CHALMERS
UNIVERSITY OF TECHNOLOGY

When is it Safe to Complete an Overtaking Maneuver? Modeling Drivers' Decision to Return After Passing a Cyclist

Downloaded from: https://research.chalmers.se, 2024-12-20 08:40 UTC

Citation for the original published paper (version of record):
Rasch, A., Flannagan, C., Dozza, M. (2024). When is it Safe to Complete an Overtaking Maneuver? Modeling Drivers' Decision to Return After
Passing a Cyclist. IEEE Transactions on Intelligent Transportation Systems, 25(11): 15587-15599.
http://dx.doi.org/10.1109/TITS.2024.3454768

N.B. When citing this work, cite the original published paper.

© 2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, or reuse of any copyrighted component of this work in other works.

(article starts on next page)

When Is It Safe to Complete an Overtaking Maneuver? Modeling Drivers' Decision to Return After Passing a Cyclist

Alexander Rasch, Carol Flannagan, and Marco Dozza

***Abstract*— For cyclists, being overtaken represents a safety risk of possibly being side-swiped or cut in by overtaking drivers. For drivers, such maneuvers are challenging–not only do they need to decide when to initiate the maneuver, but they also need to time their return well to complete the maneuver. In the presence of oncoming traffic, the problem of completing an overtaking maneuver extends to balancing head-on with side-swipe collision risks. Active safety systems such as blind-spot or forward-collision warning systems, or, more recently, automated driving features, may assist drivers in avoiding such collisions and completing the maneuver successfully. However, such systems must interact carefully with the driver and prevent false-positive alerts that reduce the driver's trust in the system. In this study, we developed a driver-behavior model of the drivers' return onset in cyclist-overtaking maneuvers that could improve such a safety system. To provide cumulative evidence about driver behavior, we used data from two different sources: test track and naturalistic driving. We developed Bayesian survival models for the two datasets that can predict the probability of a driver returning, given time-varying inputs about the current situation. We evaluated the models in an in-sample and out-of-sample evaluation. Both models showed that drivers use the displacement of the cyclist to time their return decision, which is accelerated if an oncoming vehicle is present and close. We discuss how the models could be integrated into an active-safety system to improve driver acceptance.**

***Index Terms*— Cyclist safety, overtaking, ADAS, driver model, Bayesian model, survival model.**

Received 4 October 2022; revised 2 June 2023, 31 January 2024, and 16 May 2024; accepted 29 August 2024. Date of publication 4 October 2024; date of current version 1 November 2024. This work was supported in part by the Project Modelling Interaction between Cyclists and Automobiles (MICA2) funded jointly by Vinnova, the Swedish Energy Agency, the Swedish Transport Administration, and the Swedish Vehicle Industry, through the Strategic Vehicle Research and Innovation (FFI) Program under Grant 2019-03082; and in part by the Project Modelling Interaction between Cyclists and Automobiles (MICA) collaboration with the Project Drivers in Interaction with Vulnerable Road Users (DIV) funded by Toyota Motor Europe, Autoliv, and Veoneer. The Associate Editor for this article was Y. Zhang. *(Corresponding author: Alexander Rasch.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Local Ethical Review Authority in Gothenburg, Sweden, under Reference No. Dn:600-17.

Alexander Rasch and Marco Dozza are with the Department of Mechanics and Maritime Sciences, Chalmers University of Technology, 417 56 Gothenburg, Sweden (e-mail: alexander.rasch@chalmers.se; marco.dozza@chalmers.se).

Carol Flannagan is with the University of Michigan Transportation Research Institute, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: cacf@umich.edu).

Digital Object Identifier 10.1109/TITS.2024.3454768

I. Introduction

A. Car-Cyclist Overtaking Maneuvers

CYCLING continuously increases on a global scale due to its benefits in terms of health, environment, commuting, and leisure activities [1]. The COVID-19 pandemic accelerated this increase as social distancing and lockdowns promoted choosing cycling over public transport [2], [3]. However, as the number of cyclists increases, so does the number of interactions and possible conflicts with motorized vehicles [4]. Such interactions are most critical in locations without an infrastructure that separates cyclists from motorized vehicles [5], [6], which is the case for up to 86% of all cyclist travel globally [7]. The absence of proper cyclist infrastructure is particularly evident on rural roads where cyclist travel may be less frequent than in urban areas, however, does occur for commute, leisure, or exercise purposes [8], [9], [10].

Cyclist-overtaking scenarios represent a particularly complex form of interaction between drivers, cyclists, and possibly oncoming traffic. Such maneuvers are generally divided into four phases [11]: 1) *approaching* phase, in which the driver has to decide on whether to overtake the cyclist (referred to as a *flying* maneuver if done without significant speed decrease, and possibly despite the presence of an oncoming vehicle, making it the more risky strategy [12]) or not (referred to as *accelerative* if the driver decreases speed to let oncoming traffic pass first, and then re-accelerates to pass the cyclist), 2) *steering-away* phase, in which the driver maneuvers the vehicle towards the adjacent lane to achieve enough lateral clearance to the cyclist, 3) *passing* phase, in which the vehicle passes the cyclist, and 4) *returning* phase, in which the driver steers the vehicle back into the original lane position. Each of these phases comes with different crash risks [13]. The passing and returning phases are particularly challenging as the risk of a head-on collision with oncoming traffic (highest at the start of the returning phase) needs to be balanced with the risk of side-swiping the cyclist.

B. Side-Swipe Versus Head-On: the Risks After Committing to the Maneuver

To successfully overtake a cyclist, a driver must make not only a safe and timely decision about whether to overtake the cyclist or not, but also time well the *moment of return* back to the original lane, after having passed the cyclist. Rear-end

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

crashes during the approaching phase, for instance, due to the driver failing to recognize the cyclist, account for the most severe injuries and fatalities due to high impact speeds, which is of particular concern on rural roads [8], [14]. However, recent studies have pointed out that the later phases of the maneuver frequently account for side-swipe crashes with the cyclist [15], [16]. Gildea et al. [16] also pointed out that many such collisions, especially with lower-severity injury outcomes, are underreported in crash statistics.

Much research has been done on analyzing close passes by drivers and their contributing factors [17], such as the presence and timing of oncoming vehicles [11], [13], as well as infrastructural elements such as parked cars and on-road painted cycling lanes [18]. However, previous work has not investigated when early returns happen and threaten the cyclist's safety.

In the passing and returning phase, the driver must balance the two collision threats arising from side-swiping the cyclist and heading on the possibly present oncoming traffic. A too-close passing or a too-early return may, even without direct contact, destabilize the cyclist and cause it to fall [19]. On the other hand, a too-late return may result in a head-on collision with an oncoming vehicle. Monitoring both collision threats with the cyclist and the oncoming vehicle at the same time is challenging, as the cyclist may be hard to locate accurately for the driver after the passing moment, and the time-to-collision of the oncoming vehicle may be hard to judge [20].

C. Active-Safety Systems for Cyclist-Collision Avoidance

Vehicular active-safety systems, such as collision-warning systems, can assist drivers in preventing collisions in overtaking maneuvers [21]. During the passing phase, a blind-spot detection (BSD) system, for instance, can warn the driver, and potentially even the cyclist [16], [22], of an impending side-swipe collision by monitoring the blind spots at the sides of the vehicle [23]. At the same time, a forward collision warning (FCW) system may help the driver avoid a head-on collision with the oncoming traffic [24].

However, these systems need to be carefully tuned not to become a nuisance to the driver [25], [26], [27]. Such nuisance can arise from *technical* and *perceived* false-positive activations. While sensor errors can cause technical false-positive activations, perceived ones are caused whenever the driver feels that the activation was unnecessary [28]. Triggering perceived false-positive activations may put the driver's acceptance of the system at risk and may result in reduced trust in the system or even its deactivation [29], [30], which in turn eliminates its safety benefit [26].

D. Driver Models to Improve Active-Safety Systems

Models of driver behavior may help reduce perceived false-positive activations by active-safety systems. By allowing to predict the driver's behavior, the system can recognize early if the driver does not behave according to the prediction and warn or intervene accordingly [31], [32]. By doing so, warnings and interventions may be triggered before the kinematic threat of a collision becomes high, resulting in increased safety while still being accepted. In other words, driver models may serve as a reference [33] for the machine to factor in the usual driver behavior in the threat assessment.

Ljung Aust and Dombrovski [27] suggested that warnings and interventions that happen *outside* of a driver's *comfort zone*, i.e., the region of all possible states in which the driver does not feel any discomfort, may be more acceptable than those that happen inside of it. This, in return, results in an increased safety benefit, as more collisions can be avoided while ensuring the driver's acceptance of the system activation [32]. Therefore, modeling the driver's comfort-zone behavior can be understood as a key enabler for effective active safety.

Previous research has made use of different types of driver models for different driver tasks, from cognitive models, e.g., for driver reactions to warnings [34] or take-over requests [35], to machine-learning-based models for car-following scenarios [36]. Other machine-learning-based work has aimed at modeling interactions with vulnerable road users, e.g., using recurrent neural networks for pedestrian-trajectory estimation [37], [38], or, more recently, transformer-based architectures applied to pedestrian-crossing scenarios [39].

Probabilistic driver models in particular have gained attention, as they not only allow for predicting the driver's decision, but also its uncertainty [25]. A probabilistic model can predict the probability of a driver's action, allowing system designers to choose activation-threshold probabilities that optimize the system's true or false positive rate [32].

Just like in the earlier phases of overtaking maneuvers, a probabilistic driver model may enhance an active-safety system by allowing better-timed and yet acceptable activations, even in the later phases of an overtaking maneuver [16], [32]. However, such work has been missing to date, and this study aims at filling this gap. At the same time, *automated-driving* systems may benefit from such a model by predicting how a human driver would behave in the scenario and steer the vehicle accordingly since human-like driving may be preferred by passengers [12], [40].

E. Contribution

This study developed a computational, probabilistic model of driver behavior that can help understand and predict how drivers time their return onset after having passed a cyclist. We fitted Bayesian discrete-time survival models to relate a set of relevant factors for drivers' behavior to their return onset. We used two different datasets that complement each other and discuss their differences. Furthermore, we discuss how the model can support active-safety systems and automated vehicles in making well-timed decisions on when to alert drivers to best compromise the risk of a side-swipe collision with the cyclist and a head-on collision with the oncoming traffic.

II. Methods

A. Datasets

We obtained two datasets collected in two different environments: 1) data from a test-track (TT) experiment (Fig. 1 a)

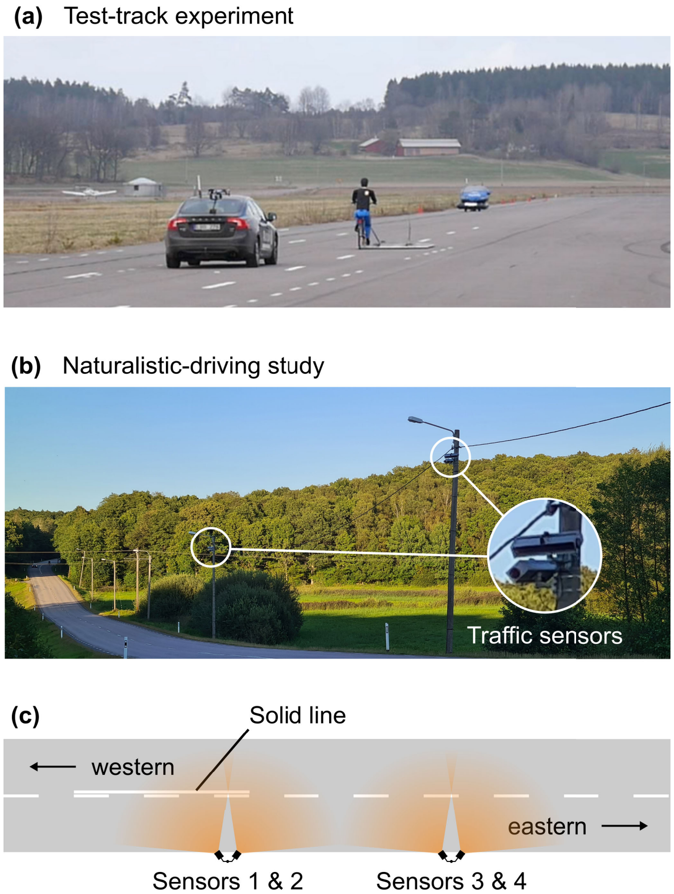


Fig. 1. Photos of the data-collection environments used in this study. Panel a shows the setup of the test-track experiment (for a video, see https://youtu.be/AixQ189hMi4). Panels b and c show the naturalistic-driving study and the locations of the four traffic sensors (for a video, see https://youtu.be/uLjw1yHNjwQ).

and 2) data from a naturalistic-driving (ND) study conducted on a rural road (Fig. 1 b, c). The two datasets capture the same scenario in different environments. By fitting the same model structure on both datasets and comparing the results, we leveraged the datasets to achieve cumulative evidence to support our results.

1) Test-Track Data: The TT data were collected in March, 2018, on V årg årda airfield, V årg årda, Sweden [13]. The data used for analyses include 18 drivers who participated in the experiment, out of which five were female. All participants were employees of Autoliv or Veoneer; however, they were not themselves involved in the development or design of safety systems. The participants were, on average, 42.9 (SD = 8.9) years old, possessed a driver's license for 24.7 (SD = 9.1) years, drove 12 (SD = 6) times per week, and 14 944 (SD = 10 205) km per year.

In the experiment, participants were instructed to keep a speed of 70 km/h and could overtake a robot cyclist traveling at 20 km/h (Fig. 1 a). The lateral position of the cyclist was controlled to be in one of two conditions: 1) cycling close to the lane edge with no overlap between ego vehicle and cyclist, 2) cycling in the center of the lane with full overlap. At the same time, an oncoming balloon vehicle was driving in the opposite direction at 40 km/h (its technical limit), and timed to meet the driver at a short and a long time gap, corresponding to 7 s and 10 s time-to-collision (TTC; measured when the ego vehicle reached 2 s TTC behind the cyclist). In the case of no oncoming traffic, the vehicle was standing still in the opposite lane (due to technical constraints), but far enough away to minimize its effect on the driver's behavior. The cyclist and the oncoming vehicle were controlled via a CHRONOS server [41]. The lane width of the recreated two-lane road was 3.75 m. The dimensions of the road users (length x width, assuming rectangular bounding boxes) were 4.63 $\times$ 2.10 m for the ego vehicle, 1.87 $\times$ 0.50 m for the cyclist, and 3.60 $\times$ 1.80 m for the oncoming vehicle. The experiment was approved by the local ethical review authority in Gothenburg, Sweden (Dn:600-17). Further details of the experiment can be found in the descriptive study reported by Rasch et al. [13].

High-precision differential GPS sensors installed on the road users recorded their position, speed, and heading. In post-processing, all signals were synchronized at a 100 Hz sampling rate to the GPS reference time. In total, 104 overtaking maneuvers were recorded.

2) Naturalistic-Driving Data: The ND data were collected on seven consecutive days in August and September 2021, on the two-lane rural road Spårhagavägen, Mölndal, Sweden. The considered road stretch was straight and about 150 m long with a lane width of about 3.60 m and a speed limit of 70 km/h (Fig. 1 b). Towards the western end of the stretch, a solid line disallowed crossing into the opposite lane for vehicles traveling in the western direction (Fig. 1 c).

Four stereovision-based smart traffic sensors OTUS3D, provided and set up by Viscando AB, Sweden,[1] installed on two light poles, continuously detected, classified and tracked road-user objects passing the road stretch (Fig. 1 b, c), delivering time-series data at a sampling rate of 25 Hz for road user type, position, speed, heading angle, and the dimensions (assuming rectangular bounding boxes) of all road users. Further details of the experiment can be found in the descriptive study reported by Rasch et al. [42].

A search algorithm detected overtaking maneuvers when 1) a passenger car and a cyclist were detected in the same time span, 2) traveling in the same direction, and 3) the car passed the cyclist.

The event selection included the following conditions:

- The overtaking vehicle was a passenger car.
- Only one cyclist was overtaken.
- The oncoming vehicle was correctly captured.
- The complete passing phase of the driver was captured.
- The overtaking driver did not "squeeze" between the cyclist and the oncoming vehicle.

Since overtaking maneuvers could occur at any location during the observed road stretch, inherently, not all phases of the overtaking maneuvers were captured. To identify whether an oncoming vehicle was present during the passing phase, we compared the extrapolated distance of the oncoming vehicle (by assuming a constant speed at the speed limit) from the ego vehicle to the sight distance available to the

[1] https://viscando.com/

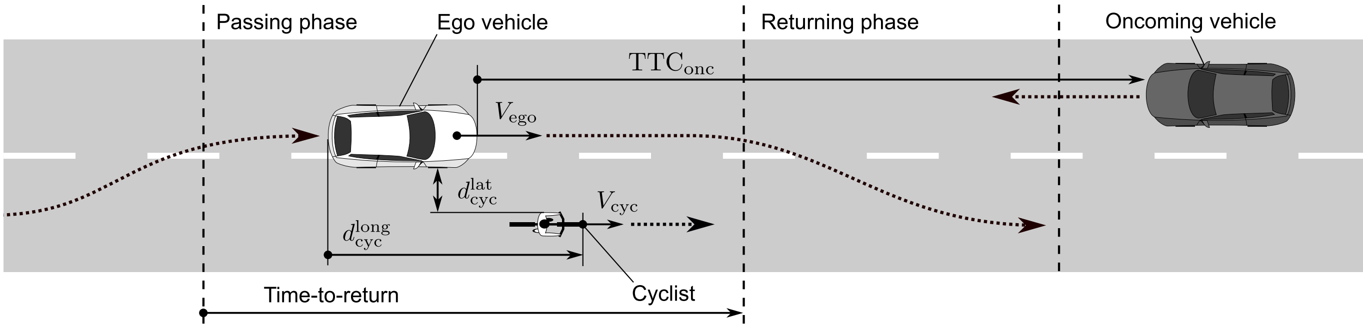


Fig. 2. Scenario of a flying overtaking maneuver with the explanatory variables for the model marked. In the figure, $d_{\mathrm{cyc}}^{\mathrm{long}}$ denotes the longitudinal displacement between the rear-end of the ego vehicle and the front end of the cyclist, $d_{\mathrm{cyc}}^{\mathrm{lat}}$ the lateral distance, $\mathrm{TTC}_{\mathrm{onc}}$ the TTC of the oncoming vehicle, and V_{ego} and V_{cyc} are the speeds of ego vehicle and cyclist, respectively.

driver. The classification and identification of the involved road users were manually verified from the anonymized videos recorded along with the trajectory data. Four false-positive events were found after manual review, a rate of 6.2%. These events were excluded from the analysis. Another seven events (10.9%) were excluded because the overtaking vehicle was not a passenger car but a transporter. These steps resulted in a dataset consisting of 127 events.

Because one goal of the analysis was to compare return onset in ND and TT data for the same type of maneuver, we made an effort to select cases from the ND sample that were comparable to those from the test track. First, only maneuvers with complete data from the passing phase (as defined in II-B.1), in which a passenger car overtook a single cyclist were included, resulting in 82 events. These events were initially divided into flying and accelerative maneuvers (as in [12]), as well as piggybacking maneuvers where the driver simply followed a lead vehicle, copying its behavior (as in [11]). However, we noted that there were five flying maneuvers in which the overtaking vehicle passed both an oncoming vehicle and the cyclist during the same passing phase. These “squeezing” events, which do not clearly fall into either the flying or accelerative categories have previously been identified by [43], [44], [45].

B. Model Variables

Fig. 2 illustrates a flying overtaking scenario of a cyclist in the presence of an oncoming vehicle. The depicted scenario shows the passing and returning phase of the maneuver [11], which follow after the driver has steered away from the collision path with the cyclist. Variables that were extracted for analysis are labeled in the figure and described in more detail in the next two subsections.

1) Response Variable (Return Onset): For each dataset, we defined the start of the passing phase at the moment in time when the ego vehicle reached 0.20 m before its lateral distance from the cyclist reached its maximum. The start of the returning phase, i.e., the return onset, was accordingly defined as the moment in time when the lateral distance from the cyclist fell back by 0.20 m from the maximum. This definition is in accordance with Rasch et al. [13]. It should be noted that this definition allowed for the passing phase to end before the driver reached the cyclist, which in turn allowed us to understand if and why such behavior happens. This definition is different from other definitions of the passing phase in previous research (e.g., Dozza et al. [11], and Kovaceva et al. [46]) that defined the passing phase with a static, spatial zone around the cyclist.

TABLE I
OVERVIEW OF THE EXPLANATORY VARIABLES USED IN THE MODELS

Acronym	Explanatory variable	Type	Definition
$d_{\mathrm{cyc}}^{\mathrm{long}}$	Longitudinal displacement ego-cyclist	Continuous	Longitudinal displacement ego-vehicle rear to cyclist front
$d_{\mathrm{cyc}}^{\mathrm{lat}}$	Lateral distance ego-cyclist	Continuous	Lateral distance ego-vehicle right side to cyclist left side
$V_{\mathrm{ego,cyc}}^{\mathrm{rel}}$	Relative speed ego-cyclist	Continuous	Relative speed at the beginning of the passing phase
OP	Presence of oncoming vehicle	Binary (0 = absent, 1 = present)	An oncoming vehicle is visible to the driver in the passing phase
$\mathrm{TTC}_{\mathrm{onc}}$	Time-to-collision to oncoming vehicle	Continuous	The front-to-front TTC of the oncoming vehicle that is present during the passing phase

2) Explanatory Variables: We included a set of explanatory variables in the models that relate to certain perceptual inputs that we hypothesized to influence the driver’s decision to initiate the returning phase (Fig. 2). These variables were inspired by previous studies on overtaking duration [10], [47] and other safety metrics during overtaking [13]. The four variables, which are summarized in Table I, include: 1) longitudinal displacement to the cyclist, related to the risk of side-swiping in a too-early return; 2) lateral distance to the cyclist, which contributes to the risk of side-swiping due to too narrow passing [48]; 3) presence of an oncoming vehicle, which represents a head-on collision risk to a vehicle that has not returned yet; and 4) relative speed between vehicle and cyclist at the start of the maneuver.

We extracted these variables from both datasets according to the same definition. For the TT data, we used the GPS data to estimate the variables, while for the ND data, we used the filtered data estimated from the traffic sensors’ cameras.

We extracted the longitudinal displacement from the ego vehicle's rear end to the cyclist's front end ($d_{\text{cyc}}^{\text{long}}$) (see Fig. 2 for illustration). The displacement is negative as long as the ego vehicle's rear end is behind the cyclist's front end and positive after that. The lateral distance to the cyclist was measured between the right side of the ego vehicle and the left side of the cyclist's bounding box ($d_{\text{cyc}}^{\text{lat}}$). Relative speed between the ego vehicle and the cyclist was measured at the onset of the passing phase ($V_{\text{ego,cyc}}^{\text{rel}} = V_{\text{ego}} - V_{\text{cyc}}$; Fig. 2). The oncoming vehicle was marked as present or absent (OP $= 0$ if absent, OP $= 1$ if present). For those cases in which it was present, TTC_{onc} measures the TTC between the front end of the ego vehicle and the front end of the oncoming vehicle, calculated as the ratio between the front-to-front distance and relative speed.

C. Bayesian Discrete-Time Survival Models

From both datasets, we selected only flying maneuvers for modeling, in which the driver overtook the cyclist without a significant decrease in speed [12], [28]. This was done to achieve comparable results between TT and ND data and to limit the scope of the models to the most critical situations in which an oncoming vehicle may have been present when passing the cyclist.

For both datasets, we fitted survival models to relate the explanatory variables to the response variable. Survival models are a common type of time-to-event model, more commonly used in medical applications to understand and predict how long an individual survives under different, possibly time-dependent, conditions and treatments. Survival models enable the prediction of events in a probabilistic manner while allowing the inclusion of time-dependent variables [49]. In the transportation field, survival models have been applied to problems concerning the duration of, for instance, overtaking maneuvers [47], [50].

The fundamental elements of a survival model are the hazard $h(t)$ and the survival function $S(t)$. While the hazard function expresses the instantaneous rate of event occurrences at a given time t, the survival function expresses the probability that the event of interest has not occurred yet before time t [49].

Discrete-time survival models represent an easy-to-use framework for modeling problems that concern the occurrence of an event, as in our case. In contrast to continuous-time survival models, their discrete-time counterparts treat the hazard as constant in each time interval. The hazard at the time t_i is defined as the conditional probability of the event at the time t_i, given no event occurred beforehand:

$$h(t_i) = \Pr(\mathrm{T} = t_i | \mathrm{T} \geq t_i). \tag{1}$$

The survival function is defined as:

$$\begin{aligned} S(t_i) &= \Pr(\mathrm{T} \geq t_i) \\ &= (1 - h(t_1)) \cdot (1 - h(t_2)) \cdots (1 - h(t_{i-1})) \\ &= \prod_{j=1}^{i-1} (1 - h(t_j)). \end{aligned} \tag{2}$$

As expressed in Eq. 2, the survival function expresses the probability of survival up to time t_i, which requires survival in all previous time intervals. Discretizing time in this way avoids increased model complexity that arises from fitting baseline hazards to the underlying time itself or assumptions about the proportionality of hazards [51]. By treating hazard as a probability, a discrete-time survival model can be fitted as a simple logistic regression where the outcome is the occurrence of an event: zero if the event did not occur within a time interval, and one if it did [52], [53].

In a Bayesian setting, a discrete-time survival model can be expressed as a mixed-effects model through a Bernoulli regression with a logit link function [53]:

$$\begin{aligned} \text{event}_i &\sim \text{Bernoulli}(p_i) \\ \text{logit}\, p_i &= \log \frac{p_i}{1 - p_i} = \text{logit}\, h(t_i) = \mathbf{X}_i \beta + \mathbf{Z}_i \mathbf{u}, \\ \mathbf{u} &\sim N(\mathbf{0}, \Sigma). \end{aligned} \tag{3}$$

In Eq. 3, the *population-level effects* (also often referred to as fixed effects) are represented by $\mathbf{X}_i$ with corresponding parameter vector β, and the *group-level effects* (often referred to as random effects) are represented by $\mathbf{Z}_i$ with the group-specific parameters included in the vector $\mathbf{u}$. $\mathbf{Z}_i$ is composed of zeros and ones to map the i-th observation to the corresponding group. The group-level parameters are usually modeled as samples from a zero-centered multivariate normal distribution with unknown covariance Σ [54]. The response variable event_i denotes whether the return onset happened ($\text{event}_i = 1$) during the i-th time period, or not ($\text{event}_i = 0$). p_i denotes the probability of the event happening during the i-th period, which is equal to the hazard h_i.

In this work, we used Bayesian discrete-time survival models to understand and predict the decision of a driver to return after having initiated the passing phase. This problem maps well to survival analysis since there is one event of interest (the return onset), and its onset can be predicted as a probability. For the reporting of the Bayesian models, we follow the guidelines by Kruschke [55].

$$\begin{aligned} \text{event}_i &\sim \text{Bernoulli}(p_i) \\ \text{logit}\, p_i &= \beta_0 + \beta_1 d_{\text{cyc},i}^{\text{long}} + \beta_2 d_{\text{cyc},i}^{\text{lat}} + \beta_3 V_{\text{ego,cyc},i}^{\text{rel}} \\ &\quad + \beta_4 \text{OP}_i + \beta_5 \text{OP}_i \text{TTC}_{\text{onc},i} + \mathbf{Z}_i \mathbf{u}_{\text{ID}}, \\ \mathbf{u}_{\text{ID}} &\sim N(\mathbf{0}, \sigma_{\text{ID}} I), \end{aligned} \tag{4}$$

where the subscript i denotes the time period (0.01 and 0.04 s long intervals for the TT and ND data, respectively).

The fitted parameters for the variables are β_0 for the intercept and β_1, β_2, β_3, β_4 and β_5 for the other population-level variables (Eq. 4). The group-level effect $u_{\text{ID},i}$ was only included in the TT model due to the repetitions by each participant and represents the influence of an individual participant with identifier ID in sample i on the intercept of the model. In the end, the effect of the driver on the intercept is characterized by the standard deviation σ_{ID} of a zero-centered normal distribution. For the ND model, we did not include this effect since we assumed that all overtaking maneuvers were performed by different drivers.

In this work, we investigated the full posterior distribution of the fitted parameters for inference purposes [55]. Inference on which parameters have a clear effect may help guide future modeling efforts and set requirements for active-safety systems as to which signals are necessary for the model to function. We identified parameters with a clear influence where the 95% highest density interval (HDI) did not include zero [56]. Furthermore, we quantified the probability of direction of the parameters' posterior distributions, defined as the proportion of the distribution that has the same sign as the distribution's median. This index quantifies the probability of a parameter being either positive or negative, yielding a value between 50% and 100% [56].

To assess whether the addition of the group-level effect for individual drivers (Eq. 4) improved the model, we used approximate leave-one-out cross-validation (LOOCV) [57], to compare the model with vs without the group-level effect. We identified the better model from a clear increase in expected log predictive density (ELPD), greater than the estimated standard error of the increase, implying a greater predictive accuracy. To understand how much individual drivers differ in their behavior, we assessed the proportion of variance explained by the group-level effect driver ID through the intraclass-correlation coefficient (ICC) for the TT model using the following equation [58]:

$$\mathrm{ICC}_{\mathrm{TT}} = \frac{\sigma_{\mathrm{ID}}^2}{\sigma_{\mathrm{ID}}^2 + \pi^2/3}. \tag{5}$$

The models were fitted in R version 4.3.2 (2023-10-31) with the brms package, version 2.20.4 [54]. We used weakly-informative prior distributions for all parameters in both models [54]. For both models, we ran the No-U-Turn (NUTS) Markov chain Monte-Carlo (MCMC) sampling algorithm to fit the parameters, with four chains, each of which ran for 2000 iterations and included a warmup period of 1000 iterations that were discarded afterward. From observing Rhat values close to 1.00 and visual inspections of the trace plots, we ensured that the MCMC chains had converged.

D. Model Assumptions

Singer and Willett [52] mention three assumptions that discrete-time survival models rely on: 1) the *linear additivity* assumption, 2) the *proportionality* assumption, and 3) the *no unobserved heterogeneity* assumption. We actively checked the first assumption by splitting it into a linearity and an additivity assumption. For testing linearity, we tried whether replacing continuous, linear explanatory variables with categorical, non-linear variables improved the model through LOOCV. To achieve convergence, we were only able to split continuous variables into binary categories, but this should capture notable non-linearity for assumption-checking purposes. Those models, however, did not appear to be an improvement over the reported models (the ΔELPD magnitude was within its standard error). We tested the additivity assumption by checking different interaction terms between pairs of explanatory variables in both models (one for TT and one for ND data). We verified through LOOCV that none of the alternative models with interactions clearly outperformed the presented model under the constraint of having the same model structure for both TT and ND data.

The second, proportionality assumption, is that the effect of covariates on the hazard is constant over time. We tested this assumption by including interactions of the model's variables with the time indicator [52]; none of such model variants could converge with the MCMC algorithm.

Finally, the third assumption of no unobserved heterogeneity (or missing covariates) is very challenging to test in this context (see, e.g., an extensive discussion in [59]). In this case, we do not have additional covariates measured; therefore, we can only declare this as an assumption and rely on the robustness of Bayesian (mixed-effects) models.

E. Model Evaluation

Survival models are usually evaluated based on 1) *discrimination*, i.e., their ability to distinguish individuals with low and high risk of experiencing the event of interest, and 2) *calibration*, which assesses how well its predictions agree with the observations in the data [60]. We used the receiver operating characteristic (ROC) curve and the accompanying area under the ROC curve (AUC) to quantify the discriminative performance of our models. We used Hosmer-Lemeshow-style plots to assess the calibration of the models [61], [62]. This method sorts the predicted hazard from each time step into deciles and plots the mean of each decile against the mean of the corresponding observed hazards. We used the root mean square error (RMSE) between predicted and observed hazards over all deciles to quantify the calibration performance of the models.

For both discrimination and calibration of the models, we employed an *in-sample* evaluation and an *out-of-sample* evaluation. We consider the full-sample model to be the model of interest and the model from which we make predictions and inferences (based on the coefficients, for example). In-sample statistics apply to that model and are typical of performance metrics for modeling in inferential contexts. To ensure that the model performance does not result from overfitting, we also conducted out-of-sample validations using 10-fold cross-validation, while splitting the data according to the individual overtaking events. The similarity between these two metrics indicates that overfitting is not substantial. If in-sample performance is much better than out-of-sample performance, overfitting is suspected.

III. Results

A. Data Summary

The TT and ND data consisted of 104 (77 flying and 27 accelerative) and 82 (48 flying, three accelerative, 26 piggybacking, and five squeezing) maneuvers, respectively. For modeling, we only used the flying maneuvers from each dataset. Table II shows an overview of the values of the variables at the return onset in the datasets used for modeling, and Fig. 3 the corresponding distributions. Both datasets contained similar passing-phase durations (Fig. 3 a), however, the durations in the TT data were, on average, slightly shorter

TABLE II

SUMMARY OF THE DATA CONTAINED IN THE TWO DATASETS, INCLUDING THE EXPLANATORY VARIABLES USED FOR MODELING. ALL DATA ARE MEASURED AT THE RETURN ONSET. ALL CONTINUOUS VARIABLES ARE SUMMARIZED AS MEAN (STANDARD DEVIATION); ALL CATEGORICAL VARIABLES AS NUMBER OF SAMPLES PER LEVEL (PERCENTAGE). SQUARE BRACKETS INDICATE THE RANGE OF VALUES ([MIN, MAX])

Characteristic	Test track, N = 77	Naturalistic driving, N = 48
Passing-phase duration (s)	1.90 (0.83) [1.04, 6.08]	2.56 (0.99) [0.80, 5.44]
Long. displacement of cyclist (m)	6.3 (6.1) [-7.1, 23.8]	5.4 (12.3) [-19.6, 50.0]
Lat. distance to cyclist (m)	2.0 (0.5) [1.0, 3.4]	1.8 (0.5) [0.7, 2.8]
Relative speed to cyclist (km/h)	44.1 (8.1) [18.6, 59.2]	48.4 (15.3) [16.1, 96.9]
Presence of oncoming vehicle (-)		
Absent	36 (47%)	30 (62%)
Present	41 (53%)	18 (38%)
Time-to-collision to oncoming (s)	5.82 (1.26) [2.56, 7.77]	3.30 (1.70) [0.19, 6.29]

than in the ND data (Table II). In both TT and ND data, some drivers started to return before having passed the cyclist, indicated by the negative tails of the distributions of the longitudinal displacement of the cyclist (Fig. 3 b). The lateral distances recorded in the TT data were, on average, slightly shorter than those recorded in the ND data (Fig. 3 c, Table II). The TT data contained more events without an oncoming vehicle (Tab. II). Both datasets had similar values of the relative speed between ego vehicle and cyclist (Fig. 3 d).

B. Model Parameters

Fig. 4 shows the posterior distributions of the model coefficients. For both models, the longitudinal displacement of the cyclist had a clear positive effect on the hazard (Fig. 4 a), i.e., with an increased displacement, the probability of the return event increases. For the TT model, the magnitude of the parameter was clearly larger than that of the ND model. The lateral distance did not have a clear effect in the ND model (Fig. 4 b); however, for the TT model, the probability of the parameter being negative was 92.6%. None of the models revealed a clear effect of the relative speed between ego vehicle and cyclist at the start of the passing phase (Fig. 4 c). However, for both models, the presence of the oncoming vehicle had a positive effect on the hazard (Fig. 4 d; 88.6% for the TT model and 100.0% for the ND model), resulting in earlier returns when an oncoming vehicle is present. The TTC of the oncoming vehicle did not have a clear effect for the TT model (57.8%), but it did for the ND model (99.7%), resulting in earlier returns for shorter TTCs (Fig. 4 e). The LOOCV revealed that the TT model with the group-level effect for individual drivers was clearly better than the model without it. The ELPD increased by 22.1 with a standard error of 7.9. The calculated ICC coefficient showed that 45.1% of the variance could be explained by the effect of individual drivers on the model's intercept. Table III summarizes the numerical values of the coefficient distributions.

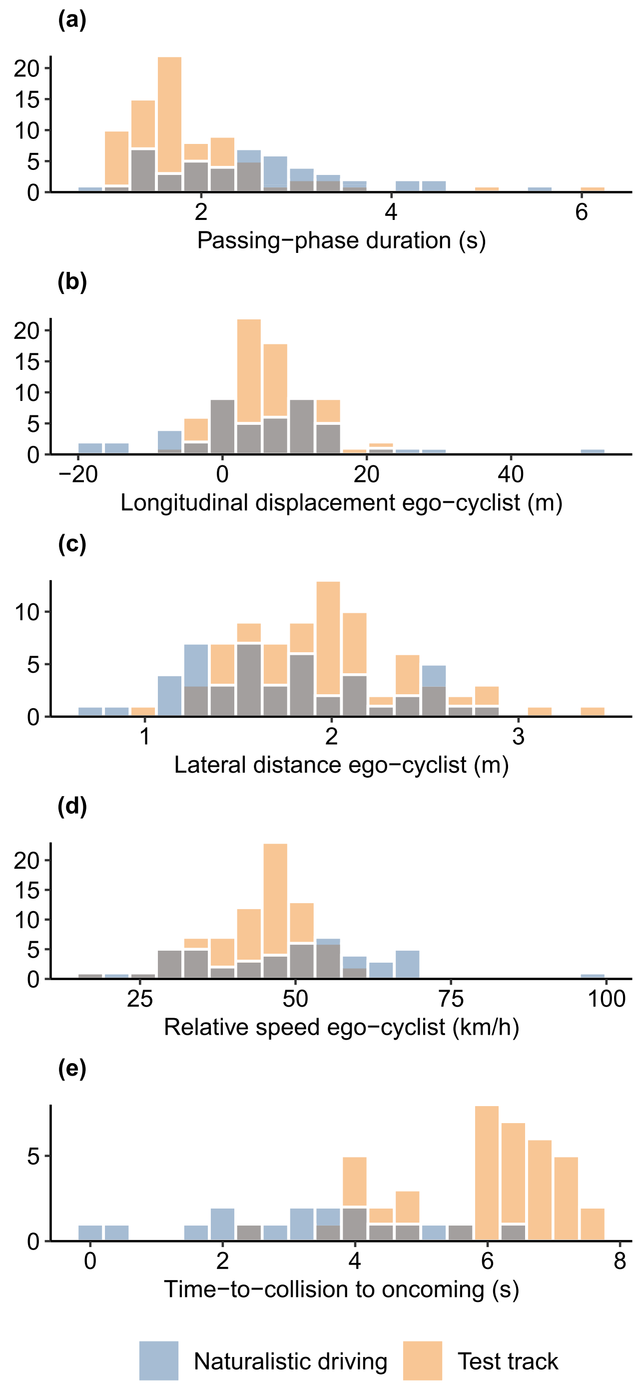


Fig. 3. Distributions of the passing-phase duration and the continuous, explanatory variables. All data are measured at the return onset.

C. Model Evaluation

Fig. 5 b shows the ROC curves for the in-sample evaluation of the discriminative ability of the models. The in-sample AUC of the TT and ND models were 0.872 and 0.837, respectively. For the out-of-sample evaluation, the 10-fold cross-validation of the TT model resulted in an average AUC of 0.887 [min 0.838, max 0.985], and for the ND model in

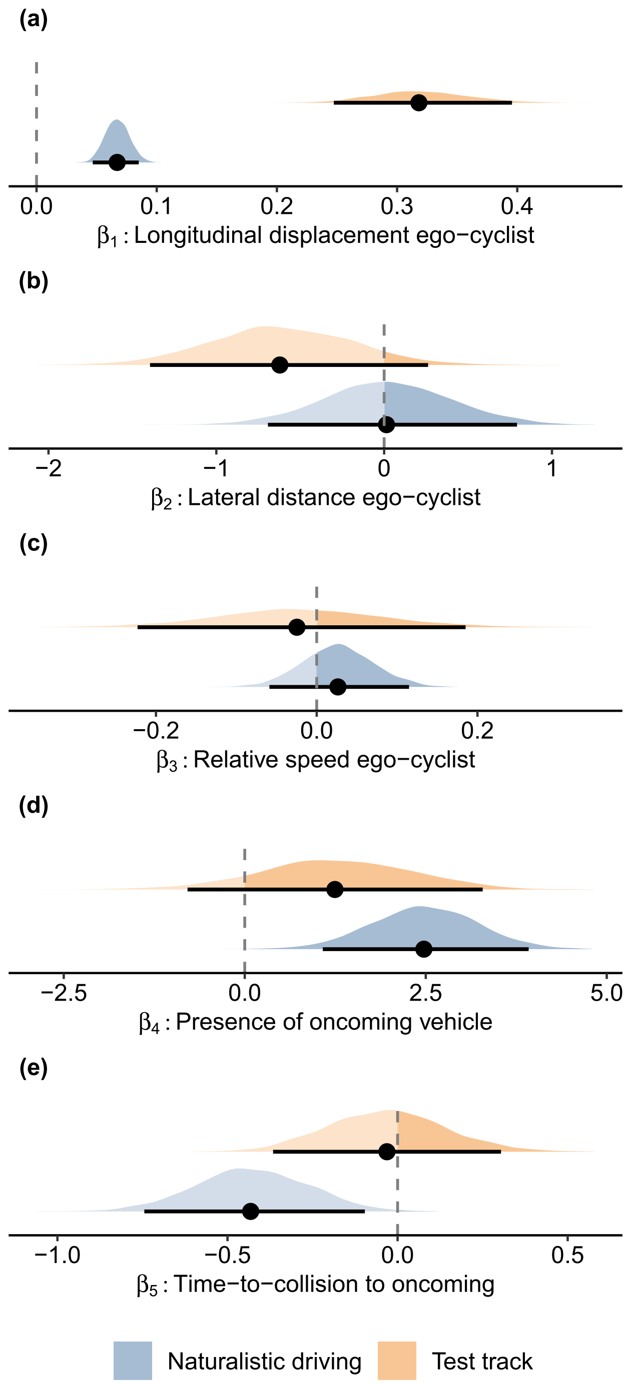


Fig. 4. Model coefficients with full posterior distribution and median (black dot) with 95% highest density interval (HDI, black horizontal bar) marked. To highlight the probability of direction, the positive part of the distributions has an increased opacity.

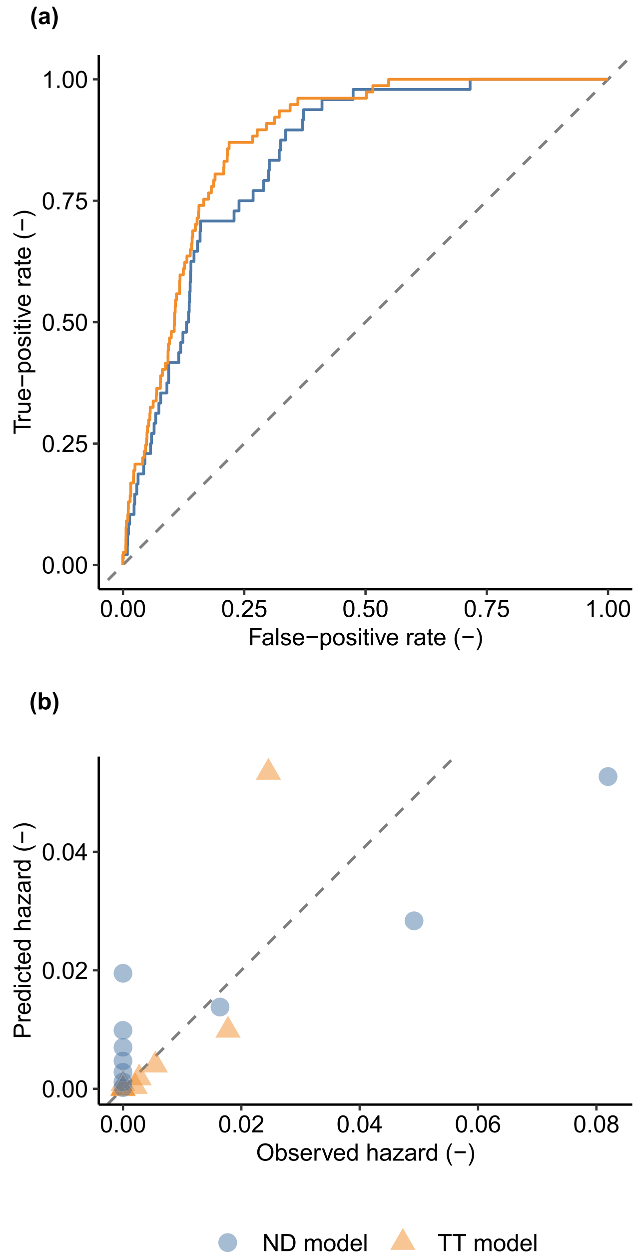


Fig. 5. Evaluation of the return-onset models fitted on test-track (TT) and naturalistic driving (ND) data. Panel a shows the receiver operating characteristic (ROC) curves for assessing the discriminative performance of the model. Panel b shows the Hosmer-Lemeshow calibration plots, i.e., the ten deciles of the hazards predicted by the models and the corresponding observed hazards.

0.841 [min 0.729, max 0.913]. Fig. 5 b shows the Hosmer-Lemeshow calibration plots for the in-sample evaluation of the models. Both models predict similar hazards to the observed ones, particularly for lower deciles, while the higher deciles of the predictions differed the most from the observations. The TT model calibration resulted in an RMSE of 0.009, while the ND model's RMSE was 0.014. The 10-fold cross-validation for the TT model resulted in an average RMSE of 0.018 [min 0.005, max 0.055], and for the ND model in 0.035 [min 0.009, max 0.148].

IV. Discussion

A. Modeling Approach

This work presents a novel application of Bayesian discrete-time survival analysis to return-onset modeling for drivers passing cyclists. Survival analysis maps well to the prediction of outcome onset time, but treating it as a discrete-time model simplifies the modeling process and simplifies the use of

TABLE III
MODEL PARAMETERS SUMMARIZED BY MEDIAN AND LOWER AND UPPER 95% HIGHEST DENSITY INTERVAL (HDI). PD DENOTES THE PROBABILITY OF DIRECTION, DEFINED AS THE PROPORTION OF THE PARAMETER POSTERIOR DISTRIBUTION THAT HAS THE SAME SIGN AS THE DISTRIBUTION MEDIAN

	Test-track model				Naturalistic-driving model			
Parameter	Median	Lower 95% HDI	Upper 95% HDI	PD (%)	Median	Lower 95% HDI	Upper 95% HDI	PD (%)
β_0	-4.97	-8.10	-1.89	100.0	-4.70	-6.38	-2.99	100.0
β_1	0.32	0.25	0.40	100.0	0.07	0.05	0.09	100.0
β_2	-0.62	-1.39	0.26	92.6	0.01	-0.69	0.79	51.5
β_3	-0.02	-0.22	0.19	58.9	0.03	-0.06	0.11	72.6
β_4	1.25	-0.79	3.29	88.6	2.47	1.08	3.92	100.0
β_5	-0.03	-0.37	0.30	57.8	-0.43	-0.74	-0.10	99.7
σ_{ID}	1.60	0.92	2.47	100.0	NA	NA	NA	NA

time-dependent covariates, which are critical in a changing traffic scenario. Further, Bayesian models have a number of advantages over maximum-likelihood models. First, they characterize all of the uncertainty in the model, rather than focusing only on the maximum likelihood model itself. In this case, we emphasize inferences about the predictive value of our parameters; such inferences are made on the basis of the full posterior distribution. Second, they can fit complex models with many parameters and are relatively unaffected by "nuisance parameters" [63]. This is useful because we are interested in inferences about the parameters but can eliminate reliance on p-values associated with predictors [64]. Finally, Bayesian models incorporate priors, which in this case were weakly informative, introducing enough shrinkage to prevent spuriously large coefficients. Future implementations of this model could explore the use of stronger priors and/or the use of these results to build priors for other datasets.

B. Driver Behavior in the Passing Phase

Our results suggest that drivers consider the longitudinal displacement of the cyclist to time their return after having passed the cyclist. If the driver is behind the cyclist, i.e., when the displacement is negative, the probability of returning decreases, while it increases once the driver has passed the cyclist. Both the ND and the TT model revealed that the presence of an oncoming vehicle increases the likelihood of returning. The ND model further showed that drivers returned earlier when the oncoming vehicle was closer. This indicates that drivers compensate for the risk of a head-on collision with the oncoming vehicle by increasing the risk of a side-swipe collision with the cyclist, which is in line with previous research on the effect of oncoming traffic [13], [65].

The fact that the TT model hinted at a possible negative effect of lateral distance on the hazard to return may have been due to the two distinct scenarios tested in the TT experiment (cycling in the middle of the lane vs. at the side of the lane). In the ND data, where the effect was not clear at all, cyclists were observed to usually cycle near the lane edge or even outside of the lane, confirming previous research [32]. Future work should try to investigate whether and how this factor may influence driver behavior.

Additionally, the relative speed of the cyclist appeared to have no influence on drivers in either dataset. This finding suggests that drivers may focus merely on the displacement of the cyclist irrespective of speed. From the cyclist's perspective, we hypothesize the absence of clear longitudinal interaction with the overtaking driver, as the cyclist cannot properly estimate its distance when the vehicle is approaching from behind and may not have enough time to react and regulate its speed once the vehicle has passed. Contrary to crossing scenarios, where cyclists may regulate their speed depending on the dynamics of the approaching vehicles, in our overtaking scenario, we assume the rider to be only concerned about controlling the speed to keep the bicycle stable [66]. In other words, we assume the driver was the main actor controlling the interaction between the vehicle and the cyclist. Although intuitively, one may think that the cyclist would adjust speed and lateral displacement depending on the proximity of the overtaking vehicle, we could not verify this observation in our ND data [46].

Both models performed well in the in-sample and out-of-sample evaluation, predicting hazards that were similar to the observed ones. The ICC from the TT data indicated that almost half (45.1%) of the variance is attributed to individual differences in timing. In spite of this challenge and the fact that the two datasets were collected in fundamentally different environments, they yielded similar results in terms of effect existence and direction. This similarity supports the cumulative evidence of this study, which one dataset alone could not have provided. The TT model complements the results from the ND model by showing the importance of accounting for individual drivers. Moreover, the TT approach allowed exploration of more variation in lateral distance to the cyclist, indicating that if ND data with more variation were collected, they would likely show a stronger effect of lateral distance.

C. Application in Active Safety and Automated Driving

The models could be used in an active-safety system, such as a warning system for overtaking maneuvers. Specifically, in the later phases of flying overtaking maneuvers, which are the more critical maneuvers since the driver has committed to complete the maneuver, safety systems may assist the driver. By providing a means to predict the timing of the driver's decision to return, our models can inform the system of the probability of a driver's reaction, based on the predicted survival function. At the same time, the model can inform about the uncertainty of this probability.

For instance, a BSD system may consult the model to understand when the driver would return to decide on when to stop warning the driver of the cyclist being in the blind spot [23]. This might result in a possibly later return and increased safety for the cyclist, while taking into account the driver's preference (possibly in the presence of an oncoming vehicle). At the same time, an FCW system may warn to return if the driver has not yet reacted by knowing that the driver-model output has reached a high probability of returning. If the driver does not react to the warning, an automated steering system might help the driver maneuver the vehicle back into the original lane to evade a head-on collision with the oncoming vehicle.

Automated driving may also benefit from the driver model, allowing the system to behave similarly to manual driving in overtaking maneuvers. By predicting the reaction of different percentiles of drivers, the model may, therefore, increase drivers' trust in the system [40]. Not only the driver/passenger might benefit from a human-like behavior of the system, but also the overtaken cyclist or the oncoming vehicle may be more comfortable. Furthermore, models like the one presented in this study, may support the virtual assessment of automated driving by providing a human-reference model that the automated system's safety performance can be compared against [67].

Thanks to the chosen Bayesian approach, the model includes the full distribution of the fitted parameters and can predict a full distribution of the probability to return. This leaves system designers the choice of selecting the appropriate percentile of the predicted probability at which the system can assume the driver to have reacted. A Bayesian approach may also favor the personalization of safety systems [68], [69], by online learning: the model parameters could be updated with new data after each overtaking maneuver by the driver.

D. Limitations

Our results for the TT model are based on a limited set of the world's driving population. This may explain some of the differences between the TT and the ND model; for instance, the difference in magnitude of the parameter related to the longitudinal displacement of the cyclist, or the influence of the lateral distance. At the same time, these discrepancies may be due to the differences between the environments, where the TT airfield might have given drivers a more comfortable, open space for maneuvering, while the ND environment was constrained by clear road edges and connecting curves at the ends of the straight segment.

The ND data are naturally confounded with other factors [70] that were not accounted for in this study. For instance, the sight distance and solid line on the ground may have played an important role in influencing the driver's decision to return after having passed the cyclist. Furthermore, the positioning of the cyclist with respect to the lane edge (cycling within vs. outside of the lane) might have influenced driver behavior but was not investigated in this work. A more comprehensive study of the causal effects behind earlier returns may be done in future work.

E. Future Work

In this work, we focused on rural roads, which have higher speeds and higher impact consequences for crashes. However, future work should investigate whether and how the models developed in this work extrapolate to urban roads which may account for additional variables such as the presence of parked cars, different types of road markings, and a different speed level [18].

Furthermore, we excluded events from the ND data in which drivers passed both the cyclist and the oncoming vehicle during the passing phase, i.e., "squeezed" through the gap between the oncoming vehicle and the cyclist. The squeezing behavior is in line with what was reported in previous work [43], [44], [45]. Future work should understand when this behavior occurs and what the implications on cyclist safety are.

The survival models used in this work are based on a standard logistic regression which could be fitted to the limited amount of data available and allowed accounting for repeated trials by participants. Once more data from overtaking maneuvers become available, future work should investigate the advantages of using deep-learning-based survival models for possibly more accurate event prediction [60], [71]. Future work may also include additional explanatory variables, such as driver-control related signals (e.g., steering-wheel angle), which may be obtained from ND studies with instrumented vehicles [46].

Future studies should also develop and evaluate a safety system based on the driver model presented in this study. For this purpose, this study's method of recognizing the onset of the passing phase and returning phase using the lateral positioning of the ego vehicle may need to be evaluated. For instance, the steering-wheel angle or heading angle may be considered alternative signals to detect the onset. Furthermore, the Bayesian model may need to be assessed in terms of calculation speed, since using all MCMC samples will increase the complexity of calculating predictions. Using the median or mode of the posterior distributions may be a first step to avoiding heavy calculations involving all MCMC samples. Variational inference may be another promising approach to reduce the computational demand, especially if more data need to be fitted [72].

The current work only addresses one particular phase of the overtaking maneuver, the passing phase, arguably the most important phase in flying maneuvers due to the high risk of side-swiping the cyclist. However, since overtaking is a complex process that not only concerns the passing phase, it is important to consider all overtaking phases. Future work may investigate the feasibility of developing predictive models that combine all phases. The complete overtaking maneuver is a long process that requires a notable amount of distance traveled. Since the ND data in this study rarely captured all phases due to the limited length of the observed road stretch, future studies may need to obtain data from a more extended stretch or instrumented vehicles [46], [73].

Finally, while the reference model presented in this paper may create a valuable anchor for a machine to estimate the extent to which a driver usually behaves, it is open to discussion whether the model predictions are optimal or even enough to warrant safety. Moreover, in some unusual situations (e.g., the cyclist suddenly changing trajectory), the reference model presented in this paper may not predict a safe maneuver and potentially mislead active safety and automation. Future work should therefore address how threat assessment should weigh the inputs from reference models with other more traditional threat assessment estimations [74] to maximize safety.

V. Conclusion

In this work, we presented a probabilistic driver model that provides insights into how drivers time their returning after having passed a cyclist. The model can predict the probability of the driver returning using Bayesian survival regression with time-dependent inputs that reflect the driver's perception of the scenario. The model performed well on two different datasets (test track and naturalistic driving) in both in-sample and out-of-sample evaluations, despite the limited data size. From the parameters of the models based on both datasets, we could infer that drivers use the longitudinal displacement of the cyclist to decide when to return, and an oncoming vehicle accelerates this timing. Such models could run in real-time on vehicles and serve as a reference to compare the driver's actual behavior with the driver's usual behavior and include this information in the threat assessment for active safety. Collision-warning systems could integrate the model to improve their acceptance and assist the driver in maneuvering the vehicle back to the original lane to ensure the safety of the driver, the cyclist, and the oncoming traffic. In addition, automated-driving systems could leverage the driver model to behave more human-like to increase passengers' trust and cyclists' comfort.

Acknowledgment

The authors would like to thank Autoliv and Veoneer for the support in conducting the test-track experiment; Viscando for planning, conducting, and data processing of the naturalistic-driving study; Christian-Nils Åkerberg Boda for support with the collection and processing of the test-track data; and Tobias Aderum, Prateek Thalya, and Jordanka Kovaceva, for relevant discussions about the model.

The work was carried out at the Chalmers University of Technology, Gothenburg, Sweden.

Appendix

Model Summaries

Table III summarizes the numerical values of the distributions of the model coefficients.

References

[1] R. Buehler and J. Pucher, *Cycling for Sustainable Cities*. Cambridge, MA, USA: MIT Press, 2021.

[2] R. Buehler and J. Pucher, "COVID-19 impacts on cycling, 2019–2020," *Transp. Rev.*, vol. 41, no. 4, pp. 393–400, Jul. 2021, doi: 10.1080/01441647.2021.1914900.

[3] B. Büchel, A. D. Marra, and F. Corman, "COVID-19 as a window of opportunity for cycling: Evidence from the first wave," *Transp. Policy*, vol. 116, pp. 144–156, Feb. 2022, doi: 10.1016/j.tranpol.2021.12.003.

[4] J. J. de Hartog, H. Boogaard, H. Nijland, and G. Hoek, "Do the health benefits of cycling outweigh the risks?" *Environ. Health Perspect.*, vol. 118, no. 8, pp. 1109–1116, Aug. 2010, doi: 10.1289/ehp.0901747.

[5] A. Kullgren, H. Stigson, A. Ydenius, A. Axelsson, E. Engström, and M. Rizzi, "The potential of vehicle and road infrastructure interventions in fatal bicyclist accidents on Swedish roads—What can in-depth studies tell us?" *Traffic Injury Prevention*, vol. 20, pp. 7–12, Jun. 2019, doi: 10.1080/15389588.2019.1610171.

[6] European Road Safety Observatory, *European Commission (2022) Road Safety Thematic Report—Cyclists*, Eur. Commission, Directorate Gen. Transp., Brussels, Belgium, 2022.

[7] *Global Status Report on Road Safety*, World Health Org., Geneva, Switzerland, 2018.

[8] H. Stigson, A. Kullgren, and L.-E. Andersson, "Rural road design according to the safe system approach," in *The Vision Zero Handbook: Theory, Technology and Management for a Zero Casualty Policy*, K. E. Björnberg, M.-Å. Belin, S. O. Hansson, and C. Tingvall, Eds., Cham, Switzerland: Springer, 2020, pp. 1–25, doi: 10.1007/978-3-030-23176-7_36-1.

[9] K. Kircher, S. Forward, and H. W. Warner, "Cycling in rural areas: An overview of national and international literature," Actors Planning Processes, Swedish Nat. Road Transp. Res. Inst., Mobility, Linköping, Sweden, Tech. Rep., 1124A, 2022.

[10] S. Moll, G. López, and A. García, "Analysis of the influence of sport cyclists on narrow two-lane rural roads using instrumented bicycles and microsimulation," *Sustainability*, vol. 13, no. 3, p. 1235, Jan. 2021, doi: 10.3390/su13031235.

[11] M. Dozza, R. Schindler, G. Bianchi-Piccinini, and J. Karlsson, "How do drivers overtake cyclists?" *Accident Anal. Prevention*, vol. 88, pp. 29–36, Mar. 2016, doi: 10.1016/j.aap.2015.12.008.

[12] H. Farah, G. Bianchi Piccinini, M. Itoh, and M. Dozza, "Modelling overtaking strategy and lateral distance in car-to-cyclist overtaking on rural roads: A driving simulator experiment," *Transp. Res. F, Traffic Psychol. Behaviour*, vol. 63, pp. 226–239, May 2019, doi: 10.1016/j.trf.2019.04.026.

[13] A. Rasch, C.-N. Boda, P. Thalya, T. Aderum, A. Knauss, and M. Dozza, "How do oncoming traffic and cyclist lane position influence cyclist overtaking by drivers?" *Accident Anal. Prevention*, vol. 142, Jul. 2020, Art. no. 105569, doi: 10.1016/j.aap.2020.105569.

[14] I. Isaksson-Hellman and J. Töreki, "The effect of speed limit reductions in urban areas on cyclists' injuries in collisions with cars," *Traffic Injury Prevention*, vol. 20, pp. 39–44, Dec. 2019, doi: 10.1080/15389588.2019.1680836.

[15] P. D. Fernández, M. Lindman, I. Isaksson-Hellman, H. Jeppsson, and J. Kovaceva, "Description of same-direction car-to-bicycle crash scenarios using real-world data from Sweden, Germany, and a global crash database," *Accident Anal. Prevention*, vol. 168, Apr. 2022, Art. no. 106587, doi: 10.1016/j.aap.2022.106587.

[16] K. Gildea, D. Hall, and C. Simms, "Configurations of underreported cyclist-motorised vehicle and single cyclist collisions: Analysis of a self-reported survey," *Accident Anal. Prevention*, vol. 159, Sep. 2021, Art. no. 106264, doi: 10.1016/j.aap.2021.106264.

[17] E. Rubie, N. Haworth, D. Twisk, and N. Yamamoto, "Influences on lateral passing distance when motor vehicles overtake bicycles: A systematic literature review," *Transp. Rev.*, vol. 40, no. 6, pp. 754–773, May 2020, doi: 10.1080/01441647.2020.1768174.

[18] B. Beck et al., "Bicycling crash characteristics: An in-depth crash investigation study," *Accident Anal. Prevention*, vol. 96, pp. 219–227, Nov. 2016, doi: 10.1016/j.aap.2016.08.012.

[19] A. Rasch, S. Moll, G. López, A. García, and M. Dozza, "Drivers' and cyclists' safety perceptions in overtaking maneuvers," *Transp. Res. F, Traffic Psychol. Behaviour*, vol. 84, pp. 165–176, Jan. 2022, doi: 10.1016/j.trf.2021.11.014.

[20] S. J. Levulis, P. R. DeLucia, and J. Jupe, "Effects of oncoming vehicle size on overtaking judgments," *Accident Anal. Prevention*, vol. 82, pp. 163–170, Sep. 2015, doi: 10.1016/j.aap.2015.05.024.

[21] P. Thalya. (2021). *Making Overtaking Cyclists Safer: Driver Intention Models in Threat Assessment and Decision-making of Advanced Driver Assistance System*. [Online]. Available: https://research.chalmers.se/publication/523041/file/523041Fulltext.pdf

[22] W. Jeon and R. Rajamani, "A novel collision avoidance system for bicycles," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2016, pp. 3474–3479, doi: 10.1109/ACC.2016.7525451.

[23] A. Silla, L. Leden, P. Rämä, J. Scholliers, M. Van Noort, and D. Bell, "Can cyclist safety be improved with intelligent transport systems?" *Accident Anal. Prevention*, vol. 105, pp. 134–145, Aug. 2017, doi: 10.1016/j.aap.2016.05.003.

[24] U. Hassein, M. Diachuk, and S. M. Easa, "In-vehicle passing collision warning system for two-lane highways considering driver characteristics," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2672, no. 37, pp. 101–112, Dec. 2018, doi: 10.1177/0361198118795004.

[25] M. Brännström, F. Sandblom, and L. Hammarstrand, "A probabilistic framework for decision-making in collision avoidance systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 637–648, Jun. 2013, doi: 10.1109/TITS.2012.2227474.

[26] N. Lübbe, "Integrated pedestrian safety assessment: A method to evaluate combinations of active and passive safety," Ph.D. dissertation, Dept. Appl. Mech., Chalmers Univ. Technol., Gothenburg, Sweden, 2015. [Online]. Available: http://publications.lib.chalmers.se/records/fulltext/225504/225504.pdf

[27] M. L. Aust and S. Dombrovski, "Understanding and improving driver compliance with safety system," in *Proc. 23th Int. Tech. Conf. Enhanced Saf. Vehicles (ESV)*, 2013, pp. 1–7. [Online]. Available: https://trid.trb.org/view/1361362

[28] A. Rasch and M. Dozza, "Modeling drivers' strategy when overtaking cyclists in the presence of oncoming traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 2180–2189, Mar. 2022, doi: 10.1109/TITS.2020.3034679.

[29] M. A. Goodrich and E. R. Boer, "Designing human-centered automation: Trade-offs in collision avoidance system design," *IEEE Trans. Intell. Transp. Syst.*, vol. 1, no. 1, pp. 40–54, Mar. 2000, doi: 10.1109/6979.869020.

[30] A. Vahidi and A. Eskandarian, "Research advances in intelligent collision avoidance and adaptive cruise control," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 3, pp. 143–153, Sep. 2003.

[31] J. Sjöberg, E. Coelingh, M. Ali, M. Brännström, and P. Falcone, "Driver models to increase the potential of automotive active safety functions," in *Proc. 18th Eur. Signal Process. Conf.*, 2010, pp. 204–208.

[32] J. Kovaceva, J. Bärgman, and M. Dozza, "On the importance of driver models for the development and assessment of active safety: A new collision warning system to make overtaking cyclists safer," *Accident Anal. Prevention*, vol. 165, Feb. 2022, Art. no. 106513, doi: 10.1016/j.aap.2021.106513.

[33] A. Morando, T. Victor, and M. Dozza, "A reference model for driver attention in automation: Glance behavior changes during lateral and longitudinal assistance," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 2999–3009, Aug. 2019, doi: 10.1109/TITS.2018.2870909.

[34] Y. Zhang, C. Wu, C. Qiao, A. Sadek, and K. F. Hulme, "A cognitive computational model of driver warning response performance in connected vehicle systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 14790–14805, Sep. 2022, doi: 10.1109/TITS.2021.3134058.

[35] R. Wei, A. D. McDonald, A. Garcia, and H. Alambeigi, "Modeling driver responses to automation failures with active inference," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18064–18075, Oct. 2022, doi: 10.1109/TITS.2022.3155381.

[36] J. Morton, T. A. Wheeler, and M. J. Kochenderfer, "Analysis of recurrent neural networks for probabilistic modeling of driver behavior," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1289–1298, May 2017, doi: 10.1109/TITS.2016.2603007.

[37] Y. Yao, E. Atkins, M. Johnson-Roberson, R. Vasudevan, and X. Du, "BiTraP: Bi-directional pedestrian trajectory prediction with multi-modal goal estimation," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 1463–1470, Apr. 2021, doi: 10.1109/LRA.2021.3056339.

[38] C. Zhang and C. Berger, "Pedestrian behavior prediction using deep learning methods for urban scenarios: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 10, pp. 10279–10301, Nov. 2023, doi: 10.1109/TITS.2023.3281393.

[39] Z. Zhang, R. Tian, and Z. Ding, "TrEP: Transformer-based evidential prediction for pedestrian intention with uncertainty," in *Proc. Conf. Artif. Intell. (AAAI)*, vol. 37, Feb. 2023, pp. 3534–3542, doi: 10.1609/aaai.v37i3.25463.

[40] G. Abe, K. Sato, and M. Itoh, "Driver trust in automated driving systems: The case of overtaking and passing," *IEEE Trans. Hum.-Mach. Syst.*, vol. 48, no. 1, pp. 85–94, Feb. 2018, doi: 10.1109/THMS.2017.2781619.

[41] L. Bjelkeflo et al. (2018). *CHRONOS Part 1, FFI—Vinnova*. [Online]. Available: https://www.vinnova.se/contentassets/3f1ea323948d43ce90732c74dc10fd19/2016-02573.pdf

[42] A. Rasch, Y. Tarakanov, G. Tellwe, and M. Dozza, "Drivers passing cyclists: How does sight distance affect safety? Results from a naturalistic study," *J. Saf. Res.*, vol. 87, pp. 76–85, Dec. 2023.

[43] I. Walker, "Drivers overtaking bicyclists: Objective data on the effects of riding position, helmet use, vehicle type and apparent gender," *Accident Anal. Prevention*, vol. 39, no. 2, pp. 417–425, Mar. 2007, doi: 10.1016/j.aap.2006.08.010.

[44] S. C. Shackel and J. Parkin, "Influence of road markings, lane widths and driver behaviour on proximity and speed of vehicles overtaking cyclists," *Accident Anal. Prevention*, vol. 73, pp. 100–108, Dec. 2014, doi: 10.1016/j.aap.2014.08.015.

[45] F. Feng, S. Bao, R. C. Hampshire, and M. Delp, "Drivers overtaking bicyclists—An examination using naturalistic driving data," *Accident Anal. Prevention*, vol. 115, pp. 98–109, Jun. 2018, doi: 10.1016/j.aap.2018.03.010.

[46] J. Kovaceva, G. Nero, J. Bärgman, and M. Dozza, "Drivers overtaking cyclists in the real-world: Evidence from a naturalistic driving study," *Saf. Sci.*, vol. 119, pp. 199–206, Nov. 2019, doi: 10.1016/j.ssci.2018.08.022.

[47] E. I. Vlahogianni, "Modeling duration of overtaking in two lane highways," *Transp. Res. F, Traffic Psychol. Behaviour*, vol. 20, pp. 135–146, Sep. 2013, doi: 10.1016/j.trf.2013.07.003.

[48] C. Gromke and B. Ruck, "Passenger car-induced lateral aerodynamic loads on cyclists during overtaking," *J. Wind Eng. Ind. Aerodynamics*, vol. 209, Feb. 2021, Art. no. 104489, doi: 10.1016/j.jweia.2020.104489.

[49] G. Rodríguez, *Lecture Notes on Generalized Linear Models*, 2007. [Online]. Available: https://grodri.github.io/glms/notes/

[50] F. Bella and F. Gulisano, "A hazard-based model of the motorcyclists' overtaking duration," *Accident Anal. Prevention*, vol. 141, Jun. 2020, Art. no. 105522, doi: 10.1016/j.aap.2020.105522.

[51] J. D. Singer and J. B. Willett, "It's about time: Using discrete-time survival analysis to study duration and the timing of events," *J. Educ. Statist.*, vol. 18, no. 2, pp. 155–195, 1993.

[52] J. D. Singer and J. B. Willett, *Applied Longitudinal Data Analysis*. London, U.K.: Oxford Univ. Press, May 2003, doi: 10.1093/acprof:oso/9780195152968.001.0001.

[53] A. S. Kurz, *Applied Longitudinal Data Analysis in BRMS and the Tidyverse, Version 0.0.1*, 2021. [Online]. Available: https://bookdown.org/content/4253/

[54] P.-C. Bürkner, "Brms: An R package for Bayesian multilevel models using Stan," *J. Stat. Softw.*, vol. 80, no. 1, pp. 1–28, 2017, doi: 10.18637/jss.v080.i01.

[55] J. K. Kruschke, "Bayesian analysis reporting guidelines," *Nature Hum. Behaviour*, vol. 5, no. 10, pp. 1282–1291, Aug. 2021, doi: 10.1038/s41562-021-01177-7.

[56] D. Makowski, M. S. Ben-Shachar, S. H. A. Chen, and D. Lüdecke, "Indices of effect existence and significance in the Bayesian framework," *Frontiers Psychol.*, vol. 10, pp. 1–14, Dec. 2019, doi: 10.3389/fpsyg.2019.02767.

[57] A. Vehtari, A. Gelman, and J. Gabry, "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," *Statist. Comput.*, vol. 27, no. 5, pp. 1413–1432, Sep. 2017, doi: 10.1007/s11222-016-9696-4.

[58] R. Moineddin, F. I. Matheson, and R. H. Glazier, "A simulation study of sample size for multilevel logistic regression models," *BMC Med. Res. Methodol.*, vol. 7, no. 1, p. 34, Dec. 2007, doi: 10.1186/1471-2288-7-34.

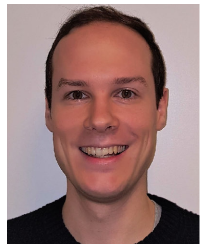
[59] R. Zaki, A. Barabadi, J. Barabady, and A. Nouri Qarahasanlou, "Observed and unobserved heterogeneity in failure data analysis," *Proc. Inst. Mech. Eng., O, J. Risk Rel.*, vol. 236, no. 1, pp. 194–207, Feb. 2022, doi: 10.1177/1748006x211022538.

[60] K. Suresh, C. Severn, and D. Ghosh, "Survival prediction models: An introduction to discrete-time modeling," *BMC Med. Res. Methodol.*, vol. 22, no. 1, p. 207, Dec. 2022, doi: 10.1186/s12874-022-01679-6.

[61] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression* (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: Wiley, 2013. [Online]. Available: https://books.google.se/books?id=64JYAwAAQBAJ

[62] M. Berger and M. Schmid, "Semiparametric regression for discrete time-to-event data," *Stat. Model.*, vol. 18, nos. 3–4, pp. 322–345, Jun. 2018, doi: 10.1177/1471082x17748084.

[63] A. P. Dawid, "A Bayesian look at nuisance parameters," *Trabajos de Estadistica Y de Investigacion Operativa*, vol. 31, no. 1, pp. 167–203, Feb. 1980, doi: 10.1007/bf02888351.
[64] R. L. Wasserstein and N. A. Lazar, "The ASA statement on p-values: Context, process, and purpose," *Amer. Statistician*, vol. 70, no. 2, pp. 129–133, Apr. 2016, doi: 10.1080/00031305.2016.1154108.
[65] G. F. Bianchi Piccinini, C. Moretto, H. Zhou, and M. Itoh, "Influence of oncoming traffic on drivers' overtaking of cyclists," *Transp. Res. F, Traffic Psychol. Behaviour*, vol. 59, pp. 378–388, Nov. 2018, doi: 10.1016/j.trf.2018.09.009.
[66] A. L. Schwab and J. P. Meijaard, "A review on bicycle dynamics and rider control," *Vehicle Syst. Dyn.*, vol. 51, no. 7, pp. 1059–1090, Jul. 2013, doi: 10.1080/00423114.2013.793365.
[67] C. Roesener, J. Hiller, H. Weber, and L. Eckstein, "How safe is automated driving? Human driver models for safety performance assessment," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–7, doi: 10.1109/ITSC.2017.8317706.
[68] M. Hasenjäger, M. Heckmann, and H. Wersing, "A survey of personalization for advanced driver assistance systems," *IEEE Trans. Intell. Vehicles*, vol. 5, no. 2, pp. 335–344, Jun. 2020, doi: 10.1109/TIV.2019.2955910.
[69] S. Schnelle, J. Wang, H.-J. Su, and R. Jagacinski, "A personalizable driver steering model capable of predicting driver behaviors in vehicle collision avoidance maneuvers," *IEEE Trans. Hum.-Mach. Syst.*, vol. 47, no. 5, pp. 625–635, Oct. 2017, doi: 10.1109/THMS.2016.2608930.
[70] J. Bärgman, "Methods for analysis of naturalistic driving data in driver behavior research," Ph.D. dissertation, Dept. Appl. Mech., Chalmers Univ. Technol., Gothenburg, Sweden, 2016. [Online]. Available: http://publications.lib.chalmers.se/records/fulltext/244575/244575.pdf
[71] M. F. Gensheimer and B. Narasimhan, "A scalable discrete-time survival model for neural networks," *PeerJ*, vol. 7, p. e6257, Jan. 2019, doi: 10.7717/peerj.6257.
[72] C. Zhang, J. Bütepage, H. Kjellström, and S. Mandt, "Advances in variational inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 2008–2026, Aug. 2019, doi: 10.1109/TPAMI.2018.2889774.
[73] P. Sun et al., "Scalability in perception for autonomous driving: Waymo open dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2446–2454.
[74] J. Dahl, G. R. de Campos, C. Olsson, and J. Fredriksson, "Collision avoidance: A literature review on threat-assessment techniques," *IEEE Trans. Intell. Vehicles*, vol. 4, no. 1, pp. 101–113, Mar. 2019, doi: 10.1109/TIV.2018.2886682.



Alexander Rasch received the B.Sc. degree in mechanical engineering from RWTH Aachen University, Aachen, Germany, in 2016, and the M.Sc. degree in systems, control, and mechatronics from the Chalmers University of Technology, Gothenburg, Sweden, in 2018, where he is currently pursuing the Ph.D. degree, with a focus on computational models for driver behavior in interaction with pedestrians and cyclists, mainly in overtaking scenarios. The goal of his research is to develop driver models that can improve advanced driver assistance systems.

Carol Flannagan received the M.A. degree in statistics and the Ph.D. degree in mathematical psychology from the University of Michigan. She is currently a Research Professor with the University of Michigan Transportation Research Institute (UMTRI) and an affiliated Associate Professor with Chalmers University. Her work in transportation research encompasses the analysis of a wide variety of transportation-related data and the development of innovative statistical methods for transportation research.

Marco Dozza received the Ph.D. degree in bioengineering from the University of Bologna, Italy, in collaboration with Oregon Health and Science University, Portland, OR, USA, in 2007. After graduation, he worked as a System Developer with Volvo Technology, a research and innovation company inside the Volvo Group. Since 2009, he has been with the Chalmers University of Technology, Gothenburg, Sweden, where he is currently a Full Professor and leads the research group for crash analysis and prevention. He is an Examiner for the course Active Safety in the Master's Program for Mobility Engineering and the course micromobility for a Sustainable Future with Chalmers Tracks.