09-04-0028

A Framework for Precautionary Safety for Automated Driving Systems—Toward Tactical Decisions Fulfilling Quantitative Risk Acceptance Criteria

Magnus Gyllenhammar,[1,2] Gabriel Rodrigues de Campos,[1] Fredrik Sandblom,[1] Martin Törngren,[2] and Jonas Fredriksson[3]

[1]*Zenseact AB, Sweden*
[2]*KTH: Kungliga Tekniska Hogskolan, Sweden*
[3]*Chalmers University of Technology, Sweden*

Abstract

Safety of Automated Driving Systems (ADSs) is arguably one of the main remaining barriers before widespread market deployment. While there exists a plethora of methods for planning a trajectory that fulfils certain constraints, what those constraints should look like, to enable effective planning of safe trajectories, is still being discussed. In this article, we generalize the concept of Precautionary Safety (PCS) and present a framework providing constraints on the tactical and operational decisions of the ADS. Such constraints consider the ADS' capabilities, the external conditions, knowledge of statistically relevant events and behaviors of other traffic actors, as well as the controllability of these events. The proposed framework enables assessment of the statistical fulfilment of quantitative risk acceptance criteria (QRACs), including requirements on accident, injury, and fatality rates. The framework further provides a means to dynamically adapt the constraints used for trajectory planning, i.e., to adapt the driving to the situation at hand. A case study, considering a possible collision scenario with a jaywalking pedestrian and a rear-end collision with a trailing vehicle, is provided to showcase the applicability and usefulness of the presented framework. The simulation-based case study displays the safety benefits from considering QRACs with multiple injury risk levels and further shows how the proposed PCS framework can be applied in practice.

History

Received: 04 Dec 2025
Revised: 25 Feb 2026
Accepted: 08 Jun 2026
e-Available: 08 Jul 2026

Keywords

Automated driving system, Quantitative safety, Quantitative risk norm, Tactical decisions, Precautionary safety

Citation

Gyllenhammar, M., Rodrigues de Campos, G., Sandblom, F., Törngren, M. et al., "A Framework for Precautionary Safety for Automated Driving Systems—Toward Tactical Decisions Fulfilling Quantitative Risk Acceptance Criteria," *SAE Int. J. of CAV* 9(4):2026, doi:10.4271/12-09-04-0028.

ISSN: 2574-0741
e-ISSN: 2574-075X

1. Introduction

Automated Driving Systems (ADSs), defined as SAE automation level 3–5 [1], promise to disrupt our transportation system with increased efficiency and safety as a result. However, for such promises to be realized, there is a need to ensure safe decision-making. Such decision-making needs not only to follow traffic rules but also to be resilient to the statistically relevant behaviors and events where other actors might violate said rules. Various formal method techniques have been suggested to guarantee that the ADS—by design—does not violate specified (traffic) rules [2] and that the ADS consequently does not expose other road users to dangerous behaviors. While this is a good, and perhaps even necessary start, these methods are (i) not well-suited to capture all statistically relevant operational uncertainties of the ADS [3–6]. Moreover, they (ii) struggle to accommodate rare cases with high-severity outcomes. Furthermore, such methods commonly (iii) do not adapt with respect to the situation awareness (SAW) [7, 8] but rather consider static assumptions defined during the design of the ADS.

Precautionary Safety (PCS) [9, 10] has been proposed to address all three (i–iii) of these shortcomings. First, it enables the ADS to adapt its driving policy to account for exposure to external events—thus capturing (i) the statistically relevant cases, including those with (ii) high-severity outcomes. Second, PCS incorporates SAW (iii), which includes the ADS's capabilities, both in terms of available resources of the vehicle platform as well as its ability to perceive the surroundings. Moreover, as highlighted in this article, PCS accounts for the controllability of the situation by considering the ADS's ability to identify and execute appropriate evasive maneuvers. The PCS framework also enables the fulfilment of Quantitative Risk Acceptance Criteria (QRACs) [11] that simultaneously considers multiple safety requirements covering accident, injury, and fatality rates, e.g., as suggested through the Quantitative Risk Norm (QRN) [12]. QRACs have been suggested, in the recent ISO 5083 [11], to cover aspects of both functional safety [13] and the safety of the intended functionality [14].

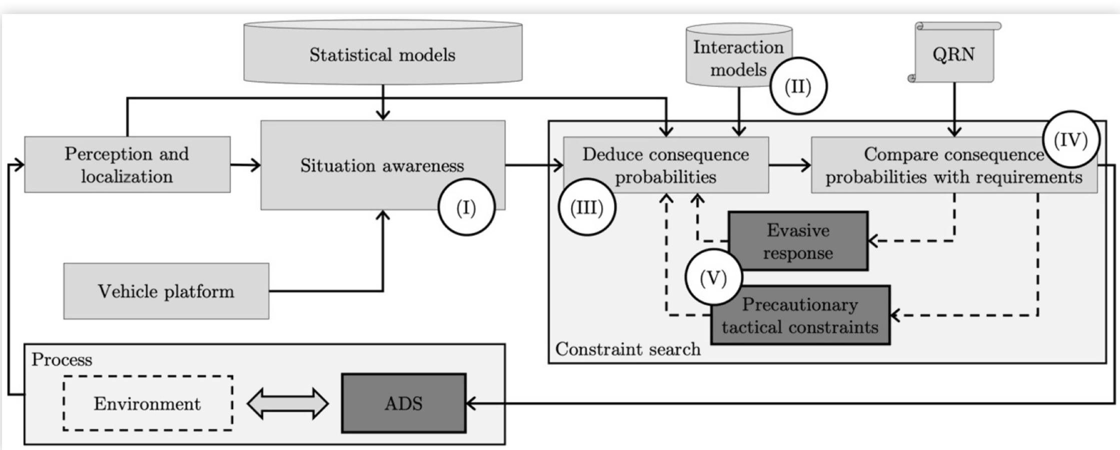
In this article, we present a PCS framework, shown in Figure 1, capable of deriving constraints on the driving policy of the ADS. The (I) SAW and the (II) interaction models are used to (III) deduce the consequence probabilities of the relevant adverse events. The deduced consequence probabilities are (IV) compared to the prescribed QRN to discern the validity of the constraints to the (V) tactical decisions and evasive responses of the ADS. The constraint search, encompassing steps (III) through (V), is iterated until the deduced consequence probabilities fulfil the requirements, after which the constraints are supplied to the ADS.

The contributions of this article are summarized as follows:

- a formalization, elaboration, and generalization of **PCS**, enabling knowledge of the QRN fulfilment at runtime. The formalization consists of:
  - a novel mathematical **formulation** for assessing the fulfilment of the QRN, and
  - a **framework**, based on this formulation, allowing the derivation of PCS constraints on the tactical decisions and evasive responses to fulfil the QRN;
- a **case study** of the framework—including collision with jaywalking pedestrians and rear-end collision with a trailing vehicle—showcasing the safety and performance benefits from the PCS approach. Moreover, the case study explores the implications of conditional exposure models.

A detailed comparison between the contributions of this article and previous work on PCS [9, 10] is given in Section 3.5.

FIGURE 1 The proposed PCS framework for deriving constraints on the tactical decisions and evasive responses of the ADS, while fulfilling QRACs.



The remainder of the article is organized as follows. First, definitions and notations central to the article are presented in Section 2 and a discussion on related works is given in Section 3. The PCS framework and the formulation supporting that framework are provided in Section 4. A case study showcasing the framework's usefulness is described in Section 5 and simulation results are presented and discussed in Section 6, where future research perspectives are also given. Finally, the conclusions are provided in Section 7.

2. Definitions and Notations

The following subsections present definitions and notations for the terms central to the remainder of the article.

2.1. Consequence Classes, ν_l

When introducing the QRN, Warg et al. [12] proposed the quantitative norm to be set as allowable frequencies of different consequences, or *consequence classes*. Each consequence class is denoted ν_l and, in practice, relates to the severity level of the incident. The QRN, however, goes beyond the severity levels commonly included in functional safety analysis of ISO26262 [13], by also encompassing quality level consequences such as vehicle damage or scared traffic participants [12].

2.2. Loss Event, L_k

An example of a *loss event* would be a collision with a jaywalker. In contrast to Warg et al. [12], that discuss incident types, we use loss events in the same way as Warg et al. use incidents. More precisely, a loss event includes a classification of the actors involved in the event, e.g., the ADS and a jaywalker (cf. Figure 4 of [12]) as well as details of the collision (e.g., impact velocity) and could be associated with multiple consequence classes.

2.3. Transfer Function, $P(\nu_l|L_k)$

There is a need to connect the loss events (L_k) and the consequence classes (ν_l). Generally, this can be modeled through a *transfer function*, $P(\nu_l|L_k)$, providing the probability of each consequence class given a loss event. The transfer function could be determined through consulting different data sources, such as, e.g., detailed crash reports. Given some selection of loss events, it might then be possible to have the transfer function $P(\nu_l|L_k)$ determined independently of the ADS implementation.

To limit the number of consequence classes relating to a specific loss event, it is useful to define the loss events more narrowly, e.g., L_1 = "Collision with jaywalker at less than 10 km/h" instead of "Collision with jaywalker." By narrowing the definitions of the loss events, we further support the more detailed PCS framework later explored in this article.

2.4. Scene, S_i

The scene, denoted by S_i, is taken to include any factor that belongs to the operational context of the ADS. Note that the factors consider in this article are by no means exhaustive. Rather, they present a suggestion of factors we have found useful in our modeling and discussions. The modeling aspects of the operational environment of the ADS are more thoroughly explored in publications related to the definition of the Operational Design Domain (ODD), see, e.g., [1, 15–18].

In defining the term scene, we draw upon the factors proposed by [19]. More specifically, we consider two main categories of factors:

- **Macroscopic factors** (similar to what one would include in an ODD): the geometry of the scene in terms of road geometry, R, and possibly surface conditions, e.g., road friction, μ; and encompass operational conditions more broadly, in the form of geographic position/road type, G, time-of-day, T, and weather, W; and
- **Microscopic factors** (pertaining to the specific operational scene): the state of the ADS, $\vec{x}$, (including velocity, acceleration, heading, etc.) and the configuration of other actors in the scene, A.

Taken together, the scene is the tuple: $S_i = \{R, \mu, G, T, W, \vec{x}, A\}$.

2.5. Adverse Event, E_j

An adverse event, $\boldsymbol{E_j}$, is an action of another traffic participant or a sudden event affecting the ability of the ADS to perform the dynamic driving task without incurring a collision. For example, it could be an aggressive cut-in into the lane of the ADS, a jaywalker suddenly crossing the road, or a tree falling over the road. The event is considered *adverse* if the resulting situation after the event is associated with a significant (e.g., above some threshold) increase in the likelihood of a loss event compared to the situation before the event occurred. In fact, a loss event should only occur if preceded by an adverse event. All non-adverse events can then be disregarded when assessing the probability of a loss event. The detailed definition of adverse events is not crucial to the case study explored in this article, but it would be useful for ensuring the efficiency of the approach outlined in this article, i.e., by avoiding analysis of non-adverse events.

3. Related Work

This section provides an overview of related work pertinent to the concepts discussed throughout the article.

3.1. Trajectory Planning

Finding suitable and safe trajectories given the scene and the control capabilities of the ADS is paramount in realizing a safe system. Such approaches can include uncertainty modeling of the obstacles [20–22], considering occlusions [23] as well as fulfilling chance-constraints using, e.g., stochastic model predictive control [24, 25]. This article does not explore the details on how to find such trajectories. Instead, the PCS framework is used to derive appropriate constraints that can subsequently be used within such trajectory optimization frameworks. For example, the set-based approaches in [20–22] consider observed obstacles and further make some assumptions on the occluded areas, but will not include all possible rare events, as this would lead to a prohibitively conservative model, as argued in [23]. The PCS framework instead ensures that relevant rare events are considered, therefore avoiding an overly conservative behavior while maintaining safety with respect to QRACs. The derived constraints from the PCS framework might also increase the feasibility of the trajectory planning frameworks to find safe trajectories, by positioning the ADS in a favorable state, anticipating possible (hidden) adverse events. Stated differently, if the trajectory planning framework is included as part of the ADS's evasive capabilities, the PCS framework can make sure to align the assumptions of the trajectory planner with the states later populated by the ADS—even when considering the risk of rare, yet catastrophic, events.

Trajectory planning decisions can also include injury minimization objectives, as explored in, e.g., [26] and [27]. Such contributions to the trajectory planning would also contribute to the final consequence estimates and would contribute valuable inputs to consider when evaluating the ADS and deciding on appropriate PCS constraints on the tactical and operational decisions.

3.2. Rule-Based (Formal) Methods

In [2], formal methods and techniques to comply with traffic rules are reviewed. Such methods can guarantee that the ADS will not cause any accidents as long as the other traffic participants also abide by those rules or assumptions, for example defined in [28]. However, this is not enough, considering that humans frequently violate traffic rules. For example, rule-based models, such as, e.g., [29–32] or collision-avoidance systems relying on worst-case assumptions [33], struggle to capture rare case events that might lead to severe outcomes. Moreover, to avoid "highly conservative behavior," formal approaches might make assumptions about the possible areas that other actors can traverse [23]. The PCS framework, proposed in this article, addresses this by explicitly considering the exposure probability of rare (edge) cases. The provided PCS constraints thus accommodate such events while still avoiding prohibitively conservative behavior. Furthermore, the PCS framework allows for statistical fulfilment of quantitative safety requirements encompassing not only the rate of collisions but also the consequences, i.e., injuries and fatalities.

3.3. The Quantitative Risk Norm

The QRN concept was proposed as a tailoring of the HARA process of ISO 26262 [13] by Warg et al. [12], with quantitative requirements on the acceptable frequency of different loss events and consequence classes as the safety target. These quantitative requirements support a shift away from a process-based safety argument toward a quantitative, evidence-based safety process, in line with the QRAC proposed in the recent ISO 5083 [11]. The PCS framework presented in this article provides a clear means for achieving such quantitative safety goals by deriving constraints for the tactical and operational decisions of the ADS.

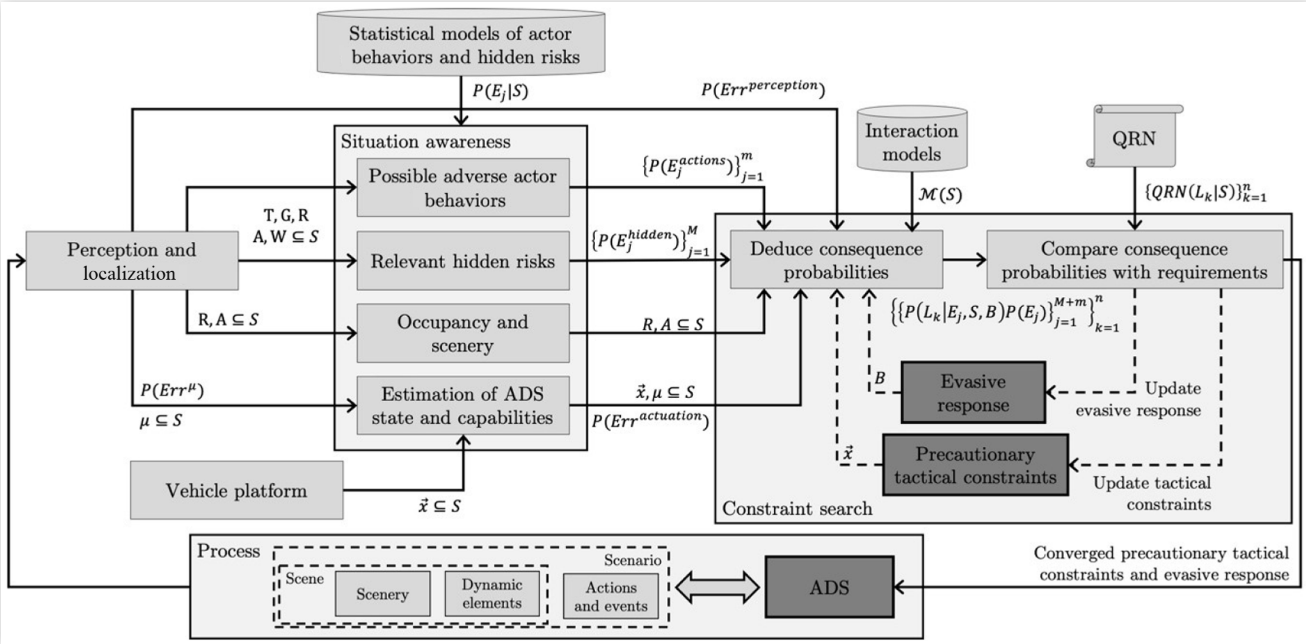
The QRN stipulates acceptable frequencies of loss events and consequence classes (cf. Figure 2 of [12]). The specific acceptable frequencies are an active topic of research and are not considered further in this article. However, it is worth noting that QRACs could include considerations beyond the classical *limit to risk of harm*. For example, Sandblom et al. [34] suggested also considering a *limit to transfer of harm* and a *limit to risk of harm in specific situations*. The latter consideration is akin to what we later discuss in this article, where the QRN level is posited per loss event for a specific scene.

For an overview of available (quantitative) risk definitions, we direct the curious reader to [35], or to [36], where a new risk acceptance criterion on the fleet level is proposed. Other aspects of risk, such as legal and ethical constraints, have also been approached but are not considered in this article. A summary of several existing safety definitions can also be found in [37].

3.4. Situation Awareness, Dynamic Risk Assessment, and Dynamic Safety Management

Situation awareness (SAW) has been proposed to support the safety assurance of an ADS by enabling the ADS to adapt its tactical decisions for safe operation [7]. The use of SAW has later come to be referred to as *tactical safety* [38, 39]. Underpinning such a SAW approach is the estimation of risks belonging to the operational situation. One way of achieving such risk estimations is called Dynamic Risk Assessment (DRA). How DRA can be realized and connected to the tactical decisions of an ADS is explored in, e.g., [40–42]. While Reich et al. [41] drew upon the integral risk approach of [43] to provide the risk estimate, there are other probabilistic approaches, such as Time-To-Critical-Collision-Probability (TTCCP) [44]. The commonality between these metrics is to consider

FIGURE 2 Details of the proposed PCS framework, supporting a process for deriving QRN fulfilling evasive responses (B) and tactical constraints ($\vec{x}$) of the ADS via loss event probabilities ($P(L_k|S,E_j,B)$) associated with the scene S and the adverse events (E_j).



trajectory distributions or predicted behaviors of the surrounding actors. In contrast to the risk estimates of [43] and [44], the PCS framework here provides a link to the fulfilment of the quantitative safety requirements of the QRN directly.

Moreover, in deriving the PCS constraints, we also consider the capabilities of the ADS. However, we not only incorporate knowledge of the internal capabilities of the ADS, as suggested in [8], but also consider error rates and uncertainties related to the perception system of the ADS, considerations that are also included in the risk map underpinning the systems proposed in [45–47].

The PCS framework presented in this article also differs from the SAW and DRA approaches, discussed above, in that all three factors supporting the HARA, i.e., exposure, controllability, and severity, are thoroughly and quantitatively integrated in our framework. For example, while the TTCCP of [44] and the integrated risk of [43] provide probabilistic estimates of the risk, but do so only through the binary lens of collision or no collision.

Based on SAW and DRA, Trapp et al. [48] proposed Dynamic Safety Management, which allows the system to "self-optimize its performance during run-time" [48]. In [49] this notion of DRA is taken one step further to create *safety awareness* by analyzing the factors underpinning the HARA. Khastgir et al. [50] also suggested dynamically calculating the factors of the HARA at runtime based on the current operational situations. Both [50] and [49] suggest that the ADS adapt to this dynamic HARA. But while [50] provides a detailed example, [49] remains rather conceptual.

Compared to [48–50], the PCS framework, presented in this article, does not rely on a predetermined hazard analysis. Instead, we directly connect the risk levels to the fulfilment of the quantitative requirements of the QRN. Not relying on a predetermined hazard analysis might be more suitable for an ADS, considering the difficulty of enumerating all possible hazards, see, e.g., [51, 52]. Furthermore, the outcome of the proposed QRN assessment is a set of constraints directly related to tactical decisions and evasive responses of the ADS, without any need to consider the integrity levels of the components used, as in [50], or the "relevant goal space," as in [49]. Moreover, we consider more nuances in assessing the factors of the HARA, including considering statistically possible adverse events—i.e., not only relying on trajectory distributions of nominal behaviors—incorporating interaction models between the ADS and other traffic actors as well as considering the evasive ability of the ADS when assessing the controllability.

3.5. Precautionary Safety

The tactical safety enabled by SAW is taken one step further through the concept of *Precautionary Safety* (PCS) [9], where a connection between a QRN and the driving policy of an ADS is established. The proposed planning module for precautionary safe decisions accounts for knowledge of the evasive capabilities of the ADS, the exposure rate to adverse events, and the performance limitations of the ADS. The PCS concept is further developed in [10], where a PCS driving velocity is derived while accounting for the uncertainties in the statistical model of event exposure, as well as the models for errors in the perception system. Neither [9] nor [10] disclose implementation details for how the PCS framework should be used in practice and how to consider multiple different events beyond the jaywalking pedestrian or the appearing moose. Both of these aspects are, however, explored in this article.

In both [9] and [10], the tactical decision is exemplified in terms of a PCS driving velocity, corresponding to the maximum velocity that satisfies the considered QRN. To evaluate the PCS velocity, the ADS is exposed to the adverse event (a jaywalker in the case of [9] and a moose in [10]) and the evasive ability of the ADS is judged to determine the impact (i.e., loss event) probabilities.

The framework presented in this article differs from the aforementioned articles on PCS in four ways, by:

- Including the selection of evasive responses within the outputs of the PCS framework;
- Explicitly considering interaction models with other traffic participants;
- Presenting a general, detailed, and mathematically grounded formulation and framework for deriving the PCS constraints, fulfilling the QRN; and
- Incorporating the use of conditional exposure models based on SAW.

Moreover, we conduct a baseline comparison as part of the case study. The case study itself is also more elaborate than those presented in [9, 10], including multiple adverse events, a detailed interaction model with the trailing vehicle, as well as conditional exposure probabilities modulated by the operational situation (i.e., the presence or absence of a standstill vehicle).

To understand how the PCS framework can be used in the broader context of an ADS safety case, please refer to [53] for a safety case fragment that builds upon the PCS principles.

Note that the PCS framework, discussed in this article, does not relate to policy-level and organizational aspects of the deployment and operations of ADS, as covered in [54].

4. PCS Framework—Assessing the QRN

The following section presents the PCS framework and the mathematical formulation for linking the ADS's performance to the fulfilment of the QRN.

To fulfil the QRN, the probability of all possible loss events for all consequence classes needs to be less than the required level stipulated in the QRN [12]. Namely,

$$\sum_{k=1}^{n} P\left(\nu_l, L_k\right) < QRN\left(\nu_l\right) \qquad \forall \nu_l \tag{1}$$

Instead of relying on fleet-level statistics and measuring the fulfilment of the QRN retrospectively, *after* the consequences have occurred, we look for a way to provide a predictive statement of the safety of the ADS. This is what we set out to do with the presented QRN assessment framework.

4.1. Formulation for QRN Assessment

It is possible to rewrite the left-hand side of (1) as:

$$\sum_{k=1}^{n} P\left(\nu_l, L_k\right) = \sum_{k=1}^{n} P\left(\nu_l | L_k\right) P\left(L_k\right) \tag{2}$$

where $P(\nu_l | L_k)$ is the *transfer function* from a specific loss event L_k to a consequence class ν_l. For some selection of loss events, the transfer function could be determined independently of the ADS implementation, as discussed in Section 2. This is a key insight, as it allows us to focus on the second factor of (2), i.e., the probability of the loss event itself, $P(L_k)$. Unfortunately, the loss event probabilities are only measurable as a lagging metric post-deployment. Thus, to achieve a predictive measure of safety, we need to consider the possible adverse events, E_j, and possible scenes, S_i, associated with the loss event. By factorizing $P(L_k)$, we can rewrite it as:

$$\begin{aligned} P\left(L_k\right) &= \sum_i \sum_j P\left(L_k, S_i, E_j, B\right) \\ &= \sum_i \sum_j P\left(L_k | S_i, E_j, B\right) P\left(E_j | S_i\right) P\left(S_i\right) \end{aligned} \tag{3}$$

where $P(S_i)$ is the probability of the ADS being in a particular scene S_i, $P(E_j | S_i)$ relates to the exposure rate of the adverse event E_j given the scene, and $P(L_k | S_i, E_j, B)$ is the ADS's ability to avoid a loss event given the scene, the adverse event and the intended evasive response of the ADS, B. In contrast to $P(L_k)$, these factors are more easily measurable or estimated.

There are three things to note with respect to (3). First, the loss event can be avoided by the ADS or the other traffic participants present in the scene. This is encapsulated by $P(L_k | S_i, E_j, B)$.

Second, the probability of an adverse event given a scene ($P(E_j | S_i)$) corresponds to the exposure of the adverse event and is closely linked to the normal usage of exposure as relied upon for the derivation of ASILs, e.g., following ISO 26262 [13] processes. This factor is independent of the ADS itself but dependent on the operational conditions of the system, such as time-of-day, weather, and geographic position, i.e., the scene. To find the exposure rates of different adverse events, one can rely on collected operational data. Either the collected data is used to construct the model at each new update of the ADS, or it is constructed continuously using, for example, Bayesian inference. An example of the latter is explored in more detail in [10]. The models could also be constructed based on data from external sources or databases, including roadside sensors, naturalistic data sets, or governmental databases. It is also useful to consult available crash databases to discern the factors leading up to the crashes and analyze whether those factors constitute adverse events. In such cases, the crash database might also provide input to the exposure model.

Third, the probability of the ADS being in the scene in the first place is highly influenced by design choices—i.e., the selection of ODD—as well as the run-time decisions of the ADS—e.g., how to position the ADS within the traffic scene and at what speed to travel. The ADS should strive to avoid combinations of scenes and adverse events where the ADS struggles to avoid the associated loss events or where there is a low controllability on the part of the other actors. Avoiding such scenes can effectively be done by removing them from the ODD altogether. However, the high-risk scenes that remain in the ODD are still possible to avoid by appropriate constraints to the tactical decisions of the ADS, which is the purpose of the PCS constraints explored in this article.

4.2. Detailed PCS Framework for QRN Assessment and Derivation of Tactical Constraints

The PCS tactical constraints of the ADS should lead the system away from scenes with a high likelihood of resulting in a loss event, i.e., where the product of the two factors $P(L_k|S_i, E_j, B)$ $P(E_j|S_i)$ is large, in comparison with the QRN. While this sounds straightforward in theory, the question remains: how does one make such information available to the ADS at runtime, i.e., how does one know which scenes pose elevated risks and which do not? For that purpose, we suggest the framework presented in Figure 1, which is detailed further in Figure 2. Note that the two decision variables for the proposed PCS framework are the PCS tactical constraints and the intended evasive response, as indicated by the two dark gray boxes of Figure 2.

The ADS (dark gray box at the bottom of Figure 2) interacts with the surrounding environment, and this process is perceived by the perception and localization system of the ADS (to the left of Figure 2). The perception and localization system provides estimates of the current scene in terms of macroscopic scene factors: *R, T, G,* and *W*, as well as the microscopic scene factor of how other actors in the scene are positioned, *A*. Together with (statistical) models for the probability of different adverse events $P(E_j|S)$, this information creates the SAW of the ADS needed to assess the fulfilment of the QRN. The SAW provides information about the possible adverse events, either as adverse behavior of other actors ($E_j^{actions}$) or adverse events related to hidden risks (E_j^{hidden}).

Together with the vehicle platform, the perception and localization system also provides input to estimate the state and capabilities of the ADS, $\vec{x}$ and μ. Furthermore, the perception system also provides estimates of the error rates for each of the provided pieces of information, denoted $P(Err\{actuation, perception\})$.

To deduce the loss event probabilities, there is also a need to know the evasive capabilities of the ADS and, moreover, the intended evasive response, *B*, related to different adverse events—captured in the term $P(L_k|E_j, S, B)$. Additionally, there is a need for models describing the interaction ($M(S)$) between the ADS and the other actors in the scene. Altogether, this information is used to deduce the loss event probability related to the current scene, *S*.

It is not only relevant to assess the loss event probability of the perceived scene but also that of “adjacent” scenes, as indicated by the reported errors from the perception system.

Furthermore, the errors in perception also play a more direct role when deducing the loss event probabilities, as in the case study explored later, where the perception system sometimes fails to detect the adverse event altogether.

How the deduction of the loss event probabilities is done in detail is dependent on multiple factors related to the ADS, its intended use case, and ODD, as well as available statistical models of the different risks and available interaction models. The deduction can then draw upon different means for simulation, modeling, and sampling to achieve a reliable estimate. This is an area extensively explored in relation to scenario-based testing, where different sampling techniques, such as, e.g., importance sampling [55], adaptive importance sampling [56], and subset simulation [57], have been proposed to speed up the coverage of the testing. Along the same vein, in [58, 59], a process for leveraging subset simulation to conclude the probability of different severity level outcomes of a scenario is described and explored. Such a process would also be highly relevant in the context of efficiently deducing the loss event probabilities discussed in this article. In the interest of clarity and focus of this article, however, a general and detailed process for the consequence deduction is deferred to future work. That said, the case study in the subsequent sections provides a concrete and detailed explanation of how it can be done for a specific example.

Once the loss event probabilities have been calculated, they can be compared to the levels stipulated in the QRN, i.e., $QRN(L_k|S)$. If they are found to be in accordance with the QRN, the analyzed precautionary tactical constraints and intended evasive response are considered safe. If not, there is a need to alter the state of the ADS ($\vec{x}$) through different tactical constraints or to change the intended evasive response (*B*).

4.3. Context-Specific QRN

As part of the budgeting process for the QRN, it makes sense to consider requirements on the loss event level, and not only on the consequence level [12]. However, one might want to take this one step further and consider context-specific QRN levels, QRN(·| S_i), where the requirements are dependent on the operational conditions of the ADS, i.e., the scene S_i. Just as with the use of loss event–level QRN requirements, deciding to use context-specific QRN requirements is a choice that should support the safety assurance efforts of the ADS. Like the loss event–level QRN, a context-specific QRN also supports fulfilling the *limit to risk of harm in specific situations* [34]. However, while the situation-specific risk is covered by a context-specific QRN, it is important to also consider the fulfilment on the consequence level. For that purpose, it is central to see how the context-specific, the loss event level, and the consequence-level QRN requirements interact. The QRN requirements on any of the former levels might be warranted in isolation without a connection to the consequence level, but there is still a need to consider the total risk of harm as captured by the consequence-level QRN.

Before looking at that connection, it is worth noting that the scenes that condition the context-specific QRN levels should be coarse. The purpose here is not to make a detailed risk budget across any combination of all the factors making up the scene. Instead, context-specific QRN requirements benefit from being coarse in the sense of imposing different allowable frequencies for radically different types of operational contexts, such as, e.g., urban environments compared to highway driving or heavy rain versus sun and dry roads. In particular, the context-specific QRN should by no means encode any of the microscopic factors of the scene. This would both render a close to infinite number of scenes to consider as well as render the interconnection between the ADS's tactical constraints and the fulfilment of the QRN much more complex than necessary.

When determining the consequence-level QRN fulfilment from the loss event QRN, one would need to sum across all possible (coarsely distinguished) scenes to determine the collective contribution of risk from the context-specific QRN requirements. Using the law of total probability gives us

$$\sum_i w_i QRN\left(\nu_l | S_i\right) \le QRN\left(\nu_l\right) \tag{4}$$

with w_i being the weights corresponding to each scene. These weights need to be selected in such a way that they over-approximate the probability of the ADS being in the scenes with high risk, to make sure that this sum is a conservative and applicable estimate of the fulfilment of $QRN(\nu_l)$ when considering the propensity of the ADS to operate in any given scene S_i. These weights are closely related to $P(S_i)$. Note, however, that it is non-trivial to ensure specific probabilities for the ADS to operate in specific scenes without imposing arbitrary limitations on what missions to accept and when the feature will be available.

An alternative is to make sure that the QRN of each scene is, on average, below that of the complete QRN, i.e., $\frac{1}{\# scenes} QRN\left(\nu_l | S_i\right) \le QRN\left(\nu_l\right), \forall S_i$. However, requiring the QRNs for all scenes to be below the overall (average) QRN might drastically impact the performance in situations where this is a much too strict limit. For example, it is reasonable that there is much more risk associated with driving at night, in heavy rain at temperatures close to 0°C, than on dry roads with good illumination—at least with respect to some classes of loss events. Furthermore, by the same token, other contexts will instead allow a larger margin in terms of risk than otherwise warranted. Consequently, while being a simple, straightforward solution to the fulfilment of the overall consequence-level QRN, it is also associated with a few unwanted side effects.

Now, circling back to the proposed QRN assessment framework. Note that the subscript *i* has been removed from the *S* in Figure 2. The reason for this is that the ADS is already operating in a particular scene. Thus, there is no need to consider all possible scenes, as is the case when we determine the overall QRN fulfilment [c.f. (4)]. In particular, the probability of being in a specific scene has already "collapsed" when the ADS operates there. A context-specific QRN for the scene, i.e. $QRN(L_k|S)$, provides the link we need between the QRN and the tactical constraints of the ADS.

4.4. Run-Time Estimate of QRN Fulfilment

Performing a thorough and detailed simulation to assess the loss event probabilities for all scenes and considering all possible adverse events at runtime in the vehicle is likely not possible unless there are analytical models or simple, reliable heuristics available. However, that need not prevent the use of such information to elicit tactical constraints for the ADS at runtime. Instead, there are three different ways through which such information can be provided:

1. Through a pre-calculated look-up table, as suggested in [60]. An example of this is also provided through the case study presented in the following sections;
2. Using a surrogate model running in the vehicle, as suggested in [61]; and
3. Make a full (possibly simulation-based) assessment on a remote server and push the results to the ADS at runtime, following the concept outlined in [62].

These different approaches would correspond to changing the "*constraint search*" box on the right-hand side of Figure 1, while the other components given in the framework would act as inputs.

5. Case Study

The applicability of the proposed framework of Figure 2 is explored through a case study presented in this section. The case study involves the potential collision with a jaywalking pedestrian and a rear-end collision with a trailing vehicle, and the ADS operating on a single-lane road, illustrated in Figure 3. The two decision variables of the ADS considered here are: the evasive response of the ADS, B, incorporating a brake delay response of 300 ms, and the tactical constraints relating to the cruising speed, v_0. An instantiation of the QRN assessment framework pertaining to the case study is given later in Figure 5, highlighting the aspects of the framework that are applicable to the case study.

The case study requires establishing models, data, and statistics to populate several elements of the proposed framework. First, the interaction models ($M(S)$), covering both the jaywalking pedestrian as well as the trailing vehicle, needs to be established. Second, the ADS's capabilities to detect and act in the presence of the adverse event also need to be modeled. Third, the QRN levels ($QRN(L_k|S)$), toward which the deduced loss event probabilities can be compared, should also be determined.

Finally, the case study is extended to consider conditional exposure rates to the adverse event ($P(E_{jaywalker}|S)$), depending on the presence or absence of a standstill vehicle.

5.1. Frontal Collisions with Jaywalking Pedestrians

For modeling the loss event probabilities with the jaywalking pedestrian, the case study relies on the models developed and presented in [10]. These include a uniform probability of the distance to the jaywalker as it appears in front of the ADS, i.e., $d_{jaywalker} \propto U(0, D_{max})$, with D_{max} set to 150 m. The perception system of the ADS is modeled by λ with three different false negative (FN) error rates for three different perception ranges, $\lambda = \{[0,20] : 10^{-6}, [20,50] : 10^{-4}, [50,150] : 10^{-3}\}$, where the ranges given in meters and the FN rate given per hour. These perception ranges are illustrated as different shades of green in Figure 3. Compared to the model of [10], the present case study also considers a brake delay of 300 ms [63] between the detection of the pedestrian and the onset of the evasive braking response, a_{ADS}. This selection of brake delay models the characteristic of mechanical braking systems and is found to adequately approximate more elaborate brake system models [63]. Moreover, the pedestrians are also modeled to pass the lane of the ADS at a speed of 5 km/h, and the lane width is taken to be 5 m.

5.2. Trailing Vehicle Interactions—Modeling Braking Response

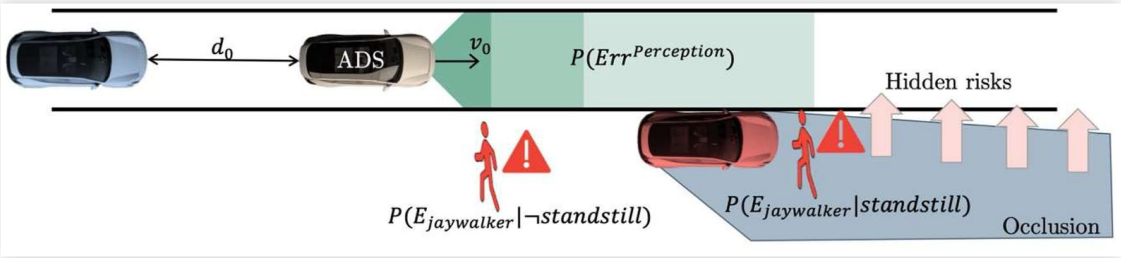
If there is a trailing vehicle present in the scene when the ADS needs to avoid collision with a jaywalker, there is a risk of a rear-end collision. For this case study, this interaction model ($M(S)$) shows the trailing vehicle models and the braking response of the driver in the trailing vehicle according to the distributions presented in [64]. Markkula et al. [64] analyzed time series data of rear-end conflicts with near-crash or crash outcomes from naturalistic data sets and provided a distribution of the driver behaviors across the analyzed data. For our modeling, we are looking for a way to determine the time at which the rear vehicle starts braking, t_B, and thus make use of the distributions of the brake onset timing, α_B, presented in Figure 10 of [64]. Furthermore, for simplicity and without loss of generality, we assume a constant acceleration of −8 m/s^2 on the part of the rear vehicle once the brake response has started. Note here that no brake delay is included, as the models from [64] considers the resulting trajectories rather than the acceleration request following the pressing of the brake pedal.

The brake onset timing, α_B, is introduced as a dimensionless quantity to normalize some of the dependencies between the situation and the response in the data analyzed by Markkula et al. [64]. The quantity is defined as

$$\alpha_B\left(t_B\right) = \frac{t_B - t_{0.2}}{t_C - t_{0.2}} \tag{5}$$

where t_B is the time of brake onset we are looking for, t_C is the time of collision if the driver does nothing to

FIGURE 3 Illustration of the considered scenario of the case study, including a jaywalking pedestrian and a trailing vehicle.



respond to the threat, and $t_{0.2}$ is a factor where the driver first sees a visual looming exceeding the level corresponding to an inverse Time-To-Collision (TTC) value of 0.2 s (related to τ^{-1} = 0.2 s^{-1} [64] (see p. 217)). For our purposes $t_{0.2}$ is taken to be the $max(0, t_C - 5)$, i.e., the time point where the TTC is equal to 5 s. The $max(\cdot)$ is used to capture the cases when the collision time is more imminent than 5 s without violating the model assumptions.

Markkula et al. split the provided distributions of the brake onset timing according to whether the driver has its eyes-on-threat or not. Note that the data does not include a significant portion of automated vehicles, but the eyes-off-threat cases are simply a means for Markkula et al. to fit a reasonable model to the data present.

Since we are looking for a means to model general driver behavior, we consider a weighted sum of the two distributions provided for eyes-on- and eyes-off-threat. The collective Cumulative Distribution Function (CDF) of this weighted sum, following the parameters given in [64], becomes:

$$F\left(t_B\right) = w_1\Phi\left(\frac{\alpha_B\left(t_B\right) - \mu_{\mathcal{N}}}{\sigma_{\mathcal{N}}}\right) + w_2\Phi\left(\frac{\ln\left(\alpha_B\left(t_B\right)\right) - \mu_{\log\mathcal{N}}}{\sigma_{\log\mathcal{N}}}\right) \tag{6}$$

where w_i are the weights, $\alpha_B(t_B)$ is the brake onset timing described above, Φ is the normal CDF, and μ and σ are the location and scale parameters of the respective eyes-on- and -off-threat distributions. For the purpose of this article, we use the car events (top panel of Figure 10 of [64]), the values for which are: $w_1 = \frac{94}{94+115}$, $w_2 = \frac{115}{94+115}$, $\mu_{\mathcal{N}}$ = 0.22, $\sigma_{\mathcal{N}}$ = 0.29, $\mu_{\log\mathcal{N}}$ = −1.40, and $\sigma_{\log\mathcal{N}}$ = 0.71. Now, for a specific case of the following distance d_0, initial velocity v_0, evasive manoeuvre a_{ADS}, and considering the QRN presented in Table 1, we can determine the specific brake onset times (t_B) that would result in each specific loss event L_k related to different impact velocities. While we have a different goal than Shin et al. [65], we acknowledge the need for considering the same cases to derive the collision conditions for our different brake onset times, i.e., whether the ADS has stopped before the collision or not.

Once the brake onset times, t_B, corresponding to the different impact velocities are determined, the corresponding probabilities, $P(L_k|brake)$, can be calculated using (6). An illustration of this process is shown in Figure 4, where the CDF is shown as a function of t_B for the values of t_C and $t_{0.2}$ corresponding to the same parameters as used for producing the results later presented in Section 6.

The final loss event probability for a rear-end loss event is also dependent on the risk of a brake event happening in the first place. Thus, the probability of a consequence is $P(L_k|brake)\,P(brake)$, with the first factors determined as described above. The probability of a brake response by the ADS in our example is the probability of a jaywalker entering the road $P(E_{jaywalker})$ plus the false positive rate ($P(Err^{FP})$) of detecting jaywalkers when there are none. Thus, the final rear-end loss event probability of a specific scene $\{d_0, v_0\} \subset S$ (if a potential jaywalker were to appear) becomes

$$P\left(L_k|,d_0|,v_0\right) = P\left(L_k|brake\right)P\left(E_{jaywalker}\right) + P\left(Err^{FP}\right) \tag{7}$$

In the results presented in this article, the $P(Err^{FP})$ = 10^{-6} h^{-1}. In the bottom half of Figure 6, the schematic steps of deducing the loss event probabilities of rear-end collisions, described in this section, are summarized visually.

One could argue that the factor to account for FNs for jaywalkers (i.e., the case when there is a jaywalker present, but the ADS fails to detect it) should also be included in (7). However, considering that the system will then crash into the jaywalker, and assuming that it would then use full braking, this factor is irrelevant for determining the rear-end loss event probabilities and is thus omitted for the purpose of this case study.

This brake response model is based on data including crashes and near crashes, meaning that the resulting brake onset times (t_B) is likely conservative. Conservative in the sense that they result in worse outcomes than would a model based on more normal driving conditions. However, when the ADS enacts a high brake response to avoid a pedestrian collision, it would indeed present a critical collision risk, warranting the conservative model for brake response. Conversely, a model based on less critical data might render a lower probability of collision. In the shift toward increasingly automated vehicles, this brake response model is nevertheless relevant as it captures the possibly late responses associated with partially automated vehicles and the potentially increased distraction rate of human drivers.

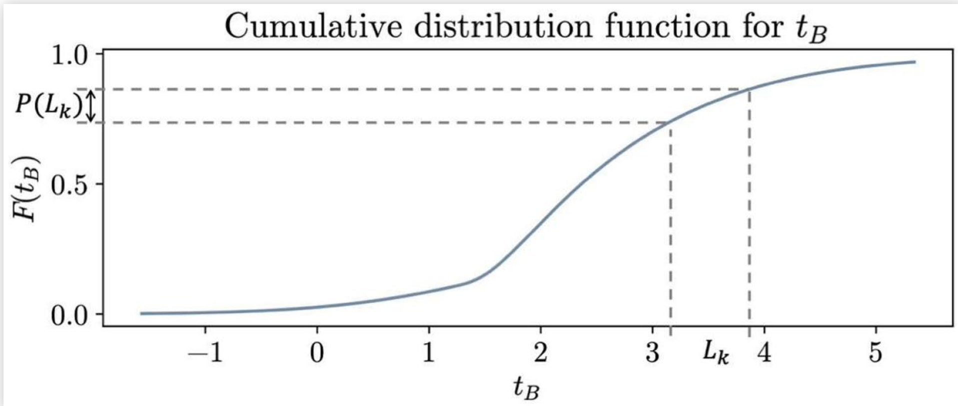
TABLE 1 QRN values for rear-end collisions considered in the case study.

Impact velocity [km/h] (i.e., loss event L_k)	Consequence (ISS) ν_l	Accident rate [h^{-1}]
≤30	Vehicle damage (N/A)	7×10^{-5}
30–47.5	Light injury (1–3)	7×10^{-6}
47.5–55	Moderate injury (4–8)	3×10^{-7}
55–65	Severe injury (>9)	5×10^{-8}
>65	High fatality (N/A)	1.1×10^{-8}

5.3. Conditional Exposure Probabilities—Standstill Vehicle

The example presented above later helps showcase the relationship between the situation around the ADS and the appropriate evasive responses and tactical constraints

FIGURE 4 Illustration of the deduction of the probability of the loss event L_k by determining the t_B:s corresponding to the impact velocities for the loss event and correlating those to the CDF for the brake onset time.



to fulfil the QRN. However, to further exemplify the uses of the proposed QRN assessment framework we also consider whether there is a vehicle standing still at the side of the road. The assumption here is that if there is a standstill vehicle, there is likely a human close by and the risk of that human (now jaywalker) walking out in front of the ADS is significantly higher than the exposure to jaywalkers in a scene without a standstill vehicle.

The probability of a jaywalker walking out on the road in front of the ADS if there is a standstill vehicle present is taken to be $P(E_{jaywalker}|standstill) = 1 \times 10^{-4}$ h^{-1}, the probability of a vehicle being standstill on the roadside is $P(standstill) = 5 \times 10^{-3}$ h^{-1} and the probability of a jaywalker if there is not a vehicle standing still is $P(E_{jaywalker}|standstill) = 1 \times 10^{-8}$ h^{-1}. These numbers represent an engineering judgement by the authors, established for the purpose of showcasing the impact of considering conditional exposure models in the context of this case study, but nevertheless thought to be realistic. That said, such rates can be estimated through, e.g., accident or incident databases, statistics from authorities, or done according to the process described and explored in [10].

When using the exposure rate to jaywalkers, conditional on the presence of a standstill vehicle, it is relevant to also account for the FN rates of detecting the standstill vehicle ($P(FN)$). Thus, $P(FN)P(standstill)P(E_{jaywalker}|standstill)$ needs to be added to $P(E_{jaywalker}|standstill)$, from above. Assuming $P(FN) = 10^{-6}$ h^{-1}, this FN factor will however become negligible ($10^{-6} \times 5 \times 10^{-3} \times 1 \times 10^{-4} = 5 \times 10^{-13}$ h^{-1}) in this example.

If we did not have the ability to discern the presence of a standstill vehicle, we would be stuck with a general probability of jaywalkers for all conditions, given by the law of total probability:

$$P\left(Ejaywalker\right) = P\left(Ejaywalker|standstill\right)P\left(standstill\right) + P\left(E_{jaywalker}|standstill\right)P\left(standstill\right) \tag{8}$$

Using the values from above: $P(E_{jaywalker}) \approx 5 \times 10^{-7}$ h^{-1}.

5.4. QRNs for the Case Study

In the case study, it is assumed that the QRN is posited for each loss event-type (collision with jaywalker or rear-end) separately. Furthermore, the QRN is split into different subcategories to better relate to different consequence classes. Indeed, this split is selected in such a way that the transfer from a loss event is assumed to be completely captured by one single consequence class. While this is a limitation to the fidelity of the modeling, it does not impact the insights garnered from the case study. Moreover, it does also not impact the details of the method for run-time deduction of loss event probabilities, which are anyhow the focus of this article.

Rather than allocating large portions of this article to derive a new QRN for the jaywalker event, we rely on the work previously done in [9], where a detailed QRN is derived and presented. The specific values are reproduced in Table A.1 in the Appendix and will, for the sake of clarity, not be further discussed here.

Following a similar process as the one used in [10], we derive a QRN for the loss event pertaining to rear-end collisions. We leverage the STRADA (Swedish Traffic Accident Data Acquisition) database, statistics, providing the number of rear-end accidents (called "upphinnande olyckor") resulting in a fatality, or different injury levels. Specifically, the non-fatal accidents are estimated by the Police on site of the accident and/or by hospital staff (if they are involved in caring for the victims of the accident) according to the Injury Severity Score (ISS) [66].

According to STRADA there are roughly 5.5 million vehicles in Sweden, assuming that each vehicle operates 300 h a year and that a vehicle operates with a trailing vehicle 50% of the time, this yields the final accident rates presented in Table 1. Before finalizing the QRN, there is a need to link the consequences ν_l to a specific loss event L_k. For that purpose, we take support from [67] (see Figure 2), where the relationship between impact speed and injury probability for rear-end collisions is presented. The specific velocity levels, presented in Table 1, have been placed across the x-axis of Figure 2 of [67], based on engineering judgements by the authors. Finally, we have been unable to find statistics for non-injury accidents separated by accident type. Thus, we assume that the accident rate for the "vehicle damage" category is simply 10 times that of the "light injury" accidents.

The authors recognize that the justification of a QRN can be more complex than what is shown here. However, since the focus of this work is to show how any QRN can be met, we leave such details for future work.

5.5. Deducing Loss Event Probabilities

The task of deducing the loss event probabilities related to each event is equivalent to understanding what proportion of the adverse event, were it to occur, that leads to each loss event. A schematic illustration of the process for loss event deduction for the jaywalker event is presented in the top of Figure 6 and the corresponding process for the rear-end collisions are given in the bottom of the figure. Once we have the process for deducing the loss event probabilities, we can evaluate different tactical decisions, i.e., the cruising velocity v_0, to determine which ones fulfil the QRN.

5.6. Limitations to the Case Study

While we have strived to design a pertinent and realistic case study for the purpose of the evaluation of the proposed framework, it is important to clarify some of the limitations and shortcoming of the proposed case study. First, the case study does not consider different time-of-day T, weather conditions W, or geographic positions G, nor different surface conditions μ. Thus, for the purpose of this example, the scene, S, only involves the presence or absence of a trailing vehicle and its distance (d_0) to the ADS (relating to the factor A of the scene), the presence or absence of a standstill vehicle, as well as the state of the ADS $\vec{x}$. More specifically, we consider the velocity, v_0, as the tactical constraint (i.e., $\vec{x} \coloneqq v_0$), and the evasive response, B, to only include longitudinal actions—in practice corresponding to the brake amplitude applied (i.e., $B \coloneqq a_{ADS}$). Moreover, the case study does not include dynamic changes to the available capabilities of the ADS across different scene factors. However, the results presented in Section 6 explore how different brake amplitudes impact the constraints to the cruising velocity, v_0. Thus, effectively showcasing the relationship between the ADS's braking capabilities and the applicable cruising velocity.

Naturally, limiting the number of conditional factors in the scene simplifies the presentation and discussion of the chosen example. Nevertheless, the analyzed case study can easily be extended, and it is quite straightforward to include more factors within the proposed framework. More precisely, such factors would be included analogously to how the presence or absence of a standstill vehicle is considered in this case study. The choices of the presented case study are the result of thoughtful reasoning that strikes the right balance between realism and the level of discussion complexity in this article, and the rationales are twofold:

1. To retain clarity and conciseness of the article. Adding additional aspects would add complexity without any additional benefits in terms of showcasing the strengths and usefulness of the proposed PCS framework; and
2. The inclusion of additional factors would require access to additional models. For example, the inclusion of more scene factors would imply knowledge of specific exposure levels and interaction models corresponding to each of the explored factors. Considering that we do not have such models available to us at this time, this also contributes to the decisions and design choices made for the case study.

Figure 5 provides an instantiation of the proposed QRN assessment framework in the context of the case study, reflecting the assumptions described above.

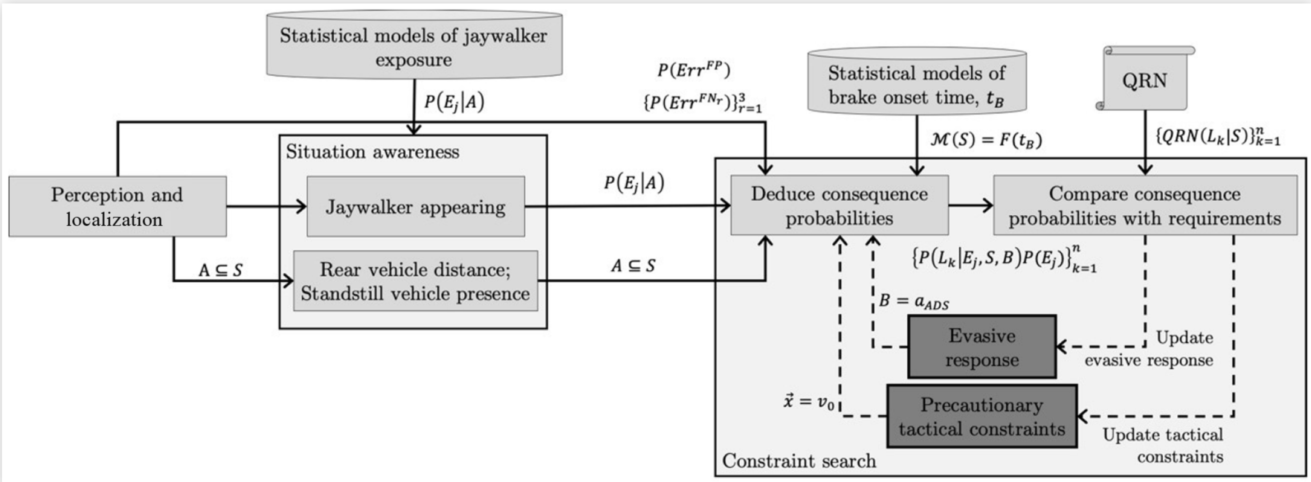
5.7. Baseline for Comparison

To understand the implications of simultaneously considering multiple consequence classes, we also introduce a baseline for benchmarking purposes, which considers an allowed collision rate with jaywalkers of 10^{-7} h^{-1} (corresponding to the severe injury risk accident rate of the QRN, presented in Table A.1) and an allowed rear-end collision rate of 3×10^{-7} h^{-1} (corresponding to moderate injuries of Table 1). The allowed collision rates are selected to provide a reasonable trade-off between performance and safety of the baseline model, as well as to showcase the benefits of using multiple consequence classes as per the QRN.

5.8. Simulation of the Case Study

To explore the case study, we have implemented a simulation in Python where the different models and statistics are realized in view of the proposed QRN assessment framework. In more detail, the implementation follows the steps shown in Figure 6 for each of the two loss event types, respectively. These steps provide an assessment of the QRN fulfilment for a set of tactical constraints (v_0) and evasive responses (a_{ADS}) given specific values for the scene (e.g., d_0 and the presence of a standstill vehicle). To illustrate the relevance of the QRN assessment framework, we also need to close the circle and consider the "constraint search" box of the framework. For that

FIGURE 5 Instantiation of the framework of Figure 2 related to the case study. The adverse event E_j only relates to the appearance of a jaywalking pedestrian. The two dark gray boxes highlight the decision variables of the ADS, namely the brake response, B, and the cruise speed, v_0.



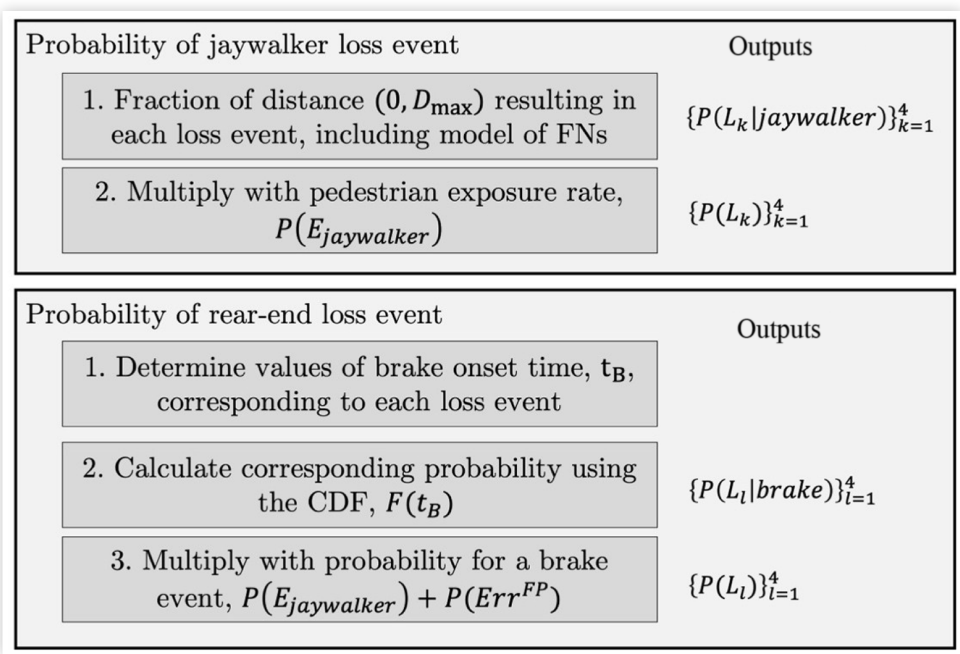
purpose, the maximum velocity (i.e., tactical constraints) of a particular scene fulfilling the QRN is derived for different evasive responses. When finding the optimal value of the velocity (i.e., maximum, conditioned on QRN fulfilment), we use binary search [68] (see §6.2.1, p. 410).

The term "simulation" here is used to mean sampling and numerically solving the needed equations and steps to arrive at the loss event probabilities and the subsequent maximum velocity. Thus, we do not rely on a specific simulation engine to determine movements of the actors and the like, but rather have a minimal set of equations and assumptions, described in the sections above, make that for us.

6. Simulation Results and Discussion

This section presents the results of the simulation-based case study, described in Section 5. First, a comparison

FIGURE 6 Depiction of the steps included in deducing the loss event probabilities related to jaywalker and rear-end loss events, respectively.



between our PCS framework and a collision rate-constrained baseline model is presented.

Second, we have a look at the simulation results of a concrete parametrization of the case study example (i) to explore the tactical constraints that fulfil the QRN for different evasive responses (i.e. braking amplitudes), (ii) to analyze the impact of different false positive rates of the perception system, and (iii) to see how the exposure rate to jaywalkers conditioned on the presence of a standstill vehicle impact the tactical constraints and the selection of evasive responses.

The ability of the ADS to adapt its precautionary constraints, at runtime, based on the implications of the proposed QRN assessment framework is discussed. As is the construction of conditional exposure models. Finally, the topic of completeness is discussed in light of the presented framework, limitations, and avenues for future work are provided.

6.1. Comparison to a Collision-Constrained Baseline

One integral part of the PCS framework is the consideration of multiple consequence classes across the QRN. This aspect is explored in Figure 7. Here, the PCS framework is compared to the baseline described in Section 5.7. Both the baseline and the PCS approach have access to the correct exposure rate of jaywalking pedestrians, as this varies across the x-axis in Figure 7. Note that the collision rates include collisions at all velocities. The collision rates are derived using the driving velocity that is most restrictive, considering both types of accidents (see the bottom panel of Figure 7).

If we first turn to the upper panel of Figure 7, where the rear-end collision rate is depicted, we see that the knowledge of a QRN leaves the PCS approach at a lower overall collision rate for low exposure rates, which also shows up as the lower allowed velocity (blue as well as orange dashed lines) in the lower panel. This trend is also true for collisions with jaywalkers, shown in the middle panel of Figure 7. It is worth noting that as the exposure rate increases, the type of accidents incurred by the baseline and the PCS approach differ. Under the assumptions of the case study, described in Section 5, almost all collisions with jaywalking pedestrians occur at cruising speed, or else the emergency braking can avoid the collision altogether. Thus, the collisions produced by the baseline are not only more abundant but also occur at higher collision velocities due to the relatively higher cruising speed compared to the PCS approach.

As the exposure is increased above $\approx 2 \times 10^{-6}$ h^{-1}, the roles change, and it is instead the baseline that produces an overall lower collision rate. The collisions produced by the PCS approach do, however, occur at lower speeds and continuously fulfil the QRN.

To the end of the x-axis, the rear-end collision rate saturates the baseline's allowed collision rate, rendering a zero allowed velocity. This is perhaps an overly conservative selection for the allowed collision rate, as it corresponds to rather severe accidents (moderate injuries, in fact). However, this collision rate was selected to showcase the nuances that the QRN-aware PCS approach can produce. When only considering collision rates, the results in the baseline produce both a higher risk profile for low exposure rates and a too conservative solution for high exposure rates.

6.2. Considering the Evasive Response

Next, the impact of different evasive responses is explored for the case study example. Here, the jaywalker exposure rate is fixed at $P(E_{jaywalker}) = 1 \times 10^{-8}$ h^{-1}, which renders the maximum velocities for different braking amplitudes (a_{ADS}), depicted in Figure 8. The blue solid line represents the velocity fulfilling the QRN for jaywalker conflicts, and we can see that a larger deceleration (to the left of the figure) results in higher allowed velocities. The yellow dash-dotted line relates to rear-end collisions. Lastly, the black dashed line is the velocity considering the fulfilment of the QRN for both types of loss events.

Following the jaywalker velocity in Figure 8, the increase in braking amplitude across the x-axis only results in an increase of just above 10 km/h. Adding lateral capabilities to detect as well as steer away would significantly improve the odds of the ADS avoiding collisions, especially those cases that occur really close to the ADS, and thus provide the main portion of the high-impact velocity collisions. However, for the purpose of showcasing the PCS framework, the present case study does not consider such maneuvers.

The presence of a trailing vehicle impacts the maximum velocity in the sense that the ADS is unable to enact really high deceleration and is thus limited to a velocity of 65 km/h, as is seen from the horizontal line of Figure 8.

Interestingly, the line corresponding to the rear-end collisions is flat across all evasive responses, which signifies the relative difficulty of mitigating the second last of the loss event classes, i.e., relating to the velocity range 55–65 km/h of Table 1.

6.3. Different False Positive Rates

Since the fulfilment of the QRN for rear-end events is highly dependent on the braking probability, described in (7), lowering such values would allow higher maximum velocities. This is illustrated in Figure 9, where the maximum velocities for different evasive responses are depicted for different values of FP rates of the perception system. Notably, for the top two lines, the braking amplitude does indeed impact the maximum velocity as the ADS is then able to fulfil the QRN for the second last and

FIGURE 7 A comparison between the PCS approach (dashed lines) and a baseline with only one consequence class, i.e., collisions (solid lines). The top panel displays the collision rate for rear-end collisions, the middle panel the collision rate with jaywalking pedestrians, and the lower panel presents the resulting maximum velocities fulfilling the respective requirements for the two approaches across the two considered collision types. The x-axis spans different exposure rates to jaywalking pedestrians. The FP rate for the perception system is set to 10^{-6} h^{-1}, the following distance d_0 = 90 m and the deceleration for both the ADS and the trailing vehicle are set to $a_{ADS} = a_{rv} = -8$ m/s^2. The dotted lines in the top and middle panels correspond to the allowed collision rate of the baseline. Note that all axes are in log-scale, except for the velocity (y-) axis in the bottom panel. Note that the maximum velocity is capped at 105 km/h.

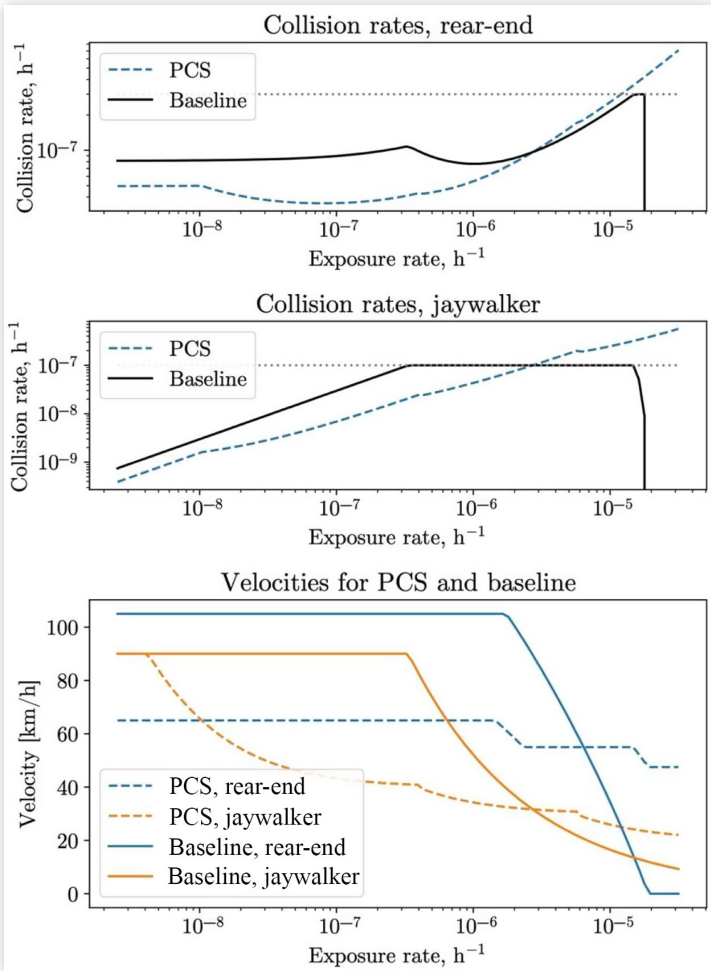


FIGURE 8 The maximum velocity fulfilling the QRN for different evasive responses a_{ADS}. The solid line only considers collision with the jaywalker. The dot-dashed line only considers the consequences of rear-end collisions (following a jaywalker event) and the black dashed line corresponds to the velocity when considering fulfilment of the QRN for both the loss event types. The exposure rate for jaywalker is set to 1 × 10^{-8} h^{-1}, the following distance d_0 = 90 m, the deceleration of the trailing vehicle $a_{rv} = -8$ m/s^2, and the FP rate of the perception system is set to 10^{-6} h^{-1}.

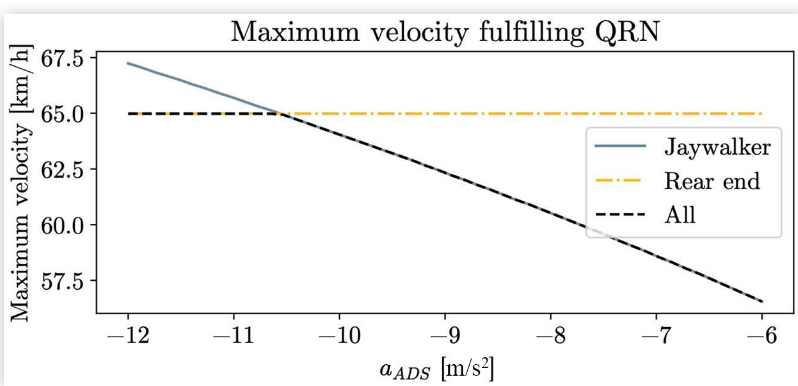
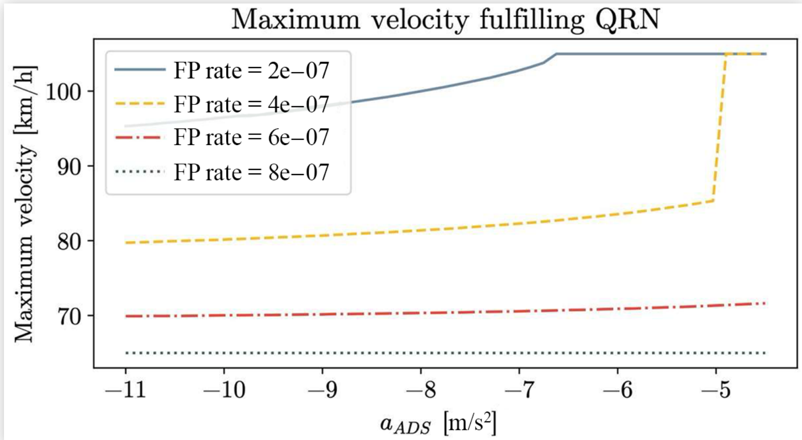


FIGURE 9 The maximum velocities and braking amplitudes, fulfilling the QRN for rear-end collisions when assuming different FP rates for the perception system. The following distance is set to d_0 = 55 m as opposed to 90 m, to get more interesting behavior for lower braking amplitudes. Note that the maximum possible velocity is capped at 105 km/h (hence the plateau of the blue solid line for lower braking amplitudes).



eventually the last consequence class, as the FP rate is reduced. Different FP rates can represent different operating conditions, e.g., night, rain, clear, or day, therefore showcasing how the PCS approach is also able to consider variable capability estimates from the ADS. Such differences could also be present in the false negative rates of the system, but for the sake of brevity, they are not explored in more detail in the presented case study.

6.4. Conditional Exposure Model

We now turn to the case where we have a conditional probability model for the exposure rates to jaywalkers, dependent on the presence of a standstill vehicle. The results for the three possible cases are shown in Figure 10, where the exposure values are taken from Section 5.3. If the ADS is unable to detect a standstill vehicle, we see that the maximum velocity matches that of Figure 8. For the case where there is a standstill vehicle present, the ADS needs to keep its velocity really low at around 15 km/h even when using high braking amplitudes, which is due to the high FP rate of the perception system in combination with the brake delay and the high exposure rate compared to the allowed accident rate of the QRN.

The middle line of Figure 10 corresponds to the case where the ADS does not have any conditional model for jaywalker exposure using the presence of a standstill vehicle. Here, the ADS is restricted to a velocity reduction of more than 30 km/h compared to the case when the ADS has access to the conditional model and detects the absence of a standstill vehicle. It is worth noting that for all three lines in Figure 10 a trailing vehicle is present. However, the rear vehicle only impacts the maximum velocity in the top line (corresponding to the case presented in Figure 8). In the other two cases, where the exposure rate to jaywalkers is higher, the QRN for a jaywalker loss event is instead the limiting factor when deriving the maximum velocity.

Other values for the different conditional probabilities will yield different quantitative results in terms of the resulting maximum velocity, but the qualitative relationship between the three lines will remain the same as depicted in Figure 10.

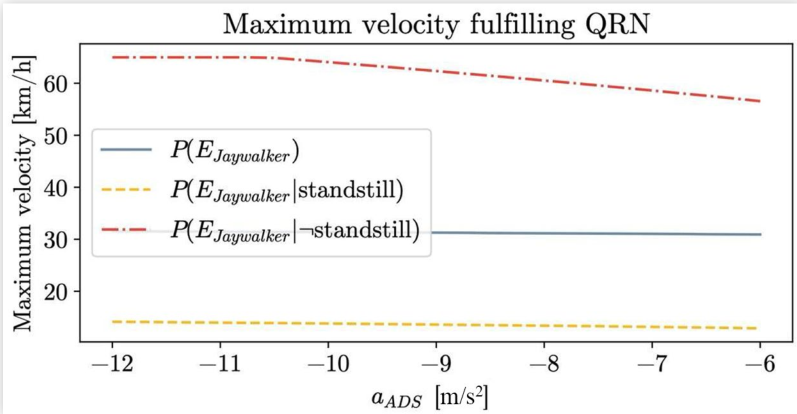
6.5. Run-time Precautionary Tactical Constraints

From the results in Figure 8, we can conclude that the ADS is able to travel slightly faster if preparing a more aggressive evasive response while detecting the *absence* of a trailing vehicle (corresponding to the left side of the figure). Conversely, if a trailing vehicle is detected, it thus augments the scene *S*, the ADS needs to lower its speed in order to fulfil the QRN.

Here, it should be noted that the plots above do not account for the FP rate of detecting the absence of a trailing vehicle. However, considering the relative closeness between the optimal (v_0 = 65 km/h, a_{ADS} = −8 m/s^2) and the applied (v_0 = 67.5 km/h, a_{ADS} = −12 m/s^2) tactical constraints/evasive response couple, any such failures would, in the context of this case study, be negligible for any reasonable FP rate and would, for this case, not impact the QRN fulfilment. In general, that term should, however, be accounted for in the analysis to ensure fulfilment of the QRN of all loss event types.

The application of the results presented here would suggest a pre-analysis of the few possibilities considered in the case study and, as such, provide the ADS with a look-up table of the appropriate tactical constraints and evasive responses considering the present operational scene. Thus, the ADS could carry out the SAW and

FIGURE 10 The maximum velocity for three different exposure levels of jaywalkers. Two corresponding to the conditional probabilities of the presence or absence of a standstill vehicle and one model independent of the standstill vehicle, the exact probabilities presented in Section 5.3. In all three cases, a trailing vehicle is present at a distance d_0 = 90 m, decelerates with a_{rv} = −8 m/s^2, and the FP rate of the perception system is set to 10^{-6} h^{-1}.



subsequently check the appropriate tactical constraints and evasive responses accordingly.

6.5.1. Dynamic Adaptation to Variations in the Scene A relevant implication of the proposed QRN assessment framework is the ability of the ADS to judge if the scene (including its own state, e.g., velocity) is likely to result in acceptable loss event probabilities. If instead the present scene is found to violate the QRN given the provided assessment results (either in the form of a look-up table or through more elaborate methods, as discussed in Section 4.4), the ADS needs to change its state to comply with the predicted QRN fulfilment. An example of such a scene would be if the ADS approaches a standstill vehicle and only detects it a few meters before passing it. Despite the late detection, the ADS should still aim to mitigate the potential accident consequences by reducing its velocity to match that of the QRN assessment (i.e., ≈15 km/h according to Figure 10). Similarly, if the ADS detects a trailing vehicle while driving at 67.5 km/h (having previously failed to detect it), it needs to adapt its velocity and the prepared evasive response.

As such, the proposed QRN assessment framework effectively grants the ADS the ability to dynamically adapt its precautionary tactical constraints to ensure continuous fulfilment of the QRN. Moreover, gathering data on the prevalence of such occasions (i.e., when the QRN is temporarily assessed to be violated due to a false SAW) could further help increase the performance for consecutive releases. Note that the risk incurred in those situations should have already been accounted for in the analysis by considering the error rates of the ADS with respect to detecting such situations. Thus, by letting the ADS adapt to dynamic changes in its SAW, it will reduce the likelihood of the ADS operating under a false impression of its surroundings. Consequently, the impact of such failures in the SAW is thus reduced and, as a result, could lead to an increased performance.

6.5.2. Errors in SAW Related to the Context-Specific QRN If the SAW makes the ADS believe it is operating in the wrong context, the applicable QRN levels might also be different. This is an important factor of residual risk that needs to be adequately analyzed. However, the impact of such errors could be mitigated through appropriate choices of which contexts to distinguish and impose different context-specific QRNs for. First, the contexts should be easily distinguishable by the ADS and preferably have redundant means of making such distinctions. Second, contexts that are prone to being mixed up could be assigned similar QRN levels to minimize the risk of a false context detection.

Furthermore, specific attention should be given toward eliminating systematic errors, since such errors would increase the risk ensuing from operating according to the wrong context-specific QRN levels. In particular, the loss event probabilities are associated with adverse events with comparatively low exposure rates. Thus, if the ADS can detect a false context classification quickly, the risk is almost negligible. However, if there are systematic errors and the ADS is left to operate in "the wrong context" for a prolonged period of time, this might have more dramatic effects on the ADS's ability to fulfil the QRN—and in turn operate safely.

Lastly, while the QRN levels might be different between different contexts, many of the scene factors will remain the same. For example, the presence of a trailing vehicle will impact the ADS in similar ways, irrespective of the context-specific QRN. Thus, assuming that the performance of the ADS is independent of the detected context, the main concern with failures in context detection is matching the deduced loss event probabilities to the wrong context-specific QRN.

6.6. Construction of Conditional Exposure Models

From Figure 10, we can conclude that it is favorable to have a model accounting for the presence of a standstill vehicle when estimating the exposure rate of a jaywalker. At least in the example considered here. Generally, it is reasonable to construct such conditional models for adverse events that restrict the performance of the ADS. However, one cannot just construct such conditional models out of thin air.

First, there is a need for data underpinning the statistics of the model. Using a Bayesian approach to determine the exposure rates, as described in [10], one still needs an additional half order of magnitude of data compared to the true value (having to do with confidence levels of the Poisson model). Thus, to estimate the exposure rate of jaywalkers in the presence of a standstill vehicle, one would need *a lot* of data from standstill vehicles to be able to say that the adverse jaywalker event rate is only 1×10^{-4} h^{-1}. More specifically, one would likely need to pass over 200,000 vehicles standing still without any jaywalker incident to conclude such a number with some confidence. However, to be able to arrive at the number of the non-conditional model (i.e., $P(E_{jaywalker}) = 5 \times 10^{-7}$ h^{-1}), one would need an even larger amount of data ($\approx 10^8$ h of data). What should be noted here is that the proposed framework for QRN assessment allows for the incorporation of such models to improve the performance of the ADS. These models do not need to be present from the beginning of operations of the ADS, but could gradually be constructed as more operational data becomes available.

Second, the construction of conditional models could, of course, be data-driven, but they can also be guided by other models. For example, considering animal exposure, one could model the exposure at different times of day. Then, a biological model for when the animal is commonly asleep or when it is likely to move about would provide valuable input to construct a conditional exposure model.

Exposure models for different times of the day might be particularly relevant if it is found that the capabilities of the ADS also change depending on the time. For example, while the ADS's perception system might be less reliable at night, the exposure rates to jaywalkers might also be less, and thus the use of a conditional model for night in this case might benefit the overall performance of the system, as the exposure rates and failure rates could somewhat counteract each other. However, note that the associated design choice of allowing reduced integrity of the ADS in relation to reduced exposure rates needs to be justified, both by adequate evidence as well as from an ethical and moral standpoint.

6.7. Completeness and Scalability

The framework proposed in this article, and presented in Figures 1 and 2, suggests evaluating the loss event probabilities and comparing these to the level posited in the QRN. But how do we know if we have considered all scenes and adverse events that might lead to a specific loss event? To ensure this completeness, it is crucial to have an exhaustive set of scenes and events when following the assessment framework proposed herein. When using the framework of this article to decide on appropriate tactical constraints and evasive responses, the completeness of the scene(s) is not relevant, as we only ever care about the scene the ADS is currently operating in. However, the adverse events need to be completely enumerated to get an accurate estimate of the loss event probabilities for the scene. Since both types of adverse events (i.e., actor behaviors and hidden risks) pertain to the operational context of the ADS and not the ADS itself, one can introduce a category for "other" events, as suggested in [12] when exhaustively enumerating the applicable loss events for the QRN. While this "other" adverse event might severely limit the initial performance of the ADS, the accumulation of operational data will gradually reduce the exposure rate. Thus, as more data becomes available, the impact of this "other" adverse event will also diminish.

In terms of the scalability of the proposed framework, it revolves around two main aspects: (1) to provide a complete assessment of the loss event probabilities in a timely manner with respect to the development cycle of the ADS and (2) to transfer the knowledge of these loss event probabilities into runtime. The first aspect (1) is, as already discussed in Section 4.2, supported by the well-established methods for scenario-based testing. However, composing generalizable interaction models for such testing and how to efficiently cover all applicable adverse events remain challenging. As for (2), the look-up table approach explored in the presented case study suggests a crude way of achieving this, but more elaborate models are possible and would likely improve the performance of the ADS and enable capturing a larger variety in the SAW. This represents a pertinent avenue for future work.

6.8. Limitations and Future Work

While we have strived to design a pertinent and realistic case study for the purpose of evaluating the proposed framework, the presented case study provides a limited validation of the proposed PCS framework, and further validation should naturally be considered in future work.

First, the case study does not consider adverse actions from actors already present in the scene (i.e., it considers just statistical models for hidden risks, see the top left of Figure 2). The inclusion of such events might also warrant a more refined simulation and sampling system than applied here. Such an approach would effectively merge the proposed QRN assessment framework with works related to scenario-based testing, as already alluded to in Section 4.2 and discussed in Section 6.7. This is our intended next step for future work.

Second, the case study does not include lateral actions of any of the actors. Adding that degree of freedom would be both crucial for realistic risk assessment and central in designing a relevant and performant ADS. The case study was nevertheless built on a one-dimensional example to facilitate clarity and conciseness of the presentation as well as to highlight the implications of the proposed framework and the associated PCS constraints. Adding additional degrees of freedom would significantly obfuscate the discussions of this article with respect to mechanisms of the proposed framework and is hence deferred to future work.

Third, future work should consider more elaborate models of vehicle dynamics, beyond the brake delay incorporated in the case study. Moreover, future evaluations should include a wider variety of scene factors and ODDs, such as, e.g., time-of-day, weather, road surface conditions, and regions with disparate traffic behavior and interaction models.

The three described limitations all pertain to the evaluation as part of the case study. They do not, however, correspond to fundamental limitations of the framework itself. Future work should focus on testing the framework while removing these three limitations from the evaluation used. One bottleneck for achieving this is, of course, the access to detailed models for event and error exposures as well as interaction models. In future work, we intend to address this bottleneck by leveraging large-scale fleet data.

In terms of limitations to the proposed framework of the article, this article does not provide concrete guidance for how to connect context-specific QRN requirements to the consequence-level QRN. This is discussed in Section 4.3, but we do not provide any definitive recommendations. Furthermore, such aspects are also not covered by the presented case study. Providing concrete strategies to enforce applicable exposures to the contexts (i.e., scenes) related to the context-specific QRNs are left for future work. The lack of concrete guidance does not, however, limit the usefulness of the proposed PCS framework in large. That said, further work is warranted to iron out these details pertaining to the use of context-specific QRN requirements within the context of the proposed PCS framework.

7. Conclusions

In this article, we formalize, generalize, and elaborate on the concept of Precautionary Safety (PCS) to supply tactical and operational decisions that can fulfil QRACs spanning multiple accident, injury, and severity levels simultaneously. In particular, the framework presented in Figure 1, and detailed in Figure 2, connects the tactical decisions and intended evasive responses to the fulfilment of the predetermined QRN. The framework itself is enabled by a mathematical formulation, where the probability of each loss event (and by extension each consequence class) is factorized into three different factors: (i) the probability of the scene; (ii) the probability of adverse events given the scene; and (iii) the loss event probability given the adverse events, the scene, and the intended evasive response. A simulation-based case study is conducted to display the utility and applicability of the presented framework. The PCS framework is also compared to a baseline, showcasing the safety and performance benefits of having a QRN-aware approach. Furthermore, the usefulness of conditional exposure models is explored and discussed. Future research should focus on investigating the scalability of this approach with respect to the full operational complexity of the ADS and its operational environment, as well as making sure to cover all applicable loss events.

Acknowledgements

The research has been supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation, and the Swedish Innovation Agency (Vinnova, through the Entice project and the TECoSA [Center for Trustworthy Edge Computing Systems and Applications]).

Declaration of AI-Assisted Technologies

The authors declare that no generative AI technologies were used in the preparation and writing of this article. In doing so, the authors take full responsibility for the authenticity and scientific integrity of this article.

Contact Information

Magnus Gyllenhammar, corresponding author
magnus.gyllenhammar@zenseact.com

References

1. SAE International, “Surface Vehicle Recommended Practice—Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles,” SAE J3016:202104, SAE J3016C, 2021.
2. Mehdipour, N., Althoff, M., Tebbens, R.D., and Belta, C., “Formal Methods to Comply with Rules of the Road in Autonomous Driving: State of the Art and Grand Challenges,” *Automatica* 152 (2023): 110692.

3. Gyllenhammar, M., Rodrigues de Campos, G., and Törngren, M., "The Road to Safe Automated Driving Systems: A Review of Methods Providing Safety Evidence," *IEEE Transactions on Intelligent Transportation Systems* 26, no. 4 (2025): 4315-4345, doi:https://doi.org/10.1109/TITS.2025.3532684.
4. Koopman, P. and Wagner, M., "Autonomous Vehicle Safety: An Interdisciplinary Challenge," *IEEE Intelligent Transportation Systems Magazine* 9, no. 1 (2017): 90-96.
5. Bolbot, V., Theotokatos, G., Bujorianu, L.M., Boulougouris, E. et al., "Vulnerabilities and Safety Assurance Methods in Cyber-Physical Systems: A Comprehensive Review," *Reliab. Eng. Syst. Saf.* 182 (2019): 179-193.
6. Riedmaier, S., Ponn, T., Ludwig, D., Schick, B. et al., "Survey on Scenario-Based Safety Assessment of Automated Vehicles," *IEEE Access* 8 (2020): 87456-87477.
7. Wardziński, A., "Safety Assurance Strategies for Autonomous Vehicles," in Harrison, M.D. and Sujan, M-A., (eds.), *International Conference on Computer Safety, Reliability, and Security (SAFECOMP)* (Berlin, Heidelberg: Springer Berlin Heidelberg, 2008), doi:https://doi.org/10.1007/978-3-540-87698-4_24.
8. Nolte, M., Jatzkowski, I., Ernst, S., and Maurer, M., "Supporting Safe Decision Making Through Holistic System-Level Representations & Monitoring–A Summary and Taxonomy of Self-Representation Concepts for Automated Vehicles," ArXiv:2007.13807, 2020.
9. Rodrigues de Campos, G., Kianfar, R., and Brännström, M., "Precautionary Safety for Autonomous Driving Systems: Adapting Driving Policies to Satisfy Quantitative Risk Norms," in *International Conference on Intelligent Transportation Systems (ITSC)*, Indianapolis, IN, 2021.
10. Gyllenhammar, M., Rodrigues de Campos, G., Sandblom, F., Törngren, M. et al., "Uncertainty Aware Data Driven Precautionary Safety for Automated Driving Systems Considering Perception Failures and Event Exposure," in *Intelligent Vehicles Symposium (IV)*, Aachen, Germany, 2022.
11. ISO, "Road Vehicles—Safety for Automated Driving Systems, Design, Verification and Validation," ISO/CD, ISO/CD 5083:2021, ISO/CD 5083, 2023.
12. Warg, F., Skoglund, M., Thorsén, A., Johansson, R. et al., "The Quantitative Risk Norm—A Proposed Tailoring of HARA for ADS," in *International Conference on Dependable Systems and Networks Workshops (DSN-W)*, Valencia, Spain, 2020.
13. ISO, "Road Vehicles—Functional Safety," ISO 26262:2018, ISO 26262, 2018.
14. ISO, "Road Vehicles—Safety of the Intended Functionality," ISO/PAS 21448:2019, ISO/PAS 21448, 2019.
15. BSI, "Operational Design Domain (ODD) Taxonomy for an Automated Driving System (ADS)—Specification," PAS 1883:2020, 2020.
16. Czarnecki, K., "Operational World Model Ontology for Automated Driving Systems–Part 1: Road Structure," Waterloo Intelligent Systems Engineering Lab (WISE) Report, 2018.
17. Czarnecki, K., "Operational World Model Ontology for Automated Driving Systems–Part 2: Road Users, Animals, Other Obstacles, and Environmental Conditions," Waterloo Intelligent Systems Engineering Lab (WISE) Report, University of Waterloo, 2018.
18. Gyllenhammar, M., Johansson, R., Warg, F., Chen, D. et al., "Towards an Operational Design Domain That Supports the Safety Argumentation of an Automated Driving System," in *European Congress on Embedded Real Time Systems*, Toulouse, France, 2020.
19. Ulbrich, S., Menzel, T., Reschka, A., Schuldt, F. et al., "Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Gran Canaria, Spain, 982-988, 2015.
20. Batkovic, I., Rosolia, U., Zanon, M., and Falcone, P., "A Robust Scenario MPC Approach for Uncertain Multi-Modal Obstacles," *IEEE Control Syst. Lett.* 5, no. 3 (2020): 947-952.
21. Koschi, M. and Althoff, M., "Set-Based Prediction of Traffic Participants Considering Occlusions and Traffic Rules," *IEEE Transactions on Intelligent Vehicles* 6, no. 2 (2020): 249-265.
22. Pek, C., Manzinger, S., Koschi, M., and Althoff, M., "Using Online Verification to Prevent Autonomous Vehicles from Causing Accidents," *Nat. Mach. Intell.* 2, no. 9 (2020): 518-528, doi:https://doi.org/10.1038/s42256-020-0225-y.
23. van der Ploeg, C., Nyberg, T., Sánchez, J.M.G., Silvas, E. et al., "Overcoming Fear of the Unknown: Occlusion-Aware Model-Predictive Planning for Automated Vehicles Using Risk Fields," *IEEE Transactions on Intelligent Transportation Systems* 25, no. 9 (2024): 12591-12604.
24. Gharavi, L., Dabiri, A., Verkuijlen, J., De Schutter, B. et al., "Proactive Emergency Collision Avoidance for Automated Driving in Highway Scenarios," *IEEE Transactions on Control Systems Technology* 33, no. 4 (2024): 1164-1177.
25. Brudigam, T., Olbrich, M., Wollherr, D., and Leibold, M., "Stochastic Model Predictive Control with a Safety Guarantee for Automated Driving," *IEEE Transactions on Intelligent Vehicles* 8, no. 1 (2023): 22-36, doi:https://doi.org/10.1109/TIV.2021.3074645.
26. Wang, Q., Zhou, Q., Lin, M., and Nie, B., "Human Injury-Based Safety Decision of Automated Vehicles," *IScience* 25, no. 8 (2022), doi:https://doi.org/10.1016/j.isci.2022.104703.
27. Parseh, M., Asplund, F., Svensson, L., Sinz, W. et al., "A Data-Driven Method Towards Minimizing Collision Severity for Highly Automated Vehicles," *IEEE Transactions on Intelligent Vehicles* 6, no. 4: 723-735.
28. Intelligent Transportation Systems Committee, "IEEE Standard for Assumptions in Safety-Related Models for Automated Driving Systems," IEEE Std 2846-2022, 2022, doi:https://doi.org/10.1109/IEEESTD.2022.9761121.

29. Shalev-Shwartz, S., Shammah, S., and Shashua, A., "On a Formal Model of Safe and Scalable Self-Driving Cars," ArXiv:1708.06374, 2017.
30. Nistér, D., Lee, H.-L., Ng, J., and Wang, Y., "The Safety Force Field," in *NVIDIA White Paper*, 2019, https://developer.download.nvidia.com/driveworks/secure/docs/DRIVE_8.0_Release_Docs/the-safety-force-field.pdf?__token__=exp=1783324528~hmac=62179a33f7cbf0f5ecc95d14a132730e7a362466829f53d06eed555c0de2de6e.
31. Censi, A., Slutsky, K., Wongpiromsarn, T., Yershov, D. et al., "Liability, Ethics, and Culture-Aware Behavior Specification Using Rulebooks," in *International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada, 2019.
32. Arechiga, N., "Specifying Safety of Autonomous Vehicles in Signal Temporal Logic," in *Intelligent Vehicles Symposium (IV)*, Paris, France, 2019.
33. Nilsson, J., Ödblom, A.C.E., and Fredriksson, J., "Worst-Case Analysis of Automotive Collision Avoidance Systems," *IEEE Trans. Veh. Technol.* 65, no. 4 (2015): 1899-1911.
34. Sandblom, F., de Campos, G.R., Harda, P., Warg, F. et al., "Choosing Risk Acceptance Criteria for Safe Automated Driving," in *International Workshop on Critical Automotive Applications: Robustness & Safety (CARS)*, Brussels, Belgium, 2024.
35. Junietz, P., Steininger, U., and Winner, H., "Macroscopic Safety Requirements for Highly Automated Driving," *Transp. Res. Rec.* 2673, no. 3 (2019): 1-10.
36. Stellet, J.E., Müller, B., and Ebel, S., "Derivation of Quantitative Risk Acceptance Criteria for Automated Driving Systems," in *15th Workshop Fahrerassistenz Und Automatisiertes Fahren*, Kloster Bonlanden, Berkheim, 139-151, 2023.
37. Koopman, P. and Widen, W., "Redefining Safety for Autonomous Vehicles," in *International Conference on Computer Safety, Reliability, and Security*, Florence, Italy, 300-314, 2024.
38. Schöner, H.-P. and Antona-Makoshi, J., "Testing for Tactical Safety of Autonomous Vehicles," in *30th Aachen Colloquium Sustainable Mobility*, Aachen, Germany, 2021.
39. Johansson, R. and Nilsson, J., "Disarming the Trolley Problem-Why Self-Driving Cars Do Not Need to Choose Whom to Kill," in *Workshop CARS 2016-Critical Automotive Applications: Robustness & Safety*, Gothenburg, Sweden, 2016.
40. Reich, J. and Trapp, M., "SINADRA: Towards a Framework for Assurable Situation-Aware Dynamic Risk Assessment of Autonomous Vehicles," in *European Dependable Computing Conference (EDCC)*, Munich, Germany, 2020.
41. Reich, J., Wellstein, M., Sorokos, I., Oboril, F., and Scholl, K.-U., "Towards a Software Component to Perform Situation-Aware Dynamic Risk Assessment for Autonomous Vehicles," in *European Dependable Computing Conference (EDCC)*, Munich, Germany, 2021.
42. Feth, P., "Dynamic Behavior Risk Assessment for Autonomous Systems," Ph.D. dissertation, Fraunhofer Verlag, 2020.
43. Eggert, J., "Risk Estimation for Driving Support and Behavior Planning in Intelligent Vehicles," *Automatisierungstechnik* 66, no. 2 (2018): 119-131.
44. Schreier, M., Willert, V., and Adamy, J., "An Integrated Approach to Maneuver-Based Trajectory Prediction and Criticality Assessment in Arbitrary Road Environments," *IEEE Transactions on Intelligent Transportation Systems* 17, no. 10 (2016): 2751-2766.
45. Gyllenhammar, M. and Sivencrona, H., Path planning in autonomous driving environments. EU Patent EP3971526B1, 2022.
46. Gyllenhammar, M. and Sivencrona, H., Risk estimation in autonomous driving environments. US Patent US12319320B2, 2022.
47. Gyllenhammar, M. and Sivencrona, H., Scenario identification in autonomous driving environments. US Patent US12084051B2, 2022.
48. Trapp, M., Schneider, D., and Weiss, G., "Towards Safety-Awareness and Dynamic Safety Management," in *European Dependable Computing Conference*, Iasi, Romania, 2018.
49. Trapp, M. and Weiss, G., "Towards Dynamic Safety Management for Autonomous Systems," in *Safety-Critical Systems Symposium*, Bristol, UK, 2019, ISBN:9781729361764.
50. Khastgir, S., Sivencrona, H., Dhadyalla, G., Billing, P. et al., "Introducing ASIL Inspired Dynamic Tactical Safety Decision Framework for Automated Vehicles," in *Intelligent Transportation Systems Conference (ITSC)*, Yokohama, Japan, 2017.
51. Sulaman, S.M., Beer, A., Felderer, M., and Höst, M., "Comparison of the FMEA and STPA Safety Analysis Methods–A Case Study," *Software Quality Journal* 27, no. 1 (2019): 349-387.
52. James Elizebeth, M., Khastgir, S., Babaev, I., Chen, S. et al., "Comparison of FTA and Stpa Approaches: A Brake-by-Wire Case Study," *SSRN Electronic Journal* (2023), doi:https://doi.org/10.2139/ssrn.4394251.
53. Gyllenhammar, M., Campos, G.R.D., and Törngren, M., "A Safety Argument Fragment Towards Safe Deployment of Performant Automated Driving Systems," in: Törngren, M., Gallina, B., Schoitsch, E., Troubitsyna, E. et al., (eds.), *Computer Safety, Reliability, and Security. SAFECOMP 2025 Workshops. SAFECOMP 2025*, Lecture Notes in Computer Science, vol. 15955 (Cham: Springer, 2026), doi:https://doi.org/10.1007/978-3-032-02018-5_15.
54. Resnik, D.B. and Andrews, S.L., "A Precautionary Approach to Autonomous Vehicles," *AI and Ethics* 4, no. 2 (2024): 403-418, doi:https://doi.org/10.1007/s43681-023-00277-6.
55. Zhao, D., Lam, H., Peng, H., Bao, S. et al., "Accelerated Evaluation of Automated Vehicles Safety in Lane-Change Scenarios Based on Importance Sampling Techniques," *IEEE Transactions on Intelligent Transportation Systems* 18, no. 3 (2016): 595-607.

56. O'kelly, M., Sinha, A., Namkoong, H., Duchi, J. et al., "Scalable End-to-End Autonomous Vehicle Testing via Rare-event Simulation," arXiv:1811.00145, 2018, https://doi.org/10.48550/arXiv.1811.00145
57. Zhang, S., Peng, H., Zhao, D., and Tseng, H.E., "Accelerated Evaluation of Autonomous Vehicles in the Lane Change Scenario Based on Subset Simulation Technique," in *International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, 2018.
58. Gyllenhammar, M., Zandén, C., Vakilzadeh, M.K., Falkovén, A. et al., Estimation of probability of collision with increasing severity level for autonomous vehicles. EU Patent EP3940487A1, 2022.
59. Gyllenhammar, M., Åsljung, D., Vakilzadeh, M.K., and de Campos, G.R., "Efficient Injury Risk Assessment for Automated Driving Systems Using Subset Simulation," in: Gallina, B., Törngren, M., and Bitsch, F., (eds.), *Computer Safety, Reliability, and Security. SAFECOMP 2025*, Lecture Notes in Computer Science, vol. 15954 (Cham: Springer, 2026), doi:https://doi.org/10.1007/978-3-032-01241-8_6.
60. Garikapati, D., Liu, Y., and Huo, Z., "Systematic Selective Limits Application Using Decision-Making Engines to Enhance Safety in Highly Automated Vehicles," *SAE Int. J. CAV* 8, no. 1 (2024): 49-70, doi:https://doi.org/10.4271/12-08-01-0005.
61. Gyllenhammar, M., Zandén, C., and Vakilzadeh, M.K., In-vehicle system for estimation of risk exposure for an autonomous vehicle. EU Patent EP4219262A1, 2022.
62. Gyllenhammar, M., Zandén, C., and Vakilzadeh, M.K., Estimation of risk exposure for autonomous vehicles. EU Patent App. EP4220581A1, European Patent Office, 2022.
63. Brännström, M., Coelingh, E., and Sjöberg, J., "Model-Based Threat Assessment for Avoiding Arbitrary Vehicle Collisions," *IEEE Transactions on Intelligent Transportation Systems* 11, no. 3 (2010): 658-669, doi:https://doi.org/10.1109/TITS.2010.2048314.
64. Markkula, G., Engström, J., Lodin, J., Bärgman, J. et al., "A Farewell to Brake Reaction Times? Kinematics-Dependent Brake Response in Naturalistic Rear-End Emergencies," *Accid. Anal. Prev.* 95 (2016): 209-226.
65. Shin, S.-G., Ahn, D.-R., Baek, Y.-S., and Lee, H.-K., "Adaptive AEB Control Strategy for Collision Avoidance Including Rear Vehicles," in *Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, 2872-2878, 2019.
66. Baker, S.P., O'Neill, B., Haddon, W. Jr., and Long, W.B., "The injury Severity Score: A Method for Describing Patients with Multiple Injuries and Evaluating Emergency Care," *Journal of Trauma and Acute Care Surgery* 14, no. 3 (1974): 187-196.
67. Jurewicz, C., Sobhani, A., Woolley, J., Dutschke, J. et al., "Exploration of Vehicle Impact Speed–Injury Severity Relationships for Application in Safer Road Design," *Transportation Research Procedia* 14 (2016): 4247-4256.
68. Knuth, D.E., *The Art of Programming*, 2nd ed. Vol. 3 (Reading, Massachusetts, USA: Addison Wesley, 1998).
69. Peden, M.M., "World Report on Road Traffic Injury Prevention," World Health Organization, 2004.

Appendix: QRN Values for Jaywalker Collisions

TABLE A.1 QRN values for jaywalker collisions, reproduced from Rodrigues de Campos et al. [9], based on [69].

Impact velocity [km/h] (i.e., Loss event L_k)	Consequence, ν_l	Accident rate [h^{-1}]
≤10	Minor accident	1×10^{-5}
10–20	Light injury risk	1×10^{-6}
20–30	Severe injury risk	1×10^{-7}
30–40	Low fatality risk	1×10^{-8}
>40	High fatality risk	1×10^{-9}